Extending the Integrated Land-Use and Transport model framework: Spatial econometrics and policy evaluation

PRESENTED IN 12/05/2014
SCHOOL OF RURAL AND SURVEYING ENGINEERING
LABORATORY OF TRANSPORTATION ENGINEERING

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

TO OBTAIN THE DEGREE DOCTOR OF PHILOSOPHY

by

Dimitrios Efthymiou

accepted after the proposition of the jury:

Assist. Prof. Constantinos Antoniou, Doctoral research supervisor, NTUA
Prof. Basil Psarianos, Supervising committee member, NTUA
Prof. Paul Waddell, Supervising committee member, UC Berkeley
Prof. Antonis Stathopoulos, NTUA
Prof. Amalia Polydoropoulou, University of Aegean
Prof. Kay Axhausen, ETH
Prof. Michel Bierlaire, EPFL

Athens
2014

All models are wrong, but some are useful.
— George E. P. Box

Acknowledgements

I would first like to express my great appreciation to my supervisor Professor Constantinos Antoniou, for his advice, guidance and enthusiastic encouragement; he gave me the opportunity to live a unique experience during my doctoral studies.

I would also like to thank Professor Michel Bierlaire for giving me the opportunity to visit TRANSP-OR lab at EPFL for a long period of my studies, learning lots of things in many fields.

I wish to thank the other members of the doctoral supervising committee, Professor Paul Waddell for his valuable comments and advice on Integrated Land-Use and Transport modeling, and Professor Basil Psarianos for his support and advice in transportation engineering, since the years of my Diploma. Moreover, Professor Kay Axhausen, Professor Amalia Polydoropoulou and Professor Ioannis Golias for their assistance and valuable feedback.

I would also like to thank my fellow researchers at the TRANSP-OR lab, Amanda, Antonin, Aurelie, Bilal, Eva, Flurin, Jianghang, Jingmin, Marija, Nitish, Tomáš and Xinjun, the great office-mates Bilge and Ricardo, Sophia from LAVOC, and the best Administrative Assistant Marianne Ruegg, with many of whom we had good times beyond academics.

I am grateful to the Special Fund for Research Grants of NTUA, the Swiss Government Excellence scholarship, EPFL, and the Sustaincity project, for the financial support provided for my doctoral research.

At last, I would like to thank Alexia for her patience and support all these years. Most of all I thank my family for their support and encouragement; my brother Alexandros, my mother Sofia, and my father Lakis who left us before the completion of this Dissertation.

Athens, 12 May 2014 D. E.

Abstract

In this research, the methodological framework of the Integrated Land-Use and Transport (LUTI) models is extended at every level. The objective of developing and implementing LUTI models is to predict the direct and indirect impacts of –transport and land-use– policies, on the environment, the society and the economy. The proposed improvements aim to increase the predicting capabilities of the current LUTI models, exploiting the strengths of microsimulation in three dimensions, agents, space and time, in order to render them flexible platforms for policy evaluation. The effects of the current economic crisis are discussed and explored throughout the doctoral dissertation.

Aiming to decrease the required budget for a LUTI model development, public on-line data are used to a large part of the analysis. Moreover, a graph-theoretic solution for associations generation in synthetic simulation is suggested. Different types of spatial econometric models are used for the development of real estate price models, which form fundamental component of every LUTI model. Urban quality indicators (i.e. accessibility, population segregation, economic viability, available open space, housing affordability, land-use and social mix, and building density) are effectively employed manifesting the benefits of trans-disciplinary collaboration in urban planning. In this research, a policy evaluation methodology based on distributions rather than single aggregate measures of quality indicators is proposed.

The results indicate that spatial econometrics effectively remove the spatial autocorrelation and achieve higher accuracy than the traditional linear regression, in predicting the dwelling prices. The impact of transportation infrastructure locations on real estate purchase prices and rents differs, depending on the type of the transit system. Qualitative transit infrastructure has preserved the real estate prices at higher levels during the crisis. Synthetic populations and real, on-line, crowdsourced data can efficiently be used for the development of LUTI models. Finally, agent-based LUTI models provide an opportunity for the development of an improved, flexible policy evaluation platform.

Περίληψη

Κίνητρο για την έρευνα. Οι αστικές περιοχές είναι πολύπλοκες συνθέσεις οικονομικών, κοινωνικών, περιβαλλοντικών και πολιτικών χαρακτηριστικών που αλληλεπιδρούν και εξελίσσονται. Σε μια προσπάθεια μείωσης της πολυπλοκότητας και δημιουργίας ενός ξεκάθαρου και κατανοητού πλαισίου λειτουργίας τους, κατά καιρούς έχουν προταθεί θεωρητικά και μαθηματικά μοντέλα. Η θεωρητική προσέγγιση του προβλήματος μπορεί κάποιες φορές να ρίχνει φως στον τρόπο αλληλεπίδρασης μεταξύ των μεταφορών και της αστικής ανάπτυξης, ωστόσο, τα αποτελέσματα βασίζονται σε υπεραπλουστευμένες παραδοχές. Απο την άλλη, με μαθηματικά μοντέλα και τεχνικές προσομοίωσης μπορεί να επιτευχθεί πιο αξιόπιστη αναπαράσταση της κατάστασης, ωστόσο τα αποτελέσματα χρειάζεται να απλοποιηθούν με σκοπό να γίνουν περισσότερο κατανοητά. Τα Γεωγραφικά Συστήματα Πληροφοριών (ΓΣΠ) συμβάλλουν προς αυτή την κατεύθυνση.

Προγραμματισμένες αστικές πολιτικές και απρόβλεπτες πολιτικές καταστάσεις –όπως η οικονομική κρίση που βιώνουμε– σχετίζονται άμεσα και έμμεσα με τα κοινωνικά, οικονομικά και περιβαλλοντικά χαρακτηριστικά των αστικών περιοχών. Η ανάγκη για πιο ακριβή πρόβλεψη των συνεπειών αυτών των πολιτικών στην αστική δομή, έχει αναζωπυρώσει το ενδιαφέρον για την ανάπτυξη Ολοκληρωμένων Μοντέλων Χρήσεων-Γης και Μεταφορών (ΟΜΧΓΜ). Τα τελευταία χρόνια, η ανάπτυξη ΟΜΧΓΜ έχει διευκολυνθεί απο την εξέλιξη της τεχνολογίας των υπολογιστών και του μαθηματικού προγραμματισμού· ωστόσο, πολλές προκλήσεις παραμένουν εξαιτίας των υψηλών απαιτήσεων των λεπτομερών δημογραφικών και κυκλοφοριακών μοντέλων.

Αρκετές αβεβαιότητες σχετικά με τον αποδοτικότητα και τις ικανότητες πρόλεψης παραμένουν αναπάντητες, διατηρώντας το έδαφος πρόσφορο για κριτική. Μέχρι σήμερα έχουν αναπτυχθεί ΟΜΧΓΜ για διάφορες Μητροπολιτικές περιοχές, είτε για ερευνητικούς σκοπούς, είτε για τη χάραξη πολιτικών. Τα μοντέλα αυτά διαφέρουν ως προς τη δομή και το σχεδιασμό τους, τον απαιτούμενο όγκο δεδομένων, τη λειτουργικότητα και την εφαρμοσιμότητά τους.

Παρά τη συνεχή ανάπτυξη μεθοδολογικά πολύπλοκων μοντέλων, οι ικανότητες πρόβλεψής τους δεν παρουσιάζουν σημαντική βελτίωση. Απο την άλλη, μεθοδολογίες που έχουν εφαρμοστεί επιτυχώς σε άλλους τομείς, όπως τα χωρικά οικονομετρικά μοντέλα (ΧΟΜ), δεν έχουν αξιοποιηθεί μέχρι σήμερα. Τα ωφέλη απο την συνεχώς αυξανόμενη λεπτομέρεια στη δομή των ΟΜΧΓΜ σε τρείς διαστάσεις –χώρο, χρόνο και άτομο– δεν έχουν αξιοποιηθεί επαρκώς για την ανάπτυξη επιτυχών μεθόδων αξιολόγησης πολιτικών. Επιπλέον, τα πλεονεκτήματα της χρήσης ανοιχτών δεδομένων και συνθετικών πληθυσμών παραμένουν αναξιοποίητα.

Τα ΟΜΧΓΜ συνδιάζουν μοντέλα επιλογής τοποθεσίας κατοικίας/εργασίας, αξιών ακινήτων, και κυκλοφοριακά μοντέλα, για την εκτίμηση και πρόβλεψη των κοινωνικών, οικονομικών και περιβαλλοντικών επιπτώσεων των πολιτικών, στις αστικές περιοχές. Για την αναπαράσταση των πολύπλοκων αστικών συστημάτων, απαιτούνται σύνθετες μεθοδολογίες και λεπτομερή δεδομένα. Η παράλειψη κρίσιμων στοιχείων της αστικής δομής και η έλλειψη ικανοτήτων πρόβλεψης του μοντέλου, οδηγεί σε λανθασμένες εκτιμήσεις, που σε περπτώσεις συγκοινωνιακών επενδύσεων μεταφράζεται σε απώλειες εκατομμυρίων ευρώ και κοινωνική ανισότητα.
Διατύπωση του προβλήματος και αντικείμενο της έρευνας. Τα ΟΜΧΓΜ χρησιμοποιούνται για την πρόβλεψη των επιπτώσεων συγκοινωνιακών έργων υψηλού προϋπολογισμού στην αστική ανάπτυξη· ωστόσο, και τα ίδια απαιτούν υψηλό κόστος για την υλοποίησή τους. Η συλλογή και (προ-)επεξεργασία δεδομένων απαιτούν τους περισσότερους απο τους πόρους (χρόνο και χρήμα) που χρειάζονται για την ανάπτυξη ενός ΟΜΧΓΜ.
Παρά τη συνεχή αύξηση διαθέσιμων, δωρεάν δεδομένων στο διαδίκτυο, η χρήση τους στα ΟΜΧΓΜ δεν έχει ακόμα αναγνωριστεί. Επιπλέον, τεχνικές ανάπτυξης σύνθετικών πληθυσμών αναπτύσσονται συνεχώς· ωστόσο, πολλά πρόσφατα ΟΜΧΓΜ χρησιμοποιούν ακόμα ακριβά δεδομένα απογραφών.
Νέες πολύπλοκες μεθοδολογίες για ΟΜΧΓΜ συνεχίζουν να αναπτύσσονται, χωρίς να επιτυγχάνεται η αναμενόμενη βελτίωση των ικανοτήτων πρόβλεψης. Η αδυναμία αυτή δε δικαιολογεί τον αριθμό των ερευνητών, το χρόνο και χρήμα που έχει επενδυθεί στην ανάπτυξη εργαλείων πρόβλεψης μεγαλύτερης ακρίβειας. Επιπλέον, παρά το γεγονός ότι τα ΟΜΧΓΜ προορίζονται για αξιολόγηση πολιτικών, δεν έχει υλοποιηθεί κάποιο πλαίσιο το οποίο να αξιοποιεί τα ωφέλη της μικροπροσομοίωσης στην κατεύθυνση αυτή.
Οι επιπτώσεις των αναδυόμενων βιώσιμων συστημάτων μεταφορών –όπως τα μοιραζόμενα αυτοκίνητα– στην αστική ανάπτυξη, καθώς επίσης των απρόβλεπτων οικονομικών καταστάσεων –όπως η ύφεση– δεν έχουν διερευνηθεί σε βάθος. Επιπλέον, δεν έχουν διερευνηθεί οι κατευθύνσεις χρήσης προηγμένων μοντέλων συμπεριφοράς, που συνδυάζουν μοντέλα διακριτών επιλογών και λανθανουσών μεταβλητών στην αξιολόγηση πολιτικών.
Ο σκοπός της παρούσας διδακτορικής έρευνας είναι να διερευνήσει, αναπτύξει και προτείνει μεθοδολογίες συλλογής δεδομένων, μοντελοποίησης και αξιολόγησης πολιτικών για Ολοκληρωμένα Μοντέλα Χρήσεων-Γης και Μεταφορών. Οι προτεινόμενες προσεγγίσεις αποσκοπούν στη μείωση του κόστους και την αύξηση της ικανότητας πρόβλεψης, εκμεταλλευόμενες τα πλεονεκτήματα της μικροπροσομοίωσης σε τρείς διαστάσεις: άτομο, χώρο και χρόνο.
Συνεισφορά της έρευνας. Η παρούσα διδακτορική έρευνα συμβάλλει με τα ακόλουθα στη διεύρυνση του μεθοδολογικού πλαισίου των ΟΜΧΓΜ:

- Διερευνάται η προπτική χρήσης δεδομένων συλλεγμένων απο τα πλήθη στην ανάπτυξη ΟΜΧΓΜ. Ηλεκτρονικά δεδομένα ακινήτων συλλέγονται και χρησιμοποιούνται για την ανάπτυξη προτύπων. Δεδομένα χρήσεων - γης/κάλυψης εξάγονται απο δορυφορικές απεικονίσεις με μεθόδους τηλεπισκόπισης, με σκοπό να χρησιμοποιηθούν για την ανάπτυξη μοντέλων. Σκοπεύοντας στη μείωση της εξάρτησης απο βάσεις δεδομένων υψηλού κόστους, προτείνεται μια γραφο-θεωρητική προσέγγιση για δημιουγία συσχετίσεων σε συνθετικό πληθυσμό.

- Προτείνεται ένα πλαίσιο ενσωμάτωσης χωρικών οικονομετρικών μοντέλων σε ΟΜΧΓΜ. Πιο συγκεκριμένα:
  - Με τη χρήση χωρικών οικονομετρικών μοντέλων μοντελοποιούνται οι αξίες των ακινήτων. Τα μοντέλα αυτά οδηγούν σε καλύτερη ακρίβεια συγκριτικά με τη μέθοδο των ελαχίστων τετραγώνων, και απαλείφουν αποτελεσματικά τη χωρική αυτοσυσχέτιση. Αποσκοπώντας στη μέτρηση των επιπτώσεων των συγκοινωνιακών υποδομών και πολιτικών στις αξίες αγοράς και ενοικίασης, αναπτύχθηκαν δύο περιπτωσιακές μελέτες στην Ελλάδα, για την Αθήνα και τη Θεσσαλονίκη.
  - Προτυποποιήθηκε η αλλαγή χρήσεων - γης/κάλυψης με τη χρήση χωρικού μοντέλου διακριτών επιλογών, επιτυγχάνοντας καλύτερη εφαρμογή απο τη γενικευμένη γραμμική παρεμβολή. Διερευνώνται οι επιπτώσεις συγκοινωνιακών υποδομών μεγάλης έκτασης στην αλλαγή χρήσεων - γης στην Αθήνα.
  - Αναπτύχθηκε μεθοδολογία για τη χωρική κατανομή συγκοινωνιακών υποδομών (σταθμών φόρτισης ηλεκτρικών αυτοκινήτων, μοιραζόμενων αυτοκινήτων κλπ.) που βασίζεται σε χωρικά οικονομετρικά μοντέλα και πολυκριτηριακή ανάλυση.
- Αναπτύχθηκε μεθοδολογία για ποιοτική και ποσοτική αξιολόγηση πολιτικών που βασίζεται σε δείκτες, εκμεταλλευόμενη τα οφέλη της μικροπροσομοίωσης σε τρείς διαστάσεις (άτομο, χώρο και χρόνο). Διερευνάται η χρήση ατομο-βασικών δεικτών για την προσβασιμότητα, ανισότητα, οικονομία/επενδύσεις και κοινωνική ποιότητα. Η προτεινόμενη μεθοδολογία βασίζεται σε χωρικές κατανομές και όχι σε μοναδικές, γενικευμένες μετρήσεις.
- Διερευνώνται οι επιπτώσεις των πολιτικών δημόσιων συγκοινωνιών στην αστική ανάπτυξη, χρησιμοποιώντας υβριδικά μοντέλα διακριτών επιλογών και λανθανουσών μεταβλητών.
- Διερευνώνται οι συνέπειες της χρηματοοικονομικής κρίσης (παράγωγα ποικίλων πολιτικών) στις συγκοινωνιακές υποδομές/πολιτικές και στις αξίες ακινήτων.

Προτάσεις για το μέλλον

- **Προσομοίωση με χρήση χωρικών οικονομετρικών μοντέλων** Η μεθοδολογία των χωρικών οικονομετρικών μοντέλων πρέπει να ενσωματωθεί στα ΟΜΧΓΜ της επόμενης γενιάς, ώστε να χρησιμοποιηθεί όχι μόνο για την ανάπτυξη μοντέλων εκτίμησης ακινήτων, αλλά και για πρόβλεψη. Ένας προβληματισμός που συχνά δημιουργείται στην ιδέα χρήσης χωρικών οικονομετρικών μοντέλων σε ΟΜΧΓΜ, αφορά τον αριθμό των κοντινότερων γειτόνων που πρέπει να χρησιμοποιηθουν για την κατασκευή των πινάκων βαρών. Μια ενδεχόμενη λύση στο πρόβλημα αυτό θα μπορούσε να είναι η εφαρμογή αρχικά ενός σχετικά απλού τύπου μοντέλου, όπως το χωρικό μοντέλο σφάλματος, και έπειτα η επανεκτίμησή του κάθε φορά που εκτελείται η διαδικασία προσομοίωσης, χρησιμοποιώντας διαφορετικούς πίνακες βαρών.

- **Ανάλυση των επιπτώσεων της οικονομικής κρίσης** Στην παρούσα εργασία αναλύονται οι επιπτώσεις της κρίσης στις αξίες των ακινήτων, αλλά και στην ικανοποίηση των επιβατών της δημόσιας συγκοινωνίας ως προς την ποιότητα των παρεχόμενων υπηρεσιών. Ωστόσο, το ενδιαφέρον για τη διερεύνηση των επιπτώσεων ακραίων οικονομικών καταστάσεων στη συμπεριφορά των ανθρώπων και την αστική ανάπτυξη, μεγαλώνει διαρκώς. Απο την αποκτειθείσα εμπειρία και με τη χρήση κατάλληλων δεδομένων, οι συνέπειες της κρίσης μπορούν να προσομοιωθούν σε περιβάλλον ΟΜΧΓΜ, και να επικυρωθούν, βελτιώνοντας τις δυνατότητες προβλεψης. Η γνώση του τρόπο αντίδρασης των αστικών περιοχών σε παρόμοιες καταστάσεις, θα αποτελούσε πολίτιμο εφόδιο στο μέλλον.

- **Κοινωνική ευημερία** Στην παρούσα διατριβή αναλύονται οι κατανομές δεικτών ανισότητας και προσβασιμότητας που σχηματίζονται απο δεδομένα προσομοίωσης. Ωστόσο, η κοινωνική ευημερία αποτελεί εναν οικονομικό δείκτη που αποκτά ιδιαίτερο ενδιαφέρον τα τελευταία χρόνια. Στο πλαίσιο της έρευνας του εργαστηρίου Συγκοινωνιακής Τεχνικής της Σχολής Αγρονόμων Τοπογράφων Μηχανικών τα τελευταία χρόνια, ο δείκτης Συνάρτηση Κοινωνικής Ευημερίας (ΣΚΕ) επεκτάθηκε και προσαρμόστηκε σε Ευρωπαϊκές πόλεις, ενώ αναπτύχθηκε ένα εργαλείο υπολογισμού του. Απαιτείται περεταίρω διερεύνηση της χρήσης του, καθώς και ανάπτυξη λεπτομερέστερης μορφής του.

- **Ψυχομετρικοί δείκτες** Η ανάπτυξη μεθοδολογιών μοντέλων συμπεριφοράς τα τελευταία χρόνια, όπως τα υβριδικά μοντέλα διακριτών επιλογών και λανθανουσών μεταβλητών, δημιουργούν νέες προοπτικές για την προτυποποίηση συμπεριφορών. Οι λανθάνουσες συμπεριφορές πρέπει να λαμβάνονται υπόψιν στα μοντέλα επιλογής τοποθεσίας (π.χ. κατοικίας, εργασίας), καθώς μπορεί να επηρεάσουν κατα τη λήψη τηε απόφασης.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Urban areas are complex structures of economical, social, environmental and political characteristics that interact and continuously evolve. In an attempt to reduce the complexity and create a clear and understandable coding of some aspect of their structure, several theoretical and mathematical models have been suggested. The theoretical approach of the problem may sometimes shed some light on the way of interaction between transport and urban development; however, the results are based on oversimplified assumptions that lack of trustworthiness (Waddell and Ulfarsson, 2004). On the other hand, with mathematical modelling and simulation techniques, a more reliable representation of the situation is achieved, but the results need to be simplified in order to be more comprehensible. The integration of Geographical Information Systems (GIS) in visualising the simulation outcomes, has played an important role in this regard.

Planned urban policies and unintended political situations –such as the current economic crisis– are directly and indirectly related with the social, economical and environmental characteristics of urban areas. The need for more accurate forecasting of the impact of these policies on urban structure, has enforced the interest for developing Integrated Land-Use and Transport –or Land-Use Transport Interaction– models (LUTI).

Within the last few years, LUTI model development has been facilitated by the evolution of computer technology and mathematical programming; however, it remains challenging due to the requirement of extremely disaggregate demographic, employment, building and transport data. Many uncertainties about their efficiency and forecasting capabilities remain unanswered, preserving the ground under the LUTI models fertile for criticism. In an extensive review of known LUTI models Hunt et al. (2005) points out the pros and cons of each. Curtis (2011) reviews the inefficiencies of the current LUTI models in measuring the accessibility of public transport. Until now, there have been developed LUTI models for different metropolitan areas, either for research purposes or for policy making. These models differ as to their content, structure, theory, modelling methodology, amount of data required, way of calibration and validation, functionality and applicability.

Nevertheless, despite the continuous development of methodologically complicated LUTI models, their forecasting performance is questionable due to the absence of validation evidence. Existing modelling methodologies widely applied in other research areas – such as spatial econometric models– have not been successfully implemented in LUTI models, yet. The benefits of the increasingly disaggregate form of LUTI models in three dimensions, agents, space and time, have not been exploited in favor of developing more successful policy evaluation techniques. The margin of using open, crowdsourced data and synthetic populations is still wide. Finally, advanced behavioural modeling methods used for transport policy analysis are still absent from LUTI models.

1.2 Methodological framework

Integrated Land-Use and Transport –or else Land-Use and Transport Interaction– (LUTI) models are sets of interacting sub-models that are used to forecast the urban development, after the implementation of a policy. LUTI models form the mathematical, quantitative approach for predicting the direct and indirect effects of urban planning policies on the environment, society and economy that the theoretical, qualitative approach is unable to predict itself.

LUTI models combine agent (households and jobs) location choice, transition, real estate price (rent and purchase) and development, and transport (activity and agent-based) models, in an integrated framework. The sub-models are first separately estimated using base-year data, and then applied in simulations of different time intervals (e.g. every year). For every simulated period, the dataset generated from the previous simulation is being used. Due to the complicated nature of the interacted agents in agent-based LUTI models, disaggregate social and structural data are required.

Figure 1.1 depicts the methodological framework of the LUTI model UrbanSim (Waddell, 2002), which forms the state-of-the-art in this field. The methodological improvements of the LUTI framework proposed in this dissertation, are dealing with the data collection, model development, modelling methodology and policy evaluation techniques.

Input data for LUTI models are the: structural characteristics of buildings, demographic characteristics of households, employment, past/planned infrastructure development, and travel data. However, due to the high purchase cost of official census databases, the use of alternative data such as synthetic populations and on-line real estate data should be further investigated.

The role of *real estate price model* is to predict the purchase and/or rent prices of the residential –or not– real estate at every step of the simulation process. Real estate prices determine –and are determined by– household and job location choice, and they are highly affected by the implementation of transport policies and infrastructure projects. In LUTI models, property prices are usually modelled with ordinary least squares, each as semi-log linear regression which is the case in UrbanSim (Waddell, 2002), or within a bid-choice model (Martinez and Donoso, 2010). However, with these modelling techniques, spatial autocorrelation is not taken into consideration.

The LUTI simulation output is used for transport and land-use policy assessment (e.g. cordon/area charging, public transport investment, densification policies). Policy evaluation is based on economic, sustainability and accessibility, indicators, usually computed in aggregate level (eg. Litman, 2011a). As a result, with the current policy evaluation techniques the strengths of microsimulation in three dimensions, agents, space and time, are not sufficiently exploited.

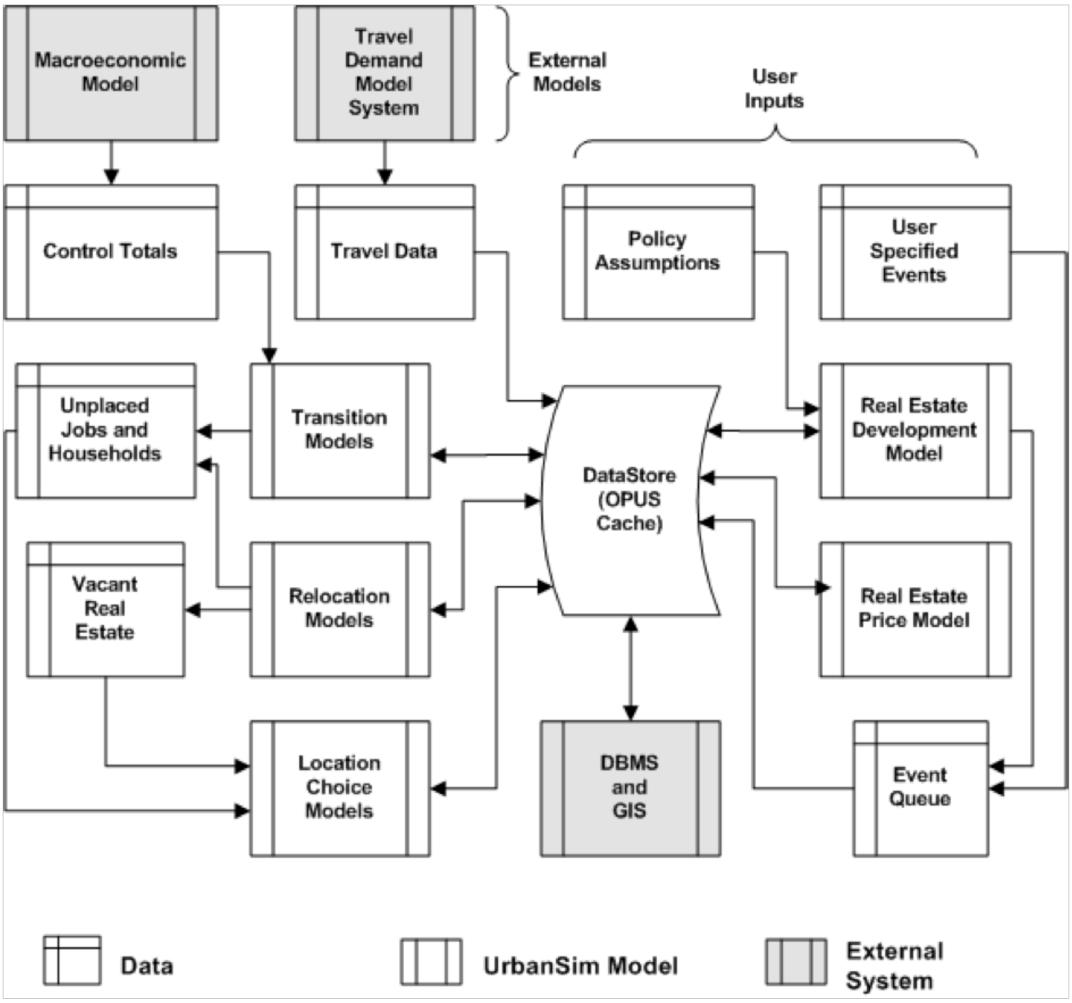


Figure 1.1: Model components and data flow. Adapted from OPUS (2011)

1.3 Problem definition and research objective

LUTI models combine location choice, real estate price and traffic models to estimate and forecast the social, economic and environmental state of urban areas, resulting by the implementation of a policy. In order to reproduce the complex multivariate structure of urban systems in computing environment, complicate modeling methodologies and extremely disaggregate data are required. The omission of critical determinants of the urban structure, and the inability of the model to make accurate predictions, result in unsuccessful forecasts. In case of transport policy evaluation, the forecasting failure amounts to multi-million euro losses and social disparity.

LUTI models are used to forecast the effects of high-budget transportation projects on urban development, but are themselves costly projects that needs to be developed. Data collection, pre-processing and processing probably require most of the resources (time and money) needed for a LUTI model development.

Despite the increasing availability of free on-line data, their use in LUTI models has not yet met the expected recognition. Moreover, population synthesis techniques are continuously emerging; however, many recent LUTI models are still using expensive census data.

New complex modeling methodologies for LUTI models keep being developed, without significant improvement of their forecasting capabilities. This weakness does not justify the number of researchers, time and money invested in the development of more accurate forecasting tools. Moreover, despite the fact that LUTI models are designed to be used for policy evaluation, a framework for policy assessment utilizing the benefits of microsimulation (on which they are based) is still absent.

The effects of the emerging sustainable transportation systems –such as carsharing– on urban development, but also the unexpected economic situations –such as recessions– have not been investigated in depth. Furthermore, not all the directions of using advanced behavioral modeling techniques that integrate discrete choice models and latent variables, have been investigated in policy analysis, yet.

The objective of this doctoral research is to investigate, develop and propose data collection, modeling and policy analysis methodologies for Integrated Land-Use and Transport models. The proposed approaches aim in cost reduction and forecasting accuracy increase, taking advantage of the strengths of microsimulation in three dimensions: agents, space and time.

1.4 Doctoral research contributions

This research makes several contributions to the state–of–the–art of Integrated Land-Use and Transport models:

- The prospect of using crowdsourced data in LUTI model development is explored. On-line real estate data are collected and used for estimation of real estate price models. Land-use/cover data extracted from satellite images by remote sensing techniques, are used for the development of land-use/cover change model. Aiming to reduce the dependence on expensive demographic databases, a graph-theoretic approach for associations generation of synthetic population is proposed.

- A framework that integrates spatial econometric models in LUTI is proposed. In particular:

  - Real estate prices are modeled using spatial econometric techniques. These models result in better fit, comparing to the traditional Ordinary Least Squares (OLS) framework, and effectively remove the spatial autocorrelation. Aiming to measure the impact of transportation infrastructure and policies on real estate purchase prices and rents, they are applied in two case studies in Greece (Athens and Thessaloniki).
  - Land-use/cover change is modeled using discrete spatial probit, resulting in better model fit than the generalized linear regression. The impact of large-scale transportation infrastructure developments on land-use change in Athens, is examined.
  - A demand-oriented approach for spatial allocation of transportation facilities (stations of electric vehicles, carsharing etc.) based on spatial econometric models and multi-criteria analysis (MCA), is proposed.

- A methodology for qualitative and quantitative policy evaluation indicators analysis, based on the strengths of microsimulation in three dimensions (agents, space and time) is developed. Agent-based indicators for accessibility (e.g. utility based logsum), inequality (e.g. Theil index), economy/investments (e.g. price-to-rent ratio) and social quality (e.g. age mix) are investigated. The proposed methodology is based on spatial distributions instead of single aggregate measurements.

- The impact of sustainable public transport policies on urban development, using hybrid choice and latent variable models is explored.

- The effects of the financial crisis (derivative of various policies) on transportation infrastructure/policies and real estate prices are investigated.

Figure 1.2 show a generalized methodological framework of LUTI model, with the contributions of this research in *italics*. Figure 1.3 presents a more detailed methodological framework with the contributions in stripped boxes.

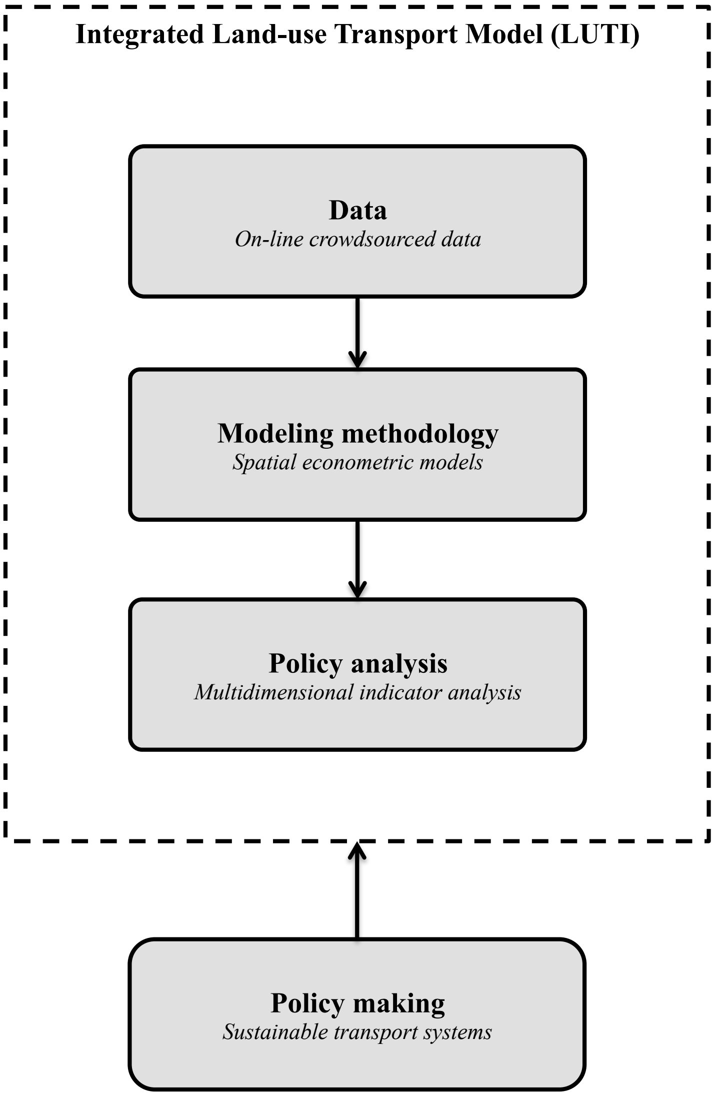


Figure 1.2: Generalized methodological framework of LUTI models

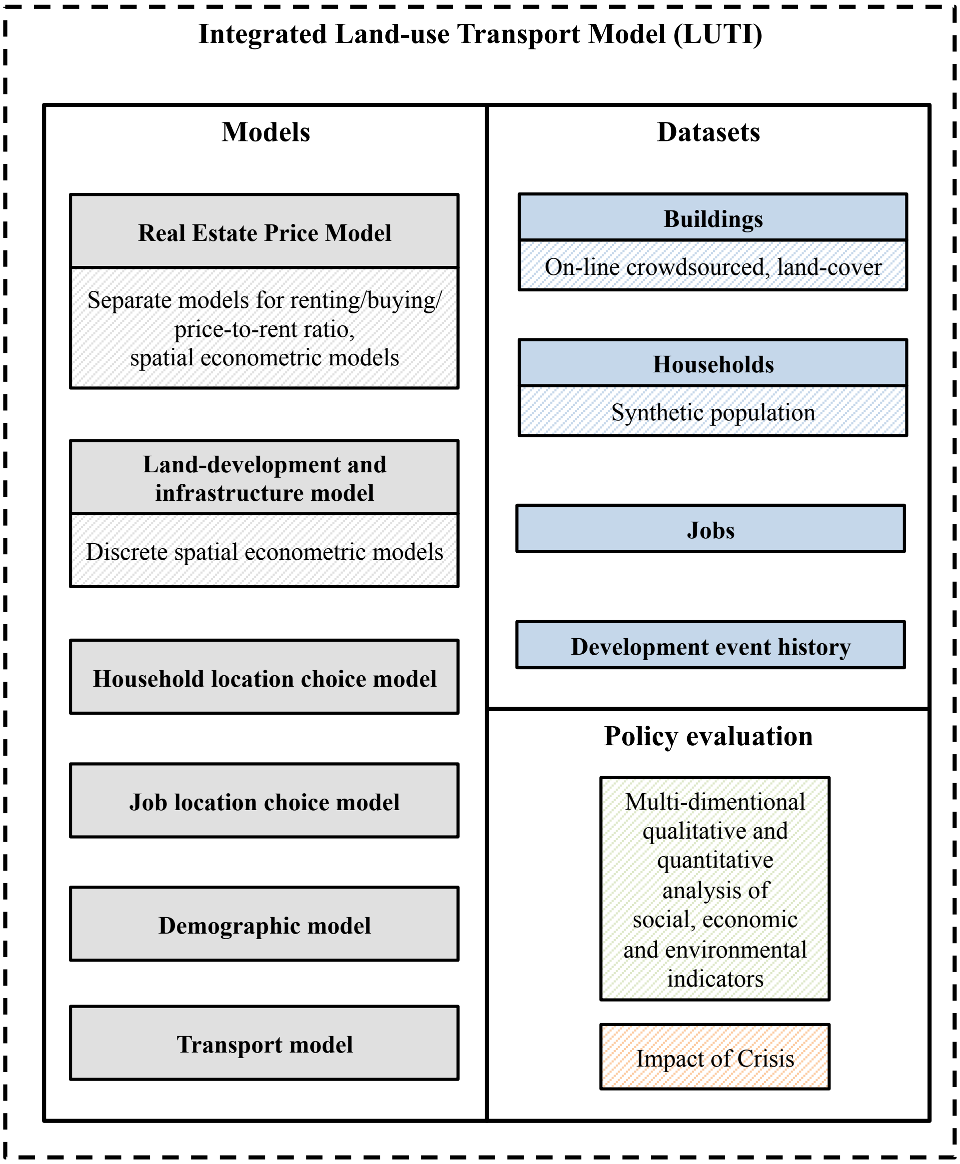


Figure 1.3: Detailed methodological framework of LUTI model - doctoral research contributions

1.5 Dissertation outline

The reminder of this dissertation is organised as follows. Chapter 2 presents the literature review of: 1) LUTI models; 2) spatial econometric models; 3) the impact of transport policies on real estate prices; 4) indicators currently being used for policy evaluation; 5) emerging public transit policies, such as carsharing and bikesharing, and 6) the behavioral modeling methodologies that are being used in this doctoral research. In Chapter 3, a number of spatial econometric model applications is presented, emphasising on the impact of transportation policies and the crisis on real estate prices. Chapter 4 presents a methodologies that can be used in different steps of LUTI model development, from data construction to qualitative and quantitative policy analysis and evaluation. In Chapter 5, the behaviour of public transit users is modeled, and the effects of the crisis on their perception about transit quality is measured. Chapter 6 presents opportunistic data sources that can be used for LUTI model development. Finally, Chapter 7 outlines the conclusions of this doctoral research.

2 Literature review

2.1 Research objective and chapter structure

The objective of this chapter is to make a review of the literature that has been studied in the context of this dissertation. Section 2.2 summarizes a number of LUTI models around the globe and synthetic population methodologies. Moreover, it presents UrbanSim & MATSim, the LUTI model that has been used for the purpose of this research, and the SustainCity project. Section 2.3 describes different types of spatial econometric models. Section 2.4 reviews a number of studies that measure the impact of transportation infrastructure and policies, such as railways, airports and highways, on real estate prices. Section 2.5 describes the indicators that are currently being used for policy evaluation, such as sustainability, economic, inequality and accessibility. Section 2.6 presents emerging public transit technologies, namely carsharing and bikesharing, and finally, Section 2.7 describes the statistical modeling methodologies that have been used throughout this dissertation.

2.2 Integrated Land-use and Transport models

2.2.1 Current situation

The first known LUTI model was developed by Lowry (1964), for Pittsburgh. Her work was fundamental for the land-use forecasting models that followed. Fleisher (1965) and Goldner (1971) reviewed the Lowry-based models in US, emphasiszing the potential of their use, while Wegener (1982), Macket (1983), Echenique (1985) and Putman (1983) were among the first who developed LUTI models for European cities.

The technical difficulties that arise for their implementation, such as restricted data availability and need for more powerful computers to handle the computationally demanding simulations, were clear from the beginning (Wegener, 2011). The improvement of computers' processing power and the introduction of new methodologies for data collection, accelerated and multiplied the applications of LUTI models, contributing to the transition to a new era.

Until now, more than twenty LUTI models (Wegener, 2004) have been developed and applied to areas of different size –from small towns to whole countries– either for research purposes or for policy planning and evaluation. These models differ as to their content, structure, theoretical background, methods of modelling, amount of required data, calibration and validation approaches, as well as their functionality and applicability. The literature on LUTI models is continuously growing. Several articles provide a thorough review (e.g. Wilson, 1998).

Martinez (1995) suggested a framework to transform travel demand into accessibility indicators, which measure the maximum surplus that the owners of land and houses can benefit from transportation activity in origins and destinations. The methodology that Martinez suggested was based on static measurements and it was later implemented within the LUTI model MUSSA for Santiago (Chile) (Martinez, 1996). MUSSA forecasts the location choice of households and firms using bid-choice theory. The performance of the model was improved in MUSSA II (Martinez and Donoso, 2010).

Kitamura et al. (1996) introduced the integrated, SAMS model framework that incorporates air-quality dynamic forecasting. Anas and Liu (2007) synthesized already developed models and created the integrated RELU-TRANS model for Chicago MSA. Farooq and Miller (2012) developed a disequilibrium-based framework for micro-simulation modeling of the space markets in the Greater Toronto area, which was applied to ILUTE (Toronto's LUTI model).

Hunt et al. (2005) review five LUTI models (ITLUP, MEPLAN, MUSSA, NYMTC-LUM and UrbanSim) and outline their capabilities in evaluating policies. Curtis (2011) outlines the inefficiencies of the current LUTI models to measure public transport accessibility and he suggests a framework applied in SNAMUTS for their measurement. The suggested accessibility indicators are: 1) degree of centrality, 2) average minimum cumulative impediment, 3) change of efficiency after the service's implementation, 4) number of residents and jobs reached with a travel–time of 30 minutes, 5) geographical distribution of transport services, 6) comparison of speed between the public and private transport and 7) network connectivity. Jonsson et al. (2011) compares the simulated scenario output generated by two models of different disaggregation level in Sweden: Lu-TRANS, and SAMPERS. They suggest that improving the computer power for model simulations is not always an obstacle, since analysis should be performed even with lower computing capabilities.

Waddell (2011) analyzes the main challenges that Metropolitan Planing Organizations (MPO) face when implementing LUTI models. These are divided in two categories, those that refer to integrated modelling, and those referring to integrated planning. In the first category belong: 1) conflicting institutions, 2) conflicting values, 3) conflicting epistemologies, 4) conflicting policies, while the second comprises: 1) transparency, 2) behavioural validity, 3) empirical validity, 4) ease of use, 5) computational performance, 6) flexibility, 7) data availability and quality and 8) uncertainty. Waddell suggests that UrbanSim is capable of overcoming these obstacles and should further be used.

Table 2.1 provides a summary of the main characteristics of the real estate price models used in six widely applied LUTI models.

Table 2.1: Real Estate Models per LUTI

LUTI Model	**Source**	**Real Estate Model**
UrbanSim	Waddell (2002)	Semi-log linear regression. Is applied last, at any time-step of the simulation (one year). The response variable is the logarithm of price, and the explanatory variables include a bundle of housing structural, neighborhood and accessibility characteristics. However, the user can modify it in the limits given by the OPUS computational framework.
MEPLAN and TRANUS	Echenique (1985)	Prices are adjusted in zones until the market 'clears' (available space of households is consumed)
ITLUP	Putman (1983)	Static equilibrium to model the demand. No endogenous prices
MUSSA	Martinez and Donoso (2010)	Bid-choice model. Prices are determined by the maximum 'bid' that people are willing to pay (WP) to maximize their surplus, and renters accept to charge to maximize their profit
ILUTE	Salvini and Miller (2005)	Individual bids
NYMTC-LUM	Anas (1998)	Equilibrium / are determined endogenously

2.2.2 Synthetic population

There are two main methods for population synthesis: *reweighting* and *imputation*, according to Williamson (2013); Lenormand and Deffuant (2012). Reweighting is the process of cloning and replicating agents from a small sample to achieve the desired population size. Imputation is the process of generating a completely new population with the same chacateristics as the sample population. Other, more comprehensive literature reviews can be found in Harland et al. (2012); Hermes and Poulsen (2012); Müller and Axhausen (2010); Ma (2011); Pritchard (2008).

Reweighting

Reweighting methods can be further subdivided into *fitting and generation*, and *combinatorial optimization*. These approaches require a reference sample of disaggregate census data as well as aggregate data for the entire population. In the fitting and generation approach, the synthetic population is created by drawing from a weighting of the reference sample (Beckman et al., 1996b). Some variants exist, such as a deterministic reweighting using generalized regression (Harding et al., 2004; Tanton et al., 2011).

The combinatorial optimization approach involves selecting a population from the reference sample and evaluating the fit to aggregate statistics through the use of tables (one per constraint). The population members are then switched out until an optimal solution is found (Voas and Williamson, 2000; Edwards and Clarke, 2009). An advantage of combinatorial optimization is relatively little variation between runs.

A disadvantage of reweighting methods, as Voas and Williamson (2000) note, is that the resulting synthetic population only produces a good fit to variables which are used as constraints. This implies that this method is oriented towards generating a synthetic population for a specific purpose. Generating a "general purpose" population is possible, but requires a large number of constraining variables. Combinatorial optimization relies heavily on random number generation, so a sufficiently robust generator is needed since a pseudorandom generator may not be adequate Voas and Williamson (1998). Additionally, the sample population must be sufficiently large and diverse, otherwise reweighting will not be successful (Huang and Williamson, 2001). More comprehensive reviews of reweighting methods can be found in Huang and Williamson (2001) and Rahman et al. (2010).

Imputation

In general, imputation methods create a population from basic known attributes (such as age and gender) and use probability distributions conditioned to known attributes to sample all other attributes. One such method is the Markov Chain Monte Carlo method, which is an extension of Iterative Proportional Fitting (IPF) proposed by Deming and

Stephan (1940). This approach first fits a *contingency table* to marginals (which can be calculated from more aggregate demographic data). Farooq et al. (2013a) use a Markov Chain Monte Carlo (MCMC) method to synthesize a population of agents and check the accuracy of this synthetic population by comparing it to full Swiss census data. The procedure in Schafer (2010) assumes the cell values to be random variables with constrained Dirichlet priors. The MCMC process is then used to fit cell values from the posterior Dirichlet. The second part of an IPF procedure uses the contingency table and its cell weights to create a synthetic population. Beckman et al. (1996b) use Inverse Transform Sampling in a Monte Carlo simulation to do this.

Validation

Many authors have commented that it is difficult to validate model results without having access to spatially-referenced microdata (Ballas et al., 1999). One option is to aggregate the synthetic data and compare it with the smallest available level of real data (Tanton et al., 2011). A common measure of fit is the normal Z score, where values over the 5% critical value are judged to be "non-fitting" (Voas and Williamson, 2001; Hynes et al., 2009; Rahman et al., 2013). Total Absolute Error (TAE), Standardized Absolute Error (SAE) and Z score are in common use as measures of fit (Ballas et al., 1999; Huang and Williamson, 2001). The best validation methods incorporate both internal (compare input to output) and external (compare to a similar, unused dataset) validation (Oketch and Carrick, 2005).

Other validation methods are regression analysis and the use of R^2 values (Ballas, 2005; Edwards and Clarke, 2009) as well as standard error around identity (SEI) (Ballas et al., 2007; Tanton et al., 2011; Tanton and Edwards, 2013).

2.2.3 UrbanSim

UrbanSim was initiated by the University of Washington in the late 1990s, and since then it has been further developed by the University of California at Berkeley, and other major institutions in US and Europe. It is provided under the GNU public license and can be downloaded from www.urbansim.org. It currently forms the state-of-the-art of LUTI models.

UrbanSim is different from other models as it adopts an approach of dynamic disequilibrium, it predicts the characteristics of the integrated transport and land use models in different time scales and because it leverages extremely disaggregated spatial information (Waddell, 2002). It can be applied in three different levels of spatial disaggregation: zones, gridcells and parcels. One of its characteristics that is regarded as both positive and negative, is the need for extremely disaggregate data. The model requires detailed data describing, among others, individual households, buildings and jobs within the area of interest. UrbanSim integrates the following types of models: 1) *Transition models*, that describe the evolution of the agents demand (households and jobs) for each simulation period; 2) *Development models*, that describe the generation of new real estate supply (residential and non-residential buildings); 3) *Relocation models*, where the agents' decision of moving from their current location is simulated; 4) *Location Choice Models*, describing the spatial allocation of the new or relocating agents across the different, already existing or new alternative locations; 5) *Price models*, that compute the prices of real estate (Waddell, 2002; Waddell et al., 2007). The models interact in each simulated period, generating a new state of the system that is used as a starting point for the simulation in the next period. More specifically, UrbanSim takes as input household, employment, building and real estate data, geo–referenced in zones, parcels or gridcells, and uses the following set of sub-models: 1) Economic Transition Model; 2) Demographic Transition Model; 3) Employment Relocation Model; 4) Household Relocation Model; 5) Employment Location Choice Model; 6) Household Location Choice Model; 7) Real Estate Development Model; 8) Land Price Model; while others have recently been added, e.g. 9) marriage model and 10) death model (Waddell, 2002). UrbanSim is probably the most widely applied LUTI model. Known case studies include Honolulu (Hawaii), Springfield (Oregon), Houston (Texas), Salt Lake City (Yutan), Zurich (Switzerland), Seattle (Washington), San Francisco (California), Paris (France) and Brussels (Belgium) (www.urbansim.org).

All location choice models in UrbanSim are of the *multinomial logit* type, where decision makers (agents) choose from a sample of available alternatives (buildings or locations) selecting the one that provides maximum utility given its attributes and price.

Multinomial logit models are applicable where there is a predefined master choice set that includes all choices available to the population (Ben-Akiva and Lerman, 1985). It is not necessary for every element of the master set to be available to all individuals; the availability of choices can be determined by other variables such as location and socioeconomic characteristics. Multinomial logit has seen wide applicability as a mode choice model, but we find that it can also be adapted to this case, a choice between different kinds of households. The model formulation is:

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j \in C_n} e^{V_{jn}}} \tag{2.1}$$

where P is the probability of choosing a particular alternative, V is the utility of each alternative, and C_n is the choice set (Luce, 1959; Marschak, 1960).

Market clearing is treated with a *first come first served* approach (Waddell, 2010) meaning that, when two agents select the same location, conflict is solved by randomly selecting one of them.

Real estate price models are also of high relevance because they describe the market value of the traded goods. In UrbanSim, real estate prices are modelled using a hedonic regression of property value per surface unit of the building, and its environment and market-level vacancy rates (Waddell and Ulfarsson, 2003), as follow:

$$\ln\left(p_{vit}\right) = \alpha + \delta\left(\frac{Q_v^s - Q_{vt}^c}{Q_v^s}\right) + \beta X_{vit} \tag{2.2}$$

Where $\ln(p_{vit})$ is the natural logarithm of price of land per surface unit for development type v at location i and time t, Q_{vt}^c is the current vacancy rate at time t, Q_v^s i is the long-term structural vacancy rate, X_{vit} is a vector of building and location attributes, and α, δ and β are estimated parameters.

In each period, new households and firms are generated by the *transition models*. Simultaneously, new supply is generated by the *development models* and distributed within locations in the city. Relocating and new agents enter the market and choose their location following the distribution defined by the location *choice models*. At the end of each period, prices are computed and all location and building attributes are updated to enter as the main input to the next period simulation.

One of the main characteristics of UrbanSim is the independent estimation process for each of the involved submodels. This is a practical advantage that simplifies the implementation of the model but also implies strong assumptions about the behavior of agents and the interdependence of the decision processes that takes place in the city. The modular structure and open source nature of the code allows to customize UrbanSim for several different circumstances and conditions, although this may require advance knowledge of the software.

2.2.4 MATSim

MATSim is an agent-based framework used for travel demand modelling. It is based on the individual person's travel plans, while it integrates information about his travel patterns, such as: mode used, time of departure, time of duration per trip and time of return. MATSim selects the optimal routes applying iterative optimisation process, an algorithm suggested by Charypar and Nagel (2005), which considers random initial departure time and duration for each activity. It is provided under the GNU public license and can be downloaded from www.matim.org.

2.2.5 SustainCity project

The goal of this European Union's project (funded by the Seventh Framework Programme [FP7]) is to address the modelling and computational issues of integrating modern mobility simulations with the latest micro-simulation land use models. The project intended to advance the state-of-the-art in the field of the microsimulation of prospective integrated models of Land-Use and Transport (LUTI). On the modelling side, the main challenges were to integrate a demographic evolution module, to add an environmental module, to improve the overall consistency and, last but not least, to deal with the multi-scale aspects of the problem: several time horizons and spatial resolutions were involved (www.sustaincity.org, Sustaincity, 2009).

The purpose of this project was the integration of UrbanSim with MATSim and METROPOLIS (De Palma et al., 1997), the creation of a new Sustainability sub-model, the implementation of a policy evaluation tool based of the Social Welfare Function (De Palma et al., 2010) and generally, its development in order to achieve a better fit at the local characteristics of European capitals. The interoperability between these systems creates a wide and flexible platform that can be used by the European governments for urban, sustainable planning and development of the European metro areas.

Before SustainCity, significant knowledge for the European conditions was gained from two case studies in Lausanne (Patterson and Hurtubia, 2008) and Brussels (Patterson and Bierlaire, 2010; Patterson et al., 2010), where the model was applied with aggregate data, and Zurich (Löchl et al., 2007).

2.3 Spatial econometric models

2.3.1 Introduction

A property can be considered as a bundle of attributes –both physical and spatial– grouped in four categories: 1) characteristics of the house (number of rooms, total surface, floor e.t.c); 2) attributes of the community (demographics, income, percentage of residents with university degree, percentage of foreigners, proportion of green surface per resident e.t.c); 3) accessibility indicators (e.g. distance from Central Business District or the proximity to a metro station) and 4) environmental quality of the region (e.g. noise or air pollution) (Picard and Antoniou, 2011).

Land location choice modelling was introduced by McFadden (1974, 1981) who suggested the Random Utility Maximisation (RUM) theory, while Alonso (1964) and Wheaton (1977) described a bid-theory approach. Lancaster (1966) first said that the utility is generated by the characteristics of the goods, while these characteristics have different values of utility at different persons. Rosen (1974) introduced the hedonic price theory in a two-part approach. In the first part a hedonic model for price is estimated and the marginal implicit price is calculated. Then, the marginal prices are used jointly with the socio-economic characteristics of the consumers to estimate the coefficients of the consumers' behaviour. That approach had been criticized by Brown and Rosen (1982), Epple and Sieg (1999) and Bartik (1987), because it assumes that the group of individuals is homogeneous. Bajari and Benkard (2001) considered different utility parameters for each individual and they generalised Rosen's model making it easier to be applied. Bajari and Kahn (2005) estimated housing demand to model the racial segregation in American cities, using a hedonic approach. They tried to explore why black households prefer to live in the city centre, while white households prefer the suburbs. They concluded that income, educational level and marriage rates are all significant variables that determine the choice to live in high human capital communities and bigger houses, usually situated in the suburbs.

Hedonic price modelling is based on the Ordinary Least Squares (OLS). One way to include spatial information, is to use the distance from the CBD or from other neighbouring characteristics –such as metro stations, lakes etc.– as variable indicators into the model. However, spatial information is not included properly, generating deficiencies because of the spatial auto-correlation and the spatial heterogeneity that remain (Anselin, 1988). According to the findings of Bitter et al. (2007), these two effects are related and should be considered jointly.

2.3.2 Types of spatial econometric models

Spatial econometrics are defined as *the collection of techniques that deal with peculiarities caused by space in the statistical analysis of regional science models* (Anselin, 2010). The correlation and covariance of the variables in different locations, is determined by the position of the data points in space. Testing the fit of spatial econometric models is a complicated problem. While the patterns and clusters can be identified, the process that leads to these patters is not always clear. As a result, there are different tests that allow one to check whether the problem of spatial autocorrelation or heterogeneity is addressed (Anselin, 2010).

Spatial econometric models need a $n \times n$ matrix of the neighbour location data points, so as to structure the covariances. These matrices can be based on *distance*, *contiguity* or *nearest neighbours*. In the first case, the matrix is formed using the data points of a given threshold of euclidean distance around each observation. In contiguity-based cases, the surrounding polygon centroids (of municipalities, states etc.) which have a common border are used, while –in the last case– a weight matrix is created by a number of k-nearest neighbour points. The spatially lagged variables created by these matrices, can be applied to the dependent variable, the independent variables, or the error term.

The size of the weighted matrix created (i.e. the number of the neighbor data points that is considered for the creation of the lags) depends on the characteristics of each case and is specified by the researcher. Lag variables can be used for the response, the observatory variables, the error term or to more than one of the aforementioned. As a result, the terms spatial lagged models, spatial cross-regressive models, spatial error models and spatial mixed/Durbin models, are used respectively for these categories. SDM is preferred when omitted variables exist, because it is can potentially reveal them. Their characteristics are described by Kissling and Gudrun (2008) but also by LeSage and Pace (2009). Anselin (2010) published an extensive literature review on the most important spatial econometric model applications of the last 30 years.

The spatial error model [defined as SEM (LeSage and Pace, 2009) or SAR$_{err}$ (Kissling and Gudrun, 2008)] assumes that spatial dependence is applied to the error term:

$$Y = X\beta + \lambda W u + \varepsilon \tag{2.3}$$

where Y is a vector, β is the $k \times 1$ vector of coefficients, X is a $n \times k$ matrix of the independent variables, λ is the spatial autoregression coefficient, u is the spatial error term (vector), W is the weights matrix and ε is the spatially independent error term.

The spatial autoregressive model [SAR (LeSage and Pace, 2009) or SAR$_{lag}$ (Kissling and Gudrun, 2008)] the autoregressive process is applied to the dependent variable Y transforming it to a *lagged* variable:

$$Y = \rho W Y + X\beta + \varepsilon \tag{2.4}$$

where Y is a vector, β is the $k \times 1$ vector of coefficients, X is a $n \times k$ matrix of the independent variables, W is the weights matrix, ε is the spatially independent error term, ρ is the spatial lag (scalar).

In spatial Durbin model [SDM (LeSage and Pace, 2009) or SAR$_{mix}$ (Kissling and Gudrun, 2008)], the spatial autoregressive process is applied both at the response and explanatory variables. This model can potentially remove the bias caused by the omitted variables.

$$Y = \rho W Y + X\beta + W X\gamma + \varepsilon \tag{2.5}$$

where Y is a vector, β is the $k \times 1$ vector of coefficients, X is a $n \times k$ matrix of the independent variables, W is the weights matrix, ε is the spatially independent error term, ρ is the spatial lag applied on Y and γ is the autoregression coefficient applied on the matrix X.

The ρ and λ values are computed by numerical optimization. The 2D surface of the likelihood function usually forms a "banana trench" (from low/high lambda to high/low rho). In order to deal with the different values, the spatial autocorrelation model [SAC (LeSage and Pace, 2009)] uses from four to nine starting points to compute the optimum result. For that purpose, it uses two weight matrices: one for the spatial lag (dependent variable) and another for the spatial error (LeSage and Pace, 2009):

$$Y = \rho W_1 Y + X\beta + u, \text{where } u = \lambda W_2 u + \varepsilon \tag{2.6}$$

2.3.3 Geographically Weighted Regression

Another spatial model used by econometricians is the Geographically Weighted Regression (GWR) introduced by Fotheringham et al. (1998) and was first applied for estimating the land prices in the London Metropolitan Area. Unlike the *global* approach of the linear regression –where an average estimation of the statistical parameters is applied equally at every observation around an area– GWR as a *local* model does not provide single statistical estimates, but spatial-aware distributions of these parameters.

GWR (Fotheringham et al., 1998; Brunsdon et al., 1998) uses the traditional regression framework with local spatial relationships:

$$y_i = \beta_0(u_i, v_i) + \sum_{i=1}^{k} \beta_k(u_i, v_i) x_i k + \varepsilon_i \tag{2.7}$$

Where: $(u_i,\ v_i)$ are the coordinates of the point i in space, and $\beta_k(u_i, v_i)$ is the known continuous function $\beta_k(u, v)$ at point i. Unlike the traditional regression, this framework considers the spatial variability of the surface.

Similar to spatial econometric models, GWR adjusts weights to a number of observations around the regression points, fitting either Gaussian or bisquare spatial kernels continuous with the distance. The number of data points that will be used for the formation of each weight matrix is determined by a bandwidth, which can be either fixed (e.g. the radius of a circle) or adaptive (e.g. percentage of total data points).

A problem that usually arises with fixed kernels is that the regression points are not always equally distributed around the surface, which means that where the data are sparse, the model uses a limited number of data points for calibration, resulting in large standard errors and "undersmoothed" surfaces (Fotheringham et al., 1998). In these cases, the method of adaptive kernels is preferred. The model differentiates the bandwidth until it includes the same number of neighbour data points at each regression point. The calculation of the fixed or adaptive kernel value can be made with the following two methodologies: the cross-validation (CV) score and the Akaike Information Criteria (AICc) optimization (minimisation). However, it can also be given manually.

AICc, the corrected AIC, is an indicator of the information distance between fitted model and the real (lower imply better fit), while it is also known as Kullback-Leibler Information Distance (Charlton and Fotheringham, 2009).

2.3.4 Discrete choice spatial econometrics

Pinkse and Slade (1998) were of the first who suggested the integration of the spatial linear regression framework with discrete choice models. Knowles (2006) reviewed the transport policies of the last 200 years focusing on their impact in shaping the geography, using discrete spatial models. Kakamu and Wago (2005) developed a Bayesian Spatial Panel Probit Model and applied it to model the Business Cycle in Japan. Kapoor et al. (2007) developed spatial panel models where they used generalized moments estimator (GMM) for the autoregressive spatial parameter. Paez (2006) introduced a spatial binary probit model with geographical weights that combines the theory of the heteroscedastic probit suggested by McMillen (1992), and GWR. The model was applied to measure the impact of transportation infrastructure on the changes of land-use. Klier and McMillen (2008) suggested a spatial logit model where the location of an auto supplier depends on the probability of being located in a neighbor country. When dealing with the problem of urban development, a first approach is to develop binary response models for panel data just by separating the "developed" from "undeveloped" land use areas (Frazier and Kockelman, 2005). Furthermore, Markov Chain microsimulation models for large-scale land use have been used to capture market dynamics and further land use forecasting extracting promising results (Zhou and Kockelman, 2011).

Spatial and temporal correlations are better expressed and interpreted using panel data from multiple time periods. Econometric models taking into account spatial parameters, such as spatial dependence, spatial autocorrelation etc. (Anselin, 1988; Elhorst, 2003) have been developed. Mixed Logit models have been estimated using panel data (Wang and Kockelman, 2006), recognizing distance-dependent correlations using spatial autoregressive models (SAR) and spatial moving average models (SMA) (Frazier and Kockelman, 2005; Benser, 2002; Miyamoto et al., 2004). Finally, the cellular automata (CA)-based SLEUTH model (Candau et al., 2002) has been designed to work with (satellite imagery) data from at least four periods. Discrete choice spatial econometrics have been used in modeling the land-use changes. Lewis and Plantinga (2007) simulated the results of different policy scenarios using spatial econometric models, by modeling the land-uses after a subsidized afforestation in the landscape. Sener et al. (2011) introduced the Generalized Spatially Correlated Logit (GSCL) model, which uses the framework of the Spatially Correlated Logit (SCL) (Bhat and Guo, 2004), taking into account the spatial term. The land-use changes in Austin (Texas) have been modeled by Wang and Kockelman (2008, 2009b) who developed the dynamic spatial ordered probit (DSOP), Wang et al. (2011), who used a geographically weighted regression for discrete response, and Wang et al. (2012), who introduced a dynamic spatial multinomial probit (DSMNP) model. The first has also been applied to measure the Ozon concentration in Austin (Wang and Kockelman, 2009a).

Spatial Probit Model

The spatial probit model is based on the following formulation:

$$y = \rho W y + X\beta + \varepsilon, \ \varepsilon \sim N(0, \sigma_\varepsilon^2 I_n) \tag{2.8}$$

or

$$y = (I - \rho W)^{-1}(X\beta + \varepsilon) \tag{2.9}$$

Where y is a $n \times 1$ vector with values 0 or 1, indicating the presence or the absence of a characteristic, X is a $k \times n$ vector of parameters, and ε is the error term.

The spatial probit model assumes that the difference between the utility to choose and not to choose ($y_i^* = U_{1i} - U_{0i}$, where $i = 1, ..., n$) follows a normal distribution.

$$y = 1, \ \text{if} y_i^* > 0 \tag{2.10}$$

$$y = 0, \ \text{if} y_i^* < 0 \tag{2.11}$$

The estimation of the on the Bayesian probit model is based on Monte-Carlo Markov-Chain (MCMC) sampling. In the Bayesian estimation, the unobserved latent utility that determined the positive (1) or negative (0) decision, is replaced by estimated parameters (y^*). A sample is generated for the structure of the y^*. The number of samples (draws) should be large enough so as convergence has been achieved. The vector is generated with the m-steps Gibbs sampler process.

The covariance matrix of the maximum likelihood estimation is computed by the term:

$$V(\hat{\theta}) = (\sum dlnL_i/d\hat{\theta})(\sum dlnL_i/d\hat{\theta}') \tag{2.12}$$

The GMM estimator of the spatial probit model selects β and ρ in order to minimize the, where Z in a matrix of instruments. In our case, the instrument is the WX term, where W is the spatial weight matrix and X is the list of the explanatory variables. The estimation begins with the initial estimates computed by a linearizes version of spatial probit, as suggested by Klier and McMillen (2008).

2.3.5 Spatial filtering

G index

A spatial statistic used for the spatial exploration of the data heterogeneity and dependence, aiming to identify similar spatial patterns, is the G statistic, introduced by Getis and Ord (1992); it indicates the association between any point i and the neighbors.

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j} \tag{2.13}$$

where w_{ij} is a symmetric spatial weight matrix with value 1 when the points i and j are within the defined distance (d) or contiguity. When the combination of (i,i) points is zero and x_i is not included in the the sum, the G^* is computed:

$$G_i^*(d) = \frac{\sum_j w_d x_j - W_i^* \bar{x}}{s\left\{\frac{[nS_{1i}^* - W_i^{*2}]}{(n-1)}\right\}^2} \tag{2.14}$$

where, $w_{ii} \neq 0$, $\bar{x}$ is the mean value of x and s^2 the variance, $W_i^* = W_i + w_{ii}$, $S_1 = \sum_j w_{ij}^2$ for $j \neq i$ and $S_{1i}^* = \sum_j w_{ij}^2$

Ravulaparthy et al. (2012) applied G^* cluster analysis to identify common patterns of house prices in Los Angeles, California. In this research, we attempt to identify clusters of dwellings with similar purchase or rent prices.

Eigenvectors' approach

Spatial filtering has been introduced by Griffith (2008) and Dray et al. (2006) aiming to account spatial autocorrelation and better understand multicolinearity. As stated in Wang et al. (2013), the advantage of using this method instead of spatial autoregressive techniques, is that it identifies clustering patterns, useful for land-use forecasting and spatial analysis. Despite the advantages of this method, its applications in transport and land-use modeling are still limited.

Spatial filtering requires the definition of the eigenvectors that represent the spatial structure of the dataset. The computation of the eigenvectors for large datasets is a time-consuming process, which forms the main disadvantage of the technique. The computed eigenvectors are then used as explanatory variables in the model, capturing the spatial correlation of the data points and leading to better fit and lower spatial-autocorrelation statistics.

Getis and Griffith (2002) compared the two filtering methodologies that they have developed, concluding that on the one hand Getis approach is simpler, less computationally demanding and more understandable, on the other, Griffith's can identify the presence of multi-collinearity, and can be applied for a spatial regression instead of the spatial autoregressive models. They applied their methods on US per-capita economic data and found that both techniques resulted in similar goodness of fit tests.

Griffith (2008) applied spatial filtering and compared the results with those of GWR, highlighting the benefits of using the technique. Tiefelsdorf and Griffith (2007) embedded the eigenvectors resulted from spatial filtering into a semi parametric statistical framework as explanatory variables, resulting in models with better fit that the traditional OLS estimation.

Spatial filtering has been applied in the research filelds of the ecology, optical imaging and economic analysis (for examples of applications see Wang et al., 2013), while the literature on transportation applications is limited. Moniruzzaman and Páez (2012) explore the benefits of using eigenvector spatial filtering in controlling the spatial autocorrelation, by modeling the transit shares in the city of Hamilton, Canada.

The research of Wang et al. (2013) forms a first attempt to apply spatial filtering in transportation and land-use research. The authors make an extensive introduction of spatial filtering literature and then apply it to model the continuous land values and binary land-uses. Comparing the results of spatial filtering OLS and GLM with spatial autoregressive (SAR) models, they prove the potential of the first to results in better goodness of fit tests and effectively remove spatial autocorrelation. They conclude that the SF approach results in more reliable marginal effects and that its use in transportation and land-use research should be further considered.

Tiefelsdorf and Boots (1995) and De Jong et al. (1984) expressed the extreme values of Moran's I as a function of eigenvectors:

$$\text{Moran's I} = \text{MI} = \frac{n}{1'W1} \cdot eigenvalue(\Omega) \tag{2.15}$$

where n is the sample size, Ω are the normalized eigenvalues. The eigenvectors are generated consecutively, starting from the one with the highest spatial autocorrelation (MI).

According to Wang et al. (2013), in land-use modeling, the identification of meaningful eigenvectors could reveal relationships that can be controlled between the land-units. The question is, which eigenvectors should be included in the regression equation? Tiefelsdorf and Griffith (2007) suggested the minimization of Z-score, a computationally demanding optimization algorithm, due to the requirement of the MI's computation at each step.

$$minz[I(\hat{\epsilon}^*)] \equiv \left[\frac{y^T M_{(X|E)} W M_{(X|E)} y}{y^T M_{(X|E)} y} - E(I(\hat{\epsilon}^*)) \right] / [var(I(\hat{\epsilon^*}))]^{1/2} \tag{2.16}$$

where $E(I(\hat{\epsilon}^*))$ is the expected MI of the residuals,

In the current study, the low goodness to fit statistics of the housing demand model resulted by the estimation of the SEM made us investigate alternative modeling solutions. For that reason, we explore the eigenvalues spatial filtering approach to model the demand.

2.3.6 Model estimation and tests

Two methodologies are generally used for the estimation of spatial econometric models: maximum likelihood, and general method of moments (GMM) (Kelejian and Prucha, 1999). When dealing with complicated data, the instrumental variables method is preferred in order to facilitate the estimation procedure.

The increase in the availability of geographically referenced (spatial) data has enforced the debate for more accurate measurements and predictions of the spatially dependent attributes. The development of research tools equivalent for spatial econometric analysis – such as R (R Development Core Team, 2014) and GeoDa (Anselin et al., 2006) – has contributed to the increase of the number of applications in different research fields.

Spatial econometric models are valid only for the time when the data were collected. Elhorst (2003) and Kapoor et al. (2007) suggested the spatial panel models that include space and time lags for both dependent and independent variables.

A number of tests to check the presence of spatial non-stationarity detected by the GWR, have been developed. The spatial variability of the independent variables is checked by the Leung tests (Leung et al., 2000). GWR is advantageous in terms of solving the issue of spatial non-stationarity; however, the estimation and calibration of this technique is complex, which usually discourages researchers from using it. The results of GWR are usually compared with OLS using the ANOVA tests but also Moran's I measure (Fotheringham et al., 1998).

Two tests are generally applied to measure the existence of spatial autocorrelation from the residuals of spatial econometric model specification: Moran's I (Anselin, 1995), and the more recently developed, APLE (Li et al., 2007).

The Moran's I is computed using the following formula:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \sum_i \sum_j w_{ij} z_i z_j / \sum_i z_i^2 \tag{2.17}$$

where z_i and z_j are the deviations from the mean, w are the weights, n is the total number of points, for i observation and j neighboring points

The APLE statistic is computed using the following formula:

$$APLE = \frac{Z'[(W + W')/2]Z}{Z'(W'W + \lambda'\lambda I/n)Z'} \tag{2.18}$$

where λ is the vector of eigenvaules of the matrix W as an estimator of ρ

2.3.7 Applications

Spatial econometric models have been widely applied in environmental sciences. There is rich literature with applications for measuring the effects of air and water quality on house prices. Brandy and Irwin (2011), in their extended review report different successful applications. Leggett and Bockstael (2000) proved that water quality affects house prices by applying weight matrices in order to consider the spatial heteroscedasticity. Kim et al. (2003) measured how sulfur dioxide (SO_2) and nitrogen dioxide (NO_2) affect house prices in Seoul (Korea). They found that the effect could be better explained by the spatial lag model. Anselin and Le Gallo (2006) applied interpolation techniques and spatial econometrics to model the effect of air quality on house prices in the South Coast. The spatial lagged model overcomes spatial autocorrelation problems, while the authors also mention that "kriging" would also be appropriate for such estimations. Brasington and Hite (2005) measured the effect of the environmental and school quality on house prices in Ohio and found that these spatial effects should not be omitted. The impact of air quality was also examined by Small and Steimetz (2007), who divided the spatial effects to *pecuniary* and *technological*. Spatial effects of the first group are welfare-neutral, while those on the second affect the welfare. They state that, in the latter case, the effect should be measured by a "spatial multiplier", which can be extracted from spatial autoregression model. Cho et al. (2006) compared local (spatial econometric) and global (linear regression, OLS) models to measure the effect of the green space amenities on housing values in Know Country, Tennessee. Hoshino and Kuriyama (2009) used spatial autocorrelation and kriging to model house prices as a function of distance from park amenities and found the superiority of the "kriging" model. Klaiber and Phaneuf (2010) published an analysis of how open areas affect household location choice and house prices, applying a theory based on social welfare. Bjorn (2009) employed Bayesian Model Averaging (BMA) and GWR to measure the impact of urban forest and vegetable cover on house prices in King Country, Washington.

Lozani-Gracia and Anselin (2011) assessed the predictive performance of price models including spatial characteristics for house valuation, in order to develop an automated technique for that purpose, since the classical regression models usually fail to predict the accurate values. They conclude that examining the prices using submarkets improves the predictive performance.

Examples of GWR applications in transportation research include the following: Chow et al. (2010) modeled the transit ridership of Florida by examining different groups of explanatory variables and values of kernels; Wang et al. (2011) compared spatial models to analyse the variation of land use change and make predictions for the future; Zhao (2004) estimated models to explain the annual daily traffic for different categories of roads, and Löchl and Axhausen (2010) examined how OLS, GWR with SAR models deal with the spatial autocorrelation, in order to establish the real estate price model for Zurich, Switzerland.

2.4 Impact of transport policies on real-estate prices

Dwellings located closer to each other share the same characteristics of the area. These spatially-dependent characteristics, affect the prices of properties located in the same area. Transportation infrastructure, for example a metro station, affects the prices of dwellings within a distance around it, either positively because of the capitalization of the transport costs in the property market, or negatively because of negative externalities such as the increase of criminality, noise pollution etc. These attributes can also be social characteristics with geographical reference. As a result, models that do not consider these local characteristics, suffer from omitted variables which possibly lead to undetectable endogeneity (Brandy and Irwin, 2011).

Real estate price data are among the main components of the integrated land-use and transport models (e.g. in UrbanSim Waddell et al., 2007) and they are used as indicators for the variations of attributes, such as housing and location. These price data can be in the form of transaction prices or rents. As rent and transaction prices are not perfectly correlated, combining two different models would provide additional information to incorporate the decision of the household in one of the two choices.

Spatial econometrics can model the real estate prices in a more efficient way –compared to simpler, commonly used approaches such as ordinary least squares (OLS)– and their integration in the integrated land-use and transport models should be further examined. Moreover, the proximity of dwellings to transportation infrastructure or policies, generates neighbour spatial relation, that should be taken into consideration when estimating real estate models. The impact of transportation infrastructure proximity –such as light rail or metro stations, railways, highways and airports– on dwellings and commercial properties in Europe, Americas and Asia, has been examined by a number of research papers summarized in Table 2.2.

2.4.1 Railways

Kockelman (1990) found that accessibility to transportation infrastructure is a determinant of the house values and rents in the San Francisco Bay Area. Al-Mosaind et al. (1993) concluded that single-family houses located within 500 meters of the light-rail transit stations of Portland (Oregon) are more expensive (c.f. Chen et al., 1997). Cervero and Landis (1993) and Benjamin and Sirmans (1996) found a positive effect of the Washington DC's metro on the house prices (c.f. Damn et al., 1980). Bowes and Ihlanfeldt (2001) analyzed the effects of rail stations on local house prices in Atlanta and found both positive and negative effects. Positive, because they attract retail businesses and reduce commuting costs, and negative because they are followed by externalities and provide access to neighborhoods for criminals. They concluded that the construction of new stations is accompanied by residential development; however, the increasing densification around the stations should be followed by government intervention in order to take control of the externalities created, both in high and low income neighborhoods. The Santa Clara (California) light-rail effect on property values was first examined by Weinberger (2000) who found that houses located within 0.5 miles from the stations have a higher lease rate than others, and later by Cervero and Duncan (2002) who found positive impacts of rail services in commercial land values. Moreover, Clower and Weinstein (2002) concluded that there is a positive linear relation of the dwelling prices and distance from light-rail stations in Dallas (Texas), while Hess and Almeida (2007) reached similar conclusions in Buffalo (New York) .

In Toronto (Canada), Bajic (1983) found that the transportation costs saved from the construction of the new (at that time) subway line were capitalized in the housing market. However, a few years later, Haider and Miller (2000) applied spatial autoregressive models and found that transportation infrastructure isn't a determinant of the values.

Henneberry (1998) examined how the house prices in Sheffield were affected by the supertram from 1988 to 1993. Debrezion et al. (2006) found that on the one hand Dutch houses proximal to rail stations are about 25% more expensive than those located 15 km away, while on the other, proximity to the railway is a negative factor probably because of the noise pollution. Moreover, commercial properties located 0.25 miles around the same stations are about 12.2% more expensive than dwellings (Debrezion et al., 2007). Martinez and Viegas (2009) examined the effects of transportation infrastructure accessibility on house prices by applying both linear and spatial econometric models. They found that better accessibility to metro stations increases house prices, while accessibility to rail stations has either negative or positive elasticities, depending on the characteristics of the line. Dorantes et al. (2011) found a positive impact of a new Metro line on house values in the South of Madrid. The negative impact of traffic externalities produced by trains, airplanes and automobilies in Netherlands, have been examined by Theebe (2004).

So et al. (1966) concluded that accessibility to minibuses in Hong Kong is a significant attribute for middle-income class households and as a result it affects the house prices. Bae et al. (2003) found that in Seoul, the proximity to the Line's 5 metro stations had a significant impact on prices, only before the line's opening. The results of a research in Shangai, China (Pan and Zhang, 2008) a city with intense transit development during the last few years, show that the rail stations shape land use development. Concerning the values, for every 100m that a house is closer to a metro station, its price increases 152 yuan (24 US$) per m^2.

The effects of the bus rapid transit (BRT) in Bogota, Colombia were examined by Rodriguez and Targa (2004), who concluded that the travel time savings are capitalized in the rental prices of houses and found an elasticity between -0.16 to -0.22 (for each 5 minutes additional walking distance from the station), and by Munoz-Raskin (2010) who found that properties close to the BRT owned by middle-income families, have higher values, while for those owned by low-income families the opposite happens. Löchl and Axhausen (2010) found a positive impact of the proximity to rail station in Zurich. Van Eggermond et al. (2011) concluded that in Singapore, the higher the distance from the MRT (Mass Rapid Transit) the lower the price.

Other papers summarizing the effects of rail transportation infrastructure on house and commercial property values, include Parsons Brinckerhoff Quade and Douglas Inc (2001), Martinez and Viegas (2009) and RICS Policy Unit (2002).

2.4.2 Airports

The impact of airports on residential property values has also been investigated in depth. In this case, while the negative externalities that are generated (e.g. noise) have negative impact (e.g. Nelson, 1980), the proximity to the airport implies higher prices, because of the increased accessibility (e.g. Lipscomb, 2003). Crowley (1973) found that the values of properties around the Toronto's International Airport vary in a period of 14 years. He found that residents were selling their houses after a "shock" period where the noise pollution was high, and then the new owners were either noise-indifferent or they probably change the land use. The result is that the prices finally reach the level they had before the 'shock' period. Feitelson (1989) analyzed the negative effects of noise produced by transportation infrastructure in Israel. Pennington et al. (1990) showed that there is not a significant effect of the aircrafts' noise in Manchester, a study that was later re-examined by Collins and Evans (1996) who considered Artificial Neutral Networks (ANNs) and reached contradictory conclusions. Cohen and Coughlin (2009) examined the case of the Atlanta International Airport and concluded that whenever the accessibility variable is omitted, the estimation is biased. Other similar studies include (Nelson, 1980), (Nelson, 2004; Cohen and Coughlin, 2008; Espay and Lopez, 2000), while for more papers on airports' noise externalities the reader can refer to Nelson (2008).

2.4.3 Highways

Cohen and Paul (2007) examined the impact of highway and airport infrastructure investments on capital asset values of U.S. manufacturing. Langley (1981) concluded that prices of houses located too close and too far from highways are low, while those located in a moderate distance are higher. Nelson (1982) examined the effects of highway noise on property values and found that house prices are discounted up to 0.63% per decibel of noise around the highways. Weinberger (2001) found that there is transportation costs capitalization in rents of properties with highway accessibility in Santa Clara. Ozbay et al. (2007) suggested an alternative methodology to estimate the marginal costs of highway transportation that incorporates the noise externalities produced by transportation. Davis and Jha (2011) proposed a framework that simulates the impacts –including the changes on house prices– of highway planning and construction on protected and low-income populations.

2.4.4 Other

Boarnet and Chalempong (2001) examined the impact of a toll road in Orange Country (California) and found it positive on house sale prices, due to the improvement of transportation accessibility. Kawamura and Mahajan (2006) applied spatial lag models to model the effect of traffic on house prices in Chicago.

Table 2.2: Selected literature on transportation infrastructure impact on property prices

Paper	Transport systam	Location	Property	Impact	Method
Nelson (1980)	Airport	six US cities	Houses	Negative	OLS
Langley (1981)	Highway	Washington D.C.	Dwellings	Higher in moderate distance	OLS
Nelson (1982)	Highway	Toronto, Spokane	Dwellings	Negative	OLS
Bajic (1983)	Subway	Toronto	Houses	Positive	OLS
Feitelson (1989)	Airport	Israel	Houses	Negative	OLS
Pennington et al. (1990)	Airport	Manchester (UK)	Dwellings	Negative	OLS
Kockelman (1990)	Work accessibility	San Francisco (CA)	Land values	Positive	OLS
Al-Mosaind et al. (1993)	Light rail	Portland (OR)	Single-family houses	Positive for distance < 0.5 km	OLS
Cervero and Landis (1993)	Urban rail	Washington DC	Dwellings	Positive	OLS
Collins and Evans (1996)	Airport	Manchester (UK)	Dwellings	Neutral	OLS
Benjamin and Sirmans (1996)	Metro	Washington DC	Houses	Positive	OLS
So et al. (1966)	Transport improvements	Hong Kong	Houses	Positive	OLS
Henneberry (1998)	Tram	Sheffield (UK)	Dwellings	Positive to neutral	OLS
Espay and Lopez (2000)	Airport	Reno-Spark, USA	Residencies	Negative	OLS
Weinberger (2000)	Light rail	Santa Clara (CA)	Dwellings	Positive for distance < 0.5 miles	OLS
Haider and Miller (2000)	Subway	Toronto (ON)	Dwellings	Neutral	Spatial
Bowes and Ihlanfeldt (2001)	Rail	Atlanta (GA)	Dwellings	Both positive and negative	OLS
Boarnet and Chalempong (2001)	Highways	Orange Country	Houses	Positive	OLS
Weinberger (2001)	Highway	Santa Clara	Rents of properties	Positive	OLS
Forkenbrock (2001)	Freight rail		Properties	Negative	OLS
Cervero and Duncan (2002)	Light rail	Santa Clara (CA)	Commercial land	Positive	OLS
Clower and Weinstein (2002)	Light rail	Dallas (TX)	Dwellings	Positive	OLS
Bae et al. (2003)	Metro	Seoul (South Korea)	Dwellings	Positive (before the opening)	OLS
Lipscomb (2003)	Airport	Atlanta, Georgia	Houses	Positive	OLS
Theebe (2004)	Trains	Netherlands	Dwellings	Negative	OLS
	Automobiles			Negative	
	Airplanes			Negative	
Nelson (2004)	Airport	Meta-analysis	Houses	Meta-analysis	OLS
Rodriguez and Targa (2004)	BRT	Bogota, Colombia	Land values	Positive	Spatial
Debrezion et al. (2006)	Rail	Netherlands	Dwellings	Positive	OLS
Kawamura and Mahajan (2006)	Highway	Chicago	Dwellings	Positive	Spatial
Hess and Almeida (2007)	Light rail	Buffalo (New York)	Dwellings	Positive	OLS
Debrezion et al. (2007)	Rail	Netherlands	Commercial	Positive	OLS
Cohen and Paul (2007)	Highway	USA	Manufacturing	Negative	Spatial
	Airport			Negative	
Pan and Zhang (2008)	Rail	Shanghai (China)	Dwellings	Positive	OLS
Nelson (2008)	Airport	Review	Properties	Negative	OLS
Martinez and Viegas (2009)	Metro	Lisbon (Portugal)	Dwellings	Positive	OLS and Spatial
	Rail			Both positive and negative	OLS and Spatial
Cohen and Coughlin (2008)	Airport	Atlanta	Houses	Negative	Spatial
Cohen and Coughlin (2009)	Airport	Atlanta	Houses	Varies	OLS
McMillen and Redfearn (2010)	EL	Chicago	Houses	Positive	Spatial
Munoz-Raskin (2010)	Bus Rapid Transit	Bogota, Colombia	Dwellings	Positive for middle-income negative for low-income families	OLS
Löchl and Axhausen (2010)	Rail	Zurich	Houses	Positive	OLS and Spatial
Van Eggermond et al. (2011)	MRT	Singapore	Houses	Positive	Spatial
Dorantes et al. (2011)	Metro	Madrid, Spain	Houses	Positive	Spatial

2.5 Policy evaluation indicators

Policy evaluation is based on single indicator measurements of the following categories: sustainability (Litman, 2011a), economic (Grazi et al., 2007), accessibility (Baradaran and Ramjerdi, 2001) and inequality measurements (Ramjerdi, 2006).

2.5.1 Sustainability

The general interest around *sustainability* and *livability* in transport development is growing. A report published by Litman (2011a) distinguishes the definitions of these two, as follows: *livability* refers to the objectives that affect the members of the community in a small scale (e.g. local pollution), while *sustainability* refers to a larger scale, such as climate change emissions.

The role of transport development in shaping sustainable cities is of major importance, because of the size of transportation infrastructure investments and their effects on economic, social and environmental development of both small and large scale environments. The cooperation of different scientific groups and sectors is needed in order to achieve the social, economic and environmental goals of sustainable transport.

Litman (2011a), indicates that the transport objectives that support the sustainability goals, are: 1) Transport system diversity, 2) system integration, 3) affordability, 4) resource (energy and land) efficiency, 5) efficient pricing and prioritization, 6) land use accessibility (smart growth), 7) operational efficiency and 8) comprehensive and inclusive planning. In more detail, the goals are: 1) economic productivity, 2) economic development, 3) energy efficiency, 4) affordability, 5) operational efficiency, 6) equity/fairness, 7) safety, security and health, 8) community development, 9) heritage protection, 10) noise prevention, 11) water pollution, 12) open space preservation, 13) good planning and 14) efficient pricing.

According to Litman (2011a), indicators are quantitative tools/variables that are used to evaluate whether the proposed policies fulfil the objectives of sustainability and achieve their goals. However, while the use of a particular set of indicators can enlighten the latent effects of a policy, evaluating it as positive for sustainable development, another set (or even the same, computed after different simulation runs) could show it harmful.

The multidimensional differentiation of the policy packages available at the policy maker, increases the difficulty of proper decision making on urban planing. As a result, the selection of suitable indicators for urban transport policy evaluation becomes very significant.

2.5.2 Economic

Two widely used sustainability, economic indicators are the Ecological Footprint (EF) and the Social Welfare (SW). The second overpasses the weaknesses of the first in failing to consider externalities (Grazi et al., 2007). SW is computed by the Social Welfare Function (SWF), which actually measures the human happiness. This indicator ranks the alternative policies suggested by the policy makers, so as to select the one to be implemented (the policy with the highest SW). The SWF has be developed, analyzed and criticized by a number of concrete research papers (eg. Arrow, 1953; Pattanaik, 1968; Coughlin and Nitzan, 1981). While SW refers to the societal preference at an aggregated level, the "cardinal social welfare" is an individual level measurement, computed by the agent's utility, which is usually the income.

Other indicators in the same context are the Gross Domestic Product (GDP) and the monetary income. Economic indicators of this category have been criticized repeatedly and their use should be careful. Comparing the amount of wealth, without taking inso consideration the way followed to be created (the level of environmental harm, the inefficiently spent of the wealth etc.), could lead to miscalculations and erroneous estimates.

2.5.3 Inequality

The term "equity" refers to the fairness of the policy impacts distribution, over the population. The impacts are indicated by the benefits and costs generated by a policy, and are usually measured by the change of the population's wealth. The inequality measurements are equity indicators that quantify the difference of the fairness within the population. They are used to evaluate policy scenarios, simulated by LUTI models. However, the literature related with their use in the context of the agent-based LUTI models is limited, if not absent.

According to Ramjerdi (2006), the inequality measurements are classified into the following three categories: 1) statistical, which measure the distribution of a characteristic in the population; these are the range, variance, measure of variation, log variance, Gini coefficient and Theil index; 2) welfare, which are based on the welfare economics and functions; these are the Kolm and Atkinson measurements; and 3) axiomatic. Ramjerdi (2006) used these measurements to assess the effects of transport policies in Oslo. She concluded that the results can be interpreted differently when using different indicators. Talen (1998), early recognized the potential benefits of visualizing the output of the inequality measurements in Geographical Information Systems, for the planners.

Viegas (2001) suggested wider perspective of equity when planning congestion charging schemes, in order to cover both horizontal and longitudinal dimensions. Franklin (2006) evaluated road pricing scenarios using among others, inequality measurements in mode choice models. Santos et al. (2008) integrated equity objectives into the accessibility maximization function for a road network design model. Levinson (2010), published a theoretical work where he attempts to interpret the effects of road pricing policies on equity. He concludes that the revenues of road pricing schemes can be used accordingly, to achieve the desired equity results.

The most applied inequality measurement is probably the Gini coefficient, which measures the statistic dispersion. It takes values between zero, where the characteristic that is measured (income or wealth) is equal everywhere, and one. It is derived from the Lorenz curve, which plots the cumulative share of income of a population (y axis) over the share of people from lowest to highest income (x axis). Gini is the ratio of the difference between the line of equality and the Lorenz curve, over the total area shaped by the line of equality. Moreover, it can be computed by the following formula:

$$G = \frac{\sum_{j\in 1}\sum_{k\in 1}|X_j(y) - X_k(y)|}{2n^2\overline{X}} \tag{2.19}$$

where n is the number of groups that belong to N and $\overline{X}$ is the characteristic that is being measured.

Although Gini is the most popular and widely cited inequality measurement, another indicator is used in this research, namely Theil index. In its general form it is equal with the difference between the maximum possible entropy of the data, and the observed. It takes values from zero (perfect equality) to ln(x). The benefit of using Theil index instead of Gini is that it can be decomposed in different subgroups, such as spatial units. For instance, Theil can be computed separately for each zone of an area –measuring the inequality at disaggregate level– and then cumulate the resulted values to compute the total inequality of the area. Elbers et al. (2005) analyzed decomposed indices to measure the income inequality in and between countries. Novotny (2007) suggested a decomposed Theil index for cross-country comparison of regional inequality. Kemel et al. (2008) used a Theil index decomposition to measure the inequality of accessibility for people in California, suggesting the application of this method for fairer allocation of transportation investments in the area.

The decomposable formulation of the Theil index is:

$$T = \sum_{g\in G}\widehat{X}_m \cdot T_g + \sum_{g\in g}\widehat{X}_g \cdot ln\left(\frac{\overline{X}_g}{\overline{\overline{X}}}\right) \tag{2.20}$$

where,

$$T_g = \frac{1}{N} \sum_{j \in N} \left(\frac{X_j}{\overline{X}} \right) \cdot ln \left(\frac{X_j}{\overline{X}} \right) \tag{2.21}$$

G is the number of groups
$\widehat{X}_g$ is the weighted characteristic of group g over all the G
$\overline{X}$ is the average value of the measured characteristic to all groups
$\overline{X}_g$ is the average value of the measured characteristic of each group
T_g is the Theil index of each group
N are the centers of region g

2.5.4 Accessibility and land-use

Accessibility measurements are theoretical formulations that indicate the ease of visiting the locations where activities are conducted. Accessibility affects the household location choice (e.g Vandenbulcke et al., 2009) and as a result the housing prices (e.g Medda, 2012; Ibeas et al., 2012). Gutiérrez and Urbano (1996) attempted to forecast the resulted, by the implementation of the Trans-European road network, increase of accessibility in Europe, using an indicator based on the impedance from country to country, and the GDP. Linneker and Spence (1996) measured the impact of the increase in accessibility resulted by the construction of the M25 London Orbital Motorway, in regional development. Geurs and Wee (2004) presented an extensive review of accessibility indicators used for land-use and transport strategies. They identify four types of components in the current accessibility indicators, that are based on: 1) land-use (e.g. the supply and demand of the opportunities distributed spatially); 2) transportation (e.g. travel time); 3) temporal (e.g. availability of the opportunities in day); 4) individual (e.g. personal characteristics). Moreover, they identify that there are four basic perspectives on measuring the accessibility, namely: 1) infrastructure-based; 2) location-based; 3) person-based; 4) utility-based. According the same research, accessibility is being used as a way of measuring the operationalization, the interpretability and communicability, as a social or economic indicator.

In its simplest form, accessibility is measured by the ease that each land-use unit can be accessed by each transport mode, as follows:

$$A_i = \sum_{j \in L} \frac{1}{f(c_{ij})} \tag{2.22}$$

Where A_i is the accessibility at location i, c_{ij} is the cost of travel from location i to location j.
The most widely applied accessibility measurement is the gravity model. In this form, accessibility is measured by considering the number of opportunities that are available to travelers, by the equation:

$$A_i = \sum_{j \in L} \frac{W_j}{f(c_{ij}, \beta)} \tag{2.23}$$

Where W_j is the number of opportunities at the location j, and β is the coefficient of the cost.

A more sophisticated way to measure accessibility is based on the utility. There are two types of utility-based accessibility measurements. Ben-Akiva and Lerman (1985) suggested the *logsum* term, which uses the denominator of the multinomial logit model. The second is based on the doubly constrained entropy model (Martinez, 1995). Banister and Berechman (2001) suggest that the accessibility is the engine behind the economic growth of an area after the implementation of a policy, because it leads to increase of the employment and productivity. Vandenbulcke et al. (2007) analyzed static accessibility indicators to evaluate different transportation models in Belgium. More recently, Martinez and Viegas (2013) applied a methodology that is widely used in botany, to model the distance-decay functions for accessibility assessment.

2.5.5 Welfare

De Jong et al. (2005) reviewed the current policy evaluation measurements, emphasizing on the rule-of-the-half and the consumer surplus, which is computed by the difference between the logsums of different scenarios. Ben-Akiva and Lerman (1985) suggested that accessibility is similar to the consumer surplus, equals to the maximum expected utility of the travelers. McFadden (2001) proved that the Hicksian and Marshallian measures of consumer surplus coincide, and a "mixed logsum" formula can represent them when the utility is linear and additive to income. Karlström (2000) suggested that McFadden's formula can only be used when the marginal utility of money is constant. The benefits of using the rule-of-the-half as a policy evaluation measure have been analyzed by Bates (2003). Von Haefen (2003) suggested a methodology for measuring the welfare, by using RUM that incorporates the individual's observed choice. Odec et al. (2003) used the logsum of the generalized cost instead of the rule-of-the-half in the consumer surplus function, and computed the welfare resulting from a congestion pricing policy in Oslo. Cherchi et al. (2004) measured the welfare using the compensating variation by resampling, the logsum and the rule-of-a-half, and found that the results are biased when compared with the correct values, resulted by a simulation of disaggregate compensating variation. Compensating variation (CV) indicates the minimum amount of money that a consumer has to be compensated after the implementation of a policy, in order to be at the same level of utility as before. Kalmanje and Kockelman (2004), measured the CV as the difference in the maximum utility between two transport policy scenarios in Austin, Texas, where the utility is the logsum of the generalized cost. Zhao et al. (2008) concluded that the mean CV result of transport policy cannot be estimated accurately even after 1 million simulations. They also suggest that the confidence intervals of the welfare measures should be computed, since they are necessary for policy analysis.

The compensating variation is computed by the formula:

$$CV_n = (1/\mu_n) ln\left(\sum_{j=1}^{J} e^{V_{nj}}\right) + C \tag{2.24}$$

Where μ_n is equal to dU_{nj}/dY_n, Y_n is the income of person n and U_n is the overall utility, and C is the constant that represents the unknown factors.

The difference of the logsums between a policy (J=1) and the base case (where J=0):

$$\Delta CS_n = (1/\mu_n)\left[ln\left(\sum_{j=1}^{J^1} e^{V^1_{nj}}\right) - ln\left(\sum_{j=1}^{J^0} e^{V^0_{nj}}\right)\right] \tag{2.25}$$

In UrbanSim, the logsum accessibilities per zone are computed by MATSim (Nicolai and Nagel, 2011), as follows:

$$A_i = \frac{1}{\beta_{Scale}} \cdot ln(\sum_{j=1}^{J} (W_j \cdot exp(-\beta_{Scale} \cdot c_{ij}))) \tag{2.26}$$

where, A_i is the workplace accessibility at location i, $i \in I$ the origins, $j \in J$ the destinations, β_{Scale} is a scale factor related to the scale of a logic model, W_j is a weight giving the number of jobs at location j, $exp(-\beta_{Scale} \cdot c_{ij})$ is a deterrence function, c_{ij} is the generalized travel cost from location i to location j.

The generalized travel cost c_{ij} is:

$$c_{ij} = (\alpha \cdot ttime) + (\beta \cdot ttime^2) + (\gamma \cdot ln(ttime)) + (\delta \cdot tdistance) + (\epsilon \cdot tdistance^2) + (\zeta \cdot ln(tdistance)) + (\eta \cdot tcost) + (\theta \cdot tcost^2) + (\iota \cdot ln(tcost)) \tag{2.27}$$

where $ttime$ is the travel time in minutes, $tdistance$ is the distance in meters, $tcost$ is the monetary travel cost, α to ι are the marginal utilities

In this research, the default values of MATSim were used: $\beta_{Scale} = 1$ and $\alpha = -12$. The other values were set to zero.

Home and work locations are distributed within a given radius from the centroid of each zone, in order to avoid that all the household and workplace locations are attached at the same link of the road network. Because of the uneven sizes of zones in our case, a radius of 500 meters was selected, representative of an average situation.

2.6 Emerging public transit policies

2.6.1 Bikesharing

Despite the growing global motorization, bike-sharing systems' demand, as a sustainable alternative transport mode, is continuously increasing. Such systems combine the advantages of bike usage, such as low cost, autonomy, flexibility, accessibility and health benefits, with the advantages of renting (as opposed to owning). Significant experience has already been gained regarding security, insurance and liability concerns, bicycle redistribution, applications of information technology systems, planning, management and pre-launch considerations.

The first bike-sharing program, the "White Bike Plan", was introduced in 1965 in Amsterdam. The scheme (that was based on fifty free, white-coloured bikes scattered around the city) was abandoned because its bicycles were soon damaged or stolen (Shaheen et al., 2010a). Other revolutionary schemes were operated in La Rochelle, France, (1974) and in Cambridge, UK, (1993). The second generation was based on coin-deposit bikes, while the third brought an additional component: the use of information technology (IT). The system design is generally based on three aspects: 1) bikes must be distinguishable; 2) docking stations must be scattered around the city, with user interface technology to allow check-in and check-out of the bicycles; 3) advanced technology should be used for security (e.g. GPS tracking and provision of users' personal information). Other considerations point toward a fourth generation of systems, including innovations such as the use of electric bikes, providing incentives to the users to return the bicycles to stations where the demand is expected to be higher (DeMaio, 2009) and improved interoperability with other modal systems (e.g. public transport and car-sharing).

In a recent study, Shaheen et al. (2011) examined the degree of adoption and behavioral response of the users of the largest bike-sharing system in the world, in Hangzou (China). The authors found that the scheme, 1.5 years after its implementation, captured modal share from the bus transit, cars, taxis and walk, while 30% of its members use it for most of their trips. The results showed that the acceptability of the program could be improved by providing bike/parking availability information, adding more stations, extending the hours of operation and offering better bike maintenance.

In Europe, the project OBIS (Optimizing Bike-Sharing in European Cities) aims to evaluate the bike-sharing systems of 10 European countries, identify the success factors and market potentials and publish a handbook to guide European decision makers for the efficient implementation of such schemes (Büttner, 2010). Another research project called PRESTO (Dufour and Ligtermoet and Partners, 2010) aims to promote cycling to everyone as a daily transport mode. The project consortium aims to develop a general Guide outlining the fundamentals of an integrated cycling policy, but also three area-specific policy guides for infrastructure, bicycle promotion and the electric-bikes (Pedelects) which could be a good sustainable solution for hilly areas.

According to Shaheen et al. (2010b) there are approximately 150 programs in more than 125 cities and 30 nations around the world, with at least 140,000 bikes. In Europe, more than 19 countries have bike-sharing systems, while in Greece the scheme was first introduced in 2010 (www.eaybike.gr), in the island of Corfu.

Another bike-related transport system is bike-rental. Companies may use it for their employees and residential units for their residents. An electric-bicycle-rental project is the E-Aix in the city of Aachen (www.ladenez.de).

The undisputed benefits of bike-enabled transport systems are sometimes not enough to persuade decision makers to exploit them and urban planners to design bicycle-friendly areas. Hendriks (2010) suggests that a way to capture their interest with future perspective is by challenging designers to give attention to bicycle routes to schools, emphasizing the safety for children.

2.6.2 Carsharing

Unstable fuel prices and increasing maintenance costs, as well as the insurance and purchase cost of a car, make car ownership a luxury that not many people can afford. Under these circumstances, car-sharing attracts more and more people. Users can enjoy the privacy of any type of car (e.g. compact car, SUV, van, and luxury) depending on their current needs, without the need and commitment of a purchase. The users of such systems typically pay a combination of (i) an amount for registration to the system database, (ii) a monthly fee and (iii) a cost of use, depending on the time used and/ or the distance traveled. Depending on the system, fuel, maintenance, insurance, parking and sometimes congestion charging costs (e.g. in the case of London's Zipcar) are included in the price. Thus, users do not understate the actual cost of the trip by miscalculating the variable costs, an error which is usually committed (Shaheen and Cohen, 2007, Walsh, 1990, Shaheen et al., 2009, Cohen et al., 2008 and Morency et al., 2008) leading to an underestimation of car travel costs.

Car-sharing was introduced in Zurich in 1948 but did not become popular until the early 1990s (Shaheen and Cohen, 2007). Experimental projects in Europe, included the Procotip (France, 1971), Witkar (Amsterdam, 1974), Green Cars (Great Britain, 1977), Bilpoolen, Vivallabil and Bilkooperativ (Sweden, 1979, 93, 85) and in USA the Short-Term Auto Rental (STAR) (California, 1983 to 1985) and the Mobility Enterprise, a research project of Purdue University from 1983 to 1986 (Shaheen and Cohen, 2007). However, the first two successful programs were in Switzerland and Germany in the late 1980s. Nowadays, car-sharing systems are operated in almost all European countries, USA, Israel, Japan, Singapore, China, Malaysia, Australia and other countries. In total, approximately 348,000 members in 600 cities of 18 countries, use such systems (Shaheen and Cohen, 2007).

Research has documented differences in the users' demographics, due to the different characteristics of each city. A study in North America (Burkhardt and Millard-Ball, 2006) found that car-sharing members are usually well-educated and environmentally aware. The highest percentage of users is between 25 and 35 years old. Because of insurance restrictions, there are not many members below 21 years of age. Half of the users have income more than 60,000$ and 72% of them live in a household with no other cars, while the average household size is 2.02 persons. Car-sharing operators suggest the use of the system particularly to those that do not drive more than 10,000 to 16,000 km per year (Shaheen and Cohen, 2007), while it is a good choice for students and low-income households (Shaheen et al., 2004). Zhou et al. (2011) found that in Austin, adults of higher income that own a car do not wish to join the scheme, while the education level does not determine the possibility of becoming a member.

Germany, Switzerland and USA have higher member-to-vehicle ratios, because of many inactive members (Shaheen and Cohen, 2007). Difficulties of car-sharing projects are usually related to the car and driver's insurance, as well as the on-street parking that is not available and free to all countries. Insurers usually hesitate to co-operate with carsharing companies, because the risks are not clearly defined. Solutions, such as the Pay-as-you-drive (PAYD) idea may offer a solution for this problem (Litman, 2011b). Shaheen et al. (2010a) examined the parking policies of carsharing in San Francisco bay Area and suggest that as the scheme expands, the public entities need to consider formal policies to allocate the parking space for it.

Carsharing companies invest on the development of new user-interface technologies in order to increase the flexibility of the users and attract more members. The results of a study that was conducted by Nerenberg et al. (1999) in San Francisco between 1996 and 1998, shows that the women who are attracted by electric car-sharing are mainly driven by environmental incentives, while men because they found interesting the technical perspective of the service, revealing that the system should not only be functional, but have technological and environmental aspects. Known carsharing providers around the world are Greenwheels, in Netherlands and Germany, CityCarClub in Sweden, Finland and Zipcar, in USA, Canada and UK. Zipcar and Flexcar, the two major US companies, were merged in 2007. The business model of these companies is usually different. CityCarShare is a non-profit company and returns the revenues to the community, while on the other hand Zipcar is for-profit. Moreover, they use vehicles of different type in their fleets.

Carsharing companies now invest on new user interface technologies for reservation, check-in and check-out and on the integration of hybrid and electric vehicles in their fleets. New carsharing services and applications improve the scheme's flexibility and competitiveness. In the age of social networks, peer-to-peer solutions are also emerging. For example, innovative services based on websites and smartphone applications (e.g. www.getaround.com) respectively to match unused vehicle periods of car owners with demand for car use by others. The owner of a car can thus rent it to other registered users of the service, for the time (s)he does not use it. The resulting arrangement (an ad-hoc network of car-owners and users) offers multiple benefits for both the groups. Naturally, many issues and challenges remain.

2.6.3 Social and environmental effects

Carsharing reduces vehicle ownership. Millard-Ball et al. (2006) found that many people canceled a car purchase or sold their car after joining a car-sharing program and each car-sharing vehicle replaces between 4 and 23 vehicles depending on the characteristics of the city. A study in North America (Martin et al., 2010) found that 60% of car-sharing members do not own a car and the system has removed between 75,000 and 94,000 vehicles from the road. The reduction of vehicle ownership implies a reduction in air pollution and traffic congestion and increases the available parking slots. Carsharing policies lead to the reduction of Vehicle Miles/Kilometers Traveled (VMT/VKT) and the GHG (Rodier and Shaheen, 2004; Rodier, 2009). In North America the reduction is 44% per car-sharing user (Shaheen et al., 2009). According to Lane (2005), car-sharing participants report increased environmental awareness after joining the program. Finally, households can save more money for their development (Ciari et al., 2009).

A way to measure the environmental impacts of bike-sharing is by multiplying the kilometers that bikes make per year with the average emissions of a motorized mode. For instance, bicycles of Velib in Paris make approximately 78,000 trips of 20 min and 312,000 km per day. If these were motorized, 57,720 kg of CO2 would have been released. Furthermore, bike-sharing systems can be followed by bicycle-lane infrastructure improvements. Bicycle riding in Paris increased 70% after the introduction of Velib (Shaheen et al., 2010b).

Table 2.3 compares car – and bike–sharing characteristics. All schemes reduce traffic congestion and alleviate the parking-shortage problem. They lead to time and money savings, lower emissions; they benefit the psychological health and improve the urban design if their stations are optimally located. Concerning the infrastructure demands, bike-sharing requires parking stations with interface technology, but also bicycle lanes across the streets. Car-sharing schemes may not require additional infrastructure, but the parking reservation is a major problem (Shaheen et al., 2010a,c), especially in the cities where it is not available and free. If the fleet includes electric vehicles, additional infrastructure is needed to provide electricity. As was mentioned before, insurance is a major problem for car-sharing companies. The latest generations of these systems use advanced information technology such as GPS tracking, smartcards, telephone and internet booking facilities (often with dedicated smartphone applications). Bike-sharing's additional advantages include higher flexibility, interoperability with transit (Rodier and Shaheen, 2008), health benefits and better accessibility.

Table 2.3: Comparison of car – and bike –sharing characteristics

		Bike-sharing	**Car-sharing**	**Electric car-sharing**
Factors affected				
Transport	Traffic	✓	✓	✓
	Parking demand	✓	●	●
Social-environmental	Emissions	✓	●	✓
	Time	✓	✓	✓
	Cost	✓	●	●
	Urban design	✓	✓	✓
Personal	Physical health	✓	×	×
	Psychological health	●	●	●
Before and after installation concerns and amenities				
Installation	Infrastructure	×	●	×
	Parking	●	×	×
	Insurance	✓	×	×
Information Technology	GPS tracking	✓	✓	✓
	Internet booking	✓	✓	✓
	Telephone booking	✓	✓	✓
Movement	Gasoline	✓	×	✓
	Electricity	●	×	✓
	Human power	✓	×	×
Other	Flexibility	✓	●	×
	Access	✓	●	●

Notes: ✓:Positive, ●: Unclear, ×: Negative

2.7 Statistical modeling

2.7.1 Factor analysis

Factor analysis aims to reduce the number of variables in a dataset, by identifying common patterns between the p variables, and revealing unobservable factors describing their correlation. Similar to principal components analysis (PCA), factor analysis is based on the correlation matrix; however, it uses a specific statistical model (Washington et al., 2003). The method was first developed in the field of psychology by Pearson and Spearman, intending to give some insight into the psychometric measurements related with unobservable variable intelligence (Johnson and Wickern, 1992). It is suggested that factor analysis should not be applied blindly to any dataset hoping to identify common patterns so as to reduce the number of attributes, but its application should be driven by theoretical motivation. According to Washington et al. (2003), factor analysis is expressed as follows.

$$\begin{aligned} X_1 - \mu &= l_{11}F_1 + l_{12}F_2 + \ldots l_{1m}F_m + \varepsilon_1 \\ X_2 - \mu &= l_{21}F_1 + l_{22}F_2 + \ldots l_{2m}F_m + \varepsilon_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_p - \mu &= l_{p1}F_1 + l_{p2}F_2 + \ldots l_{pm}F_m + \varepsilon_p \end{aligned} \tag{2.28}$$

where F are the factors, μ_i are the means and l_{ij} are the factor loadings, ε_i is associated only with the X_i and p random errors and m factor loadings are unobservable or latent.

Alternatively, in matrix notation, the model is given as:

$$(X-\mu)_{p\times 1} = L_{p\times 1}F_{m\times 1} + \varepsilon_{p\times 1} \tag{2.29}$$

These formulas have p equations and $p+m$ unknowns, which means that additional information is needed in order to obtain a unique solution. This information is given in the form of rotation models, resulting in orthogonal or oblique factor analysis models. For instance, *varimax* rotation maximizes the sum of the variances of the factor loadings, leading to orthogonal rotation. Oblique factor analysis models result in a more understandable factor structure, since they relax the restriction of uncorrelated factor loadings. The aim of using different rotations is to move the loadings closer to 1 or 0. A loading closer to 1 implies a more significant influence of the variable to the factor, while a loading closer to 0 implies low significance. As a result, extreme factor loadings are more desirable.

2.7.2 Ordered logit model

Survey respondents are usually asked to express their preferences in rating scale, usually called *Likert* (Likert, 1932). The respondents are asked to select usually among 4, 5, 7 or more alternative responses in an ordered scale, such as: very satisfied / satisfied / neutral / dissatisfied / very dissatisfied or very important / important / neutral / unimportant / totally unimportant. A multinomial logit could be used instead; however, the ordered form implies the independence of the error terms for each alternative, and therefore the Independence for Irrelevant Alternatives (IIA) assumption of the logit model (Likert, 1932). Ordered response is more similar with the alternatives close to each other, and less to the more distant alternatives. The ordered model estimates the parameters of the alternatives, as well as the threshold values between the choices. In case of four alternative ordered responses, there are three thresholds (or critical values) that separate the choices. When the utility of the respondent is between k1 and k2, he/she selects the alternative "Not Good' , while when it is more than k3, the alternative "Very Good" is selected.

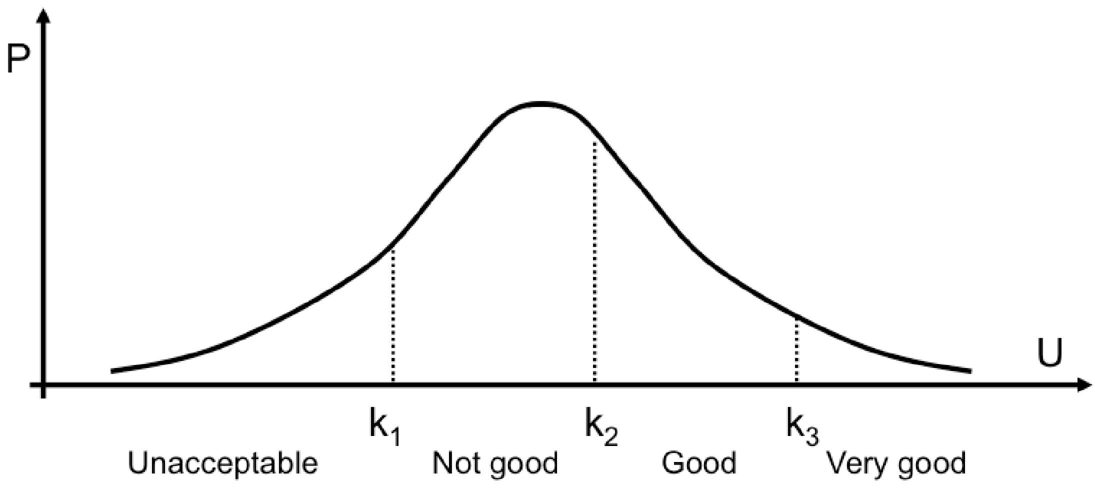


Figure 2.1: Distribution of respondents preference. Adapted from Train (2009).

2.7.3 Hybrid choice and latent variable model

Structural equation models (SEM) are used to reveal latent behaviors of the respondents. Within the last few years, SEMs have met increasing interest in transportation research. They have been applied, for example, in modeling travel behavior (Scheiner and Holz-Rau, 2007), residential location choice (Van Acker et al., 2010) and modal choice (Tyrinopoulos and Antoniou, 2013). Golob (2003) provides an extensive review of SEM applications in transport (until 2003). Walker (2001) introduced the hybrid choice and latent variable models, which integrate structural equation models in discrete choice models. In these models, latent, attitudinal explanatory variables are structured by discrete-response indicators. Since Walker (2001), behaviors such as convenience, environmental consciousness, safety, comfort and flexibility have been proved to play a crucial role in mode choice (eg. Espino et al., 2006). Likewise, the latent class models are used to reveal different groups-classes of individuals with similar tastes/choices. More recently, Atasoy et al. (2012) developed hybrid choice models with latent variables and classes to model mode choice in Switzerland. The modeling framework used in this research is as follows:

Structural equation:

$$Z^*_{ln} = X_n \lambda_l + \omega_{ln}, \omega_n \sim N(0, \Sigma_\Omega) \tag{2.30}$$

Where l is the index number of the latent variable, X is the explanatory variable, λ is the coefficient of the explanatory variables, ? is the latent variable and ω is the error term.

$$U_n = X_n \beta_l + Z^*_n \beta_2 + \varepsilon_n,\ \varepsilon_n \sim \text{standard logistic} \tag{2.31}$$

Where U is the utility, Z are the latent variables and β the coefficients of the latent variables.

Measurement model:

$$I_{rn} = Z^*_n \alpha_r + u_{rn} \tag{2.32}$$

Where r is the total number of indicators.

2.8 Discussion

LUTI models have received much –and is expected to receive more– attention in the literature. However, there is not clear evidence that the number of LUTI models around the world –and as a result the money and time spent for their development– complies with the expectations of rendering them flexible, policy evaluation tools. While the transportation research community is focused on the development of complex location-choice methodologies, the use of spatial econometrics –that are widely, successfully applied in environmental sciences– in LUTI models, has not been extensively explored, revealing a discontinuity in the literature.

Various sustainability, economic, inequality and accessibility indicators for policy evaluation have been proposed. However, the literature is focused on aggregate indicator measurements, while the strengths of microsimulation in three dimensions –agents, space and time– have not been exploited, yet.

Public transit systems attract people with different socio-economic characteristics. The behaviors/characteristics of users determine the demand and consequently the success of the system. Factor analysis, and hybrid discrete choice and latent variable models are used to investigate the latent characteristics of public transit users, and the development of more accurate prediction models.

3 Spatial econometric model applications for LUTI

3.1 Research objective and chapter structure

The objective of this chapter is to investigate the prospect of using spatial econometrics in LUTI models. More specifically, the contributions of this chapter are summarised below: Real estate price models of the two major Hellenic cities Athens and Thessaloniki are developed, particularly focusing in measuring the impact of transportation infrastructure and policies on dwellings' purchase prices and rents; the effect of the crisis on values of residencies in Athens is quantified, and disaggregate price–to–rent ratio models are developed; the efficiency of spatial econometric models with discrete response for modeling the land-use change using land-cover data is explored, by measuring the impact of large scale transportation infrastructure on urbanization; finally, a spatially-aware framework for optimal site selection of demand-oriented transport facilities –such as car/bike-sharing stations and electric car chargers– is proposed.

The rest of this chapter is structured as follows: Section 3.2 describes the real estate price models of Athens and Section 3.3 of Thessaloniki. Section 3.4 presents a price–to–rent ratio model, and Section 3.5 the impact of the economic crisis on residential values. Section 3.6 investigates the use of spatial econometrics in modeling the land-use/cover change. Finally, Section 3.7 presents a framework for optimal site selection of demand-oriented transport facilities.

3.2 Real estate prices in Athens

3.2.1 Transportation infrastructure in Athens

Athens is the capital of Greece, with a metro area –about 427 km^2– of approximately 4 million citizens (one third of the population). About 25% of them lives in the city center, where is also located the very high percentage (30%) of jobs (Milakis et al., 2008). This indicates the need for multiple transportation connections between the center and the suburbs. The city is accessible by car, rail, sea and air. The metropolitan area has as natural barrier the mountains and the sea, which imposes constraints on the planning and development of the transportation systems. The city centre is suffering from severe traffic congestion, which generates increased levels of air pollution.

During the last decade, mainly due to the Olympic Games in 2004, major transport infrastructure has been constructed. Figure 3.1 shows the main transportation infrastructure in Attica (greater area of Athens)

Public transport

The Athens metro network is composed by three lines: Line 1 is the oldest and is usually considered as a different mode, called "electric railway". It is managed by ISAP and connects Piraeus with the Northern suburbs crossing the city center. The other two lines were constructed between 1990 and 2000, when they were delivered. They are operated by AMEL, a subsidy of the Attiko Metro SA. Currently, the network is extended, while a new line (Line 4) is planned to be constructed in the near future.

Despite the range of the railway networks, the buses (ETHEL) remains the public transport system with the highest demand, because of the extended area of its service. The bus service is characterised by large delays, low frequency and unreliability. The average speed is 7 km/h in the city center and 18 km/h in the suburbs -due to the limited length of bus lanes (51km) which are usually violated- while the corresponding average speed of the European Union is 20 km/h (Hellenic Ministry of Public Works, 2009).

Road network and congestion control policies

The Attica Tollway –constructed in 2003– is a 65 km ring road of the greater Athens metro area. Prior to the current financial crisis it was serving approximately 300,000 cars per day (Halkias and Tyrogianni, 2008). The suburban railway is constructed in the central reservation of this highway.

In 1980s, a congestion control scheme aiming to restrict the number of cars entering the city center, was introduced. Cars were allowed to enter the city center only the odd or even days of the month, depending on the last digit of their license plates. This policy had the expected results only in the short-term, while in the long-term drivers bought a second car with different last digit. The increase of car ownership had negative effects on the environment, because the second car was usually older (Giaoutzi and Damianides, 1990). This response was not anticipated and rendered the measure counter-productive. However, a new measure named the "green inner ring" has been implemented since 1st September 2012, allowing the entrance of eco-friendly cars to the city centre.

Other policies aimed at eliminating the parking problem in the city center, where there are 33,000 parking lots for 50,000 cars, while it is estimated that 8,000 cars are parked illegally every day. Parking stations were constructed next to the metro stations with high daily demand, while a number of Municipalities have introduced parking pricing schemes, a measure with positive results until now (Spiliopoulou and Antoniou, 2012; Municipality of Athens, 2008).

Air and maritime transport

The Athens International Airport –constructed in 2003– is located in a low density area, 30 km from the central business district (CBD), in Spata. In Attica there are two ports (Piraeus and Rafina) and seven marinas (Alimos, Glyfada, Faliro, Zeas, Vouliagmeni, Lavrion and Floisvos).

Trip share

The Hellenic Institute of Transport Engineers (2008) found that 55 % of the Greek roads were congested in 2008, while it was estimated that ceteris paribus in 2010 the corresponding percentage would be 95% (the current financial situation has provided a temporary relief). Despite the current financial situation, the increase of gasoline price and the need for new infrastructure indicates the future trends. Athens Urban Transport Organization (2009) estimated that in 2008 the demand for transport in Athens during an average working day, was 8 million trips, while the prediction for 2012 was 10 million. Public transport holds 40% of the transport share, 50% of trips are made by motorized vehicles and the remaining 10% are made on foot. 40% of the trips are to –and from– work, 12% for shopping, 9% for leisure, 15% for personal reasons, 6% for education and 7% for social reasons (Athens Urban Transport Organization, 2009). The rush-hour peaks between 7 and 9 am in the morning and from 4 to 6 pm in the afternoon. Furthermore, about 340,000 vehicles (motorbikes are excluded) enter the CBD per day.

3.2.2 Model development

Different linear models are estimated using both the "purchase" and "rent" datasets described in Section 6.3.1: 1) Hedonic price models, where the dependent variable is the natural logarithm of price and the explanatory variables are either continuous, dummies or natural logarithms, depending on their spatial variability; 2) log-log models, following DiPasquale and Wheaton (1996), who suggest the use of both explanatory and response variables as logarithms for house price model estimations; 3) generalized additive models (GAM) using quadratically penalized likelihood maximization (penalized least squares) with splines (p-splines). The latter method (Wood, 2004) uses isotropic smooths of the independent variables and offers advanced numerical stability. The models of the above groups are referred to as Model 1, 2 and 3, respectively, while the term HPModel is used

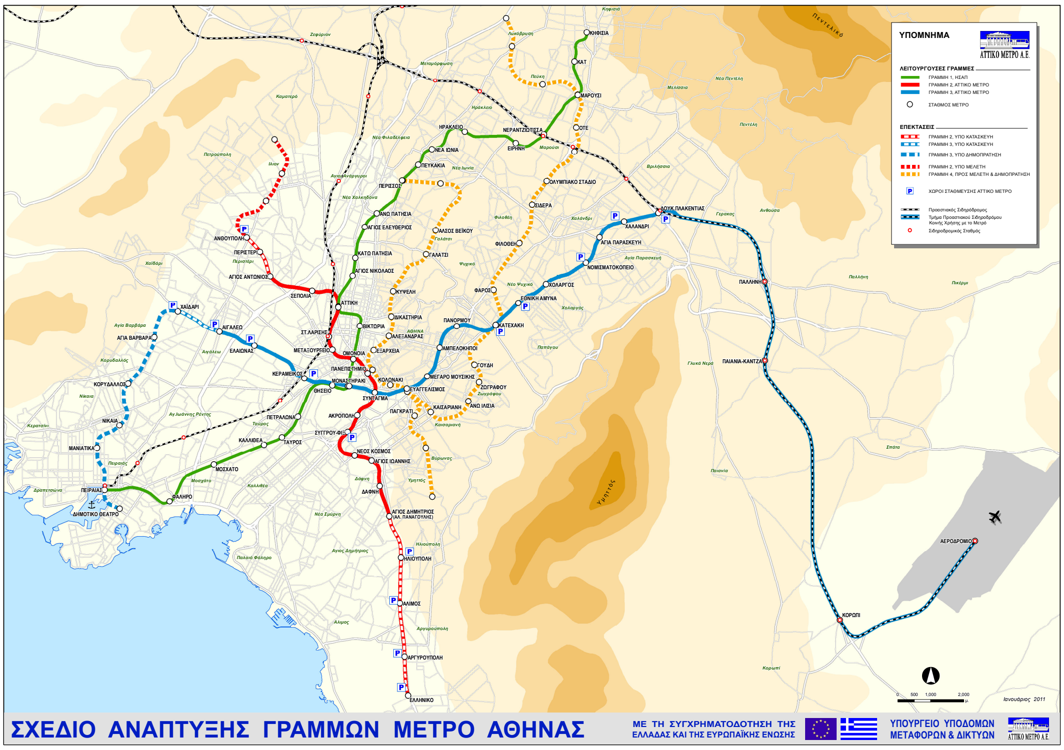


Figure 3.1: Central transport network in Athens (source: http://ametro.gr)

for the "purchase" price models and HRModel for the "rent". Table 3.1 describes the variables used in these models. The dummy variables of Model 2 were recoded from 0/1 to 1/2 so as to be specified as logarithms (as suggested in DiPasquale and Wheaton, 1996). In Model 3, many continuous explanatory variables were used in p-spline form (Table 3.1). Functions from the R package "mgcv" (Wood, 2014) are used for its estimation.

Purchase model (Model 1):

$$\begin{aligned} log(Y) = {} & \beta_{sqm} log(X_{sqm}) + \beta_{\lesssim 51} X_{\lesssim 51} + \beta_{61} X_{61} + \beta_{71} X_{71} + \beta_{81} X_{81} + \beta_{91} X_{91} + \\ & \beta_{01} X_{01} + \beta_{\gtrsim 11} X_{\gtrsim 11} + \beta_{base} X_{base} + \beta_{f1} X_{f1} + \beta_{f2} X_{f2} + \beta_{f3} X_{f3} + \beta_{f4} X_{f4} + \\ & \beta_{f5} X_{f5} + \beta_{\gtrsim f6} X_{\gtrsim f6} + \beta_{parking} X_{parking} + \beta_{openP} X_{openP} + \beta_{storage} X_{storage} + \\ & \beta_{firep} X_{firep} + \beta_{AH} X_{AH} + \beta_{AC} X_{AC} + \beta_{sfh} X_{sfh} + \beta_{seaview} X_{seaview} + \\ & \beta_{front} X_{front} + \beta_{ring} X_{ring} + \beta_{bus} X_{bus} + \beta_{metro} X_{metro} + \beta_{isap} X_{isap} + \\ & \beta_{natl} X_{natl} + \beta_{railway} X_{railway} + \beta_{sub} X_{sub} + \beta_{tram} X_{tram} + \beta_{air} X_{air} + \\ & \beta_{port} log(X_{port}) + \beta_{mar} X_{mar} + \beta_{rroad1} X_{rroad1} + \beta_{rroad2} X_{rroad2} + \\ & \beta_{rroad3} X_{rroad3} + \beta_{cbd} X_{cbd} + \beta_{NS} X_{NS} + \beta_{uni} X_{uni} + \beta_{arch} log(X_{arch}) + \\ & \beta_{coast} X_{coast} + \beta_{lowdens} X_{lowdens} + \beta_{highedu} X_{highedu} + \epsilon_i \end{aligned} \tag{3.1}$$

Rent model (Model 1):

$$
\begin{aligned}
log(Y) =& \beta_{sqm} log(X_{sqm}) + \beta_{\lesssim 51} X_{\lesssim 51} + \beta_{61} X_{61} + \beta_{71} X_{71} + \beta_{81} X_{81} + \beta_{91} X_{91} + \\
& \beta_{01} X_{01} + \beta_{\gtrsim 11} X_{\gtrsim 11} + \beta_{f1} X_{f1} + \beta_{f2} X_{f2} + \beta_{f3} X_{f3} + \beta_{f4} X_{f4} + \\
& \beta_{f5} X_{f5} + \beta_{\gtrsim f6} X_{\gtrsim f6} + \beta_{parking} X_{parking} + \beta_{openP} X_{openP} + \\
& \beta_{storage} X_{storage} + \beta_{firep} X_{firep} + \beta_{AH} X_{AH} + \beta_{AC} X_{AC} + \beta_{sfh} X_{sfh} + \\
& \beta_{seaview} X_{seaview} + \beta_{front} X_{front} + \beta_{ring} X_{ring} + \beta_{metro} X_{metro} + \\
& \beta_{isap} X_{isap} + \beta_{natl} X_{natl} + \beta_{railway} X_{railway} + \beta_{sub} X_{sub} + \beta_{air} X_{air} + \\
& \beta_{port} log(X_{port}) + \beta_{mar} X_{mar} + \beta_{rroad1} X_{rroad1} + \beta_{rroad2} X_{rroad2} + \\
& \beta_{rroad3} X_{rroad3} + \beta_{cbd} X_{cbd} + \beta_{NS} X_{NS} + \beta_{uni} X_{uni} + \beta_{arch} log(X_{arch}) + \\
& \beta_{coast} X_{coast} + \beta_{lowdens} X_{lowdens} + \beta_{highedu} X_{highedu} + \epsilon_i
\end{aligned} \tag{3.2}
$$

Moreover, the following spatial econometric models are estimated: 4) spatial error model (SEM), 5) spatial lagged model (SAR), 6) spatial Durbin model (SDM), 7) spatial autocorrelation (SAC) (LeSage and Pace, 2009) and 8) geographically weighted regression (GWR) (Brunsdon et al., 1998). The experimentation with models of different types, aims in giving results using the same explanatory variables as the linear-in-the-parameters models –where this was possible– but with superior goodness-of-fit (in terms of AIC and Moran's I). The optimal –in terms of Akaike Information Criterion (AIK) minimization– number of k-nearest neighbors, was determined using a line search approach, concluding that 3 nearest data points are required for the construction of the weight matrix used in the spatial econometric models (SAR, SEM, SDM and SAC). The variance-stabilizing coding scheme (S-coding) (Tiefelsdorf et al., 1999) was used for the generation of the nearest neighbor matrix, and the LU method –for the decomposition of a sparse matrix (LeSage and Pace, 2009)– was used for the estimation. Concerning the GWR models, a stable bandwidth of 4,500m was selected for the purchase model and a radius of 5,000m for the rent, values that maximize R^2 and minimize AIC.

Table 3.1: Specification of variables - real estate price models (Athens)

Variable	Model 1	Model 2	Model 3	Models 4-7 (Spatial)	Name
Dependent					
Price	logarithm	logarithm	logarithm	logarithm	
House Attributes					
m^2	logarithm	logarithm	logarithm	logarithm	X_{sqm}
Before 1951	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	$X_{\lesssim 51}$
1951-1960	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{51}
1961-1970	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{61}
1971-1980	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{71}
1981-1990	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{81}
1991-2000	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{91}
2001-2010	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{01}
>=2011	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	$X_{\gtrsim 11}$
Basement	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{base}
Ground floor	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{gf}
1st floor	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{f1}
2nd floor	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{f2}
3rd floor	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{f3}
4th floor	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{f4}
5th floor	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{f5}
6th floor or more	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	$X_{\gtrsim f6}$
Parking	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	$X_{parking}$
Open parking	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{openP}
Storage place	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	$X_{storage}$
Fireplace	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{firep}
Auto heat	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{AH}
A/C	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{AC}
Single family house	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{sfh}
View: Sea	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	$X_{seaview}$
Orientation: front	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{front}

Table 3.1. Specification of variables - real estate price models (Athens) (continue)

Variable	Model 1	Model 2	Model 3	Models 4-7 (Spatial)	Name
Transport attributes					
In the inner Ring	dummy, 1 if inside	dummy, log(2) if inside	dummy, 1 if inside	dummy, 1 if inside	X_{ring}
Distance from bus station	dummy, 1 if <50m	dummy, log(2) if <50m	p-spline, meters	dummy, 1 if <50m	X_{bus}
Distance from metro station	dummy, 1 if <500m	dummy, log(2) if <500m	p-spline, meters	dummy, 1 if < 500m	X_{metro}
Distance from ISAP station	dummy, 1 if < 500m	dummy, log(2) if <500m	p-spline, meters	dummy, 1 if < 500m	X_{isap}
Distance from Rail station	dummy, 1 if <500m	dummy, log(2) if <500m	p-spline, meters	dummy, 1 if < 500m	X_{natl}
Distance from railway railway	dummy, 1 if <100m	dummy, log(2) if <100m	p-spline, meters	dummy, 1 if < 100m	$X_{railway}$
Distance from suburban rail station	dummy, 1 if <2000m	dummy, log(2) if < 2000m	p-spline, meters	dummy, 1 if < 2000m	X_{sub}
Distance from tram station	dummy, 1 if <500m	dummy, log(2) if <500m	p-spline, meters	dummy, 1 if < 500m	X_{tram}
Distance from airport	dummy, 1 if <7000m	dummy, log(2) if <7000m	p-spline, meters	dummy, 1 if <7000m	X_{air}
Distance from port (Pireaus/Rafina)	logarithm	logarithm	p-spline, meters	logarithm	X_{port}
Distance from marina	dummy, 1 if < 1500m	dummy, log(2) if < 1500m	p-spline, meters	dummy, 1 if < 1500m	X_{mar}
Distance from ring-road (1)	dummy, 1 if 0-250m	dummy, log(2) if 250m		dummy, 1 if 0-0-250m	X_{rroad1}
Distance from ring-road (2)	dummy, 1 if 250-1500m	dummy, log(2) if 250m-1500m	p-spline, meters	dummy, 1 if 250m-1500m	X_{rroad2}
Distance from ring-road (3)	dummy, 1 if >1500m	dummy, log(2) if >1500m		dummy, 1 if >1500m	X_{rroad3}
Other spatial attributes					
Distance from CBD	meters	meters	p-spline, meters	meters	X_{cbd}
Northern suburbs	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{NS}
University area	dummy, 1 if TRUE	dummy, log(2) if TRUE	dummy, 1 if TRUE	dummy, 1 if TRUE	X_{uni}
Distance from archaeological sites	logarithm	logarithm	p-spline, meters	logarithm	X_{arch}
Distance from coastline	meters	meters TRUE	p-spline, meters	meters	X_{coast}
Low population density	dummy, 1 if <500/km^2, zone	dummy, log(2) if <500/km^2, zone	dummy, 1 if <500/km^2, zone	dummy, 1 if <500/km^2, zone	$X_{lowdens}$
High education level	log of % of population with university degree or higher, zone	log of % of population with university degree or higher, zone	log of % of population with university degree or higher, zone	log of % of population with university degree or higher, zone	$X_{highedu}$

In the literature, different distances from transportation infrastructure are encountered in the model variables, depending on the local characteristics of each case. For the purposes of this research, the decision of the distance has been based on the criteria of rationality and significance. A sensitivity analysis has been conducted, trying different distances in a sensible –for each mode– range. A small distance (50m) from the nearest bus station was selected to be used, meaning that it is located almost next to the dwelling, a medium distance (500m) from the urban rail stations (national railway, metro, ISAP and tram), small to medium (100m) distance from the railway –sufficient in order to capture the noise effect– a high distance from the suburban rail stations (2000m) and a very high distance (7000m) from the airport because they are located in low-density areas.

3.2.3 Results and comparison

Tables 3.2 through 3.5 summarize the model estimation results, which are interpreted and discussed in the following subsections.

Property variables

The year of construction is significant for both house purchase and rent models. Dwellings were categorized in eight different classes, based on their year of construction: 1) before 1951, 2) 1951 to 1960, 3) 1961 to 1970, 4) 1971 to 1980, 5) 1981 to 1990, 6) 1991 to 2000, 7) 2001 to 2010 and 8) in 2011 or later. Houses built between in the 50s and 60s have the lowest values, as concluded from the rent and the purchase models respectively. Rent and purchase price increase exponentially as one moves away from these decades on the time-line. Remaining older houses were possibly built during the modern-neoclassicism period that prevailed in Greece between 1940 and 1950. Many of these houses are characterized as“monuments of cultural and architectural heritage” (which may partially justify their value) while dwellings built in the 60s (the beginning of rapid urbanization in Athens) usually lack architectural interest and build quality.

The purchase price increases continuously –but not linearly– with the floor level. Moreover, the availability of a parking space, especially covered (pilotis or underground), a storage room, a fire place, auto heat and A/C, but also a front-orientation and sea-view, affect the purchase price positively. Finally, single family houses are more expensive than semi-detached and apartments.

Transport variables

The impact of externalities on house prices and rents, is evident. The values are negatively affected by the presence of ISAP or national rail station within 500m distance, arguably because of the noise, a finding similar to Brons et al. (2003) and Bowes and Ihlanfeldt (2001). Apart from that, ISAP is the oldest railway in Attica, so the demand for dwellings around its stations had probably reached high levels in the past. However, the continuous decline of its quality of service, the general increase of the citizens' average income during the last two decades –which turned them in using private transportation– and the construction of the new metro lines, have probably diversified the households location choice criteria.

The overall estimate that the prices of dwellings located close to a metro station are more expensive, is verified by the positive coefficients of the variable indicators (presence of a metro station around 500m of the houses) in both purchase and rent models. This finding comes in accordance with the results of Cervero and Landis (1993) and Benjamin and Sirmans (1996) who found positive effect of the Washington DC metro. Bajic (1983) and Bae et al. (2003) in Toronto and Seoul respectively, and Al-Mosaind et al. (1993), Hess and Almeida (2007) and Clower and Weinstein (2002), who measured the effect of light-rail in different cities in the US. The above-ground part of the metro (three stations, one of which is at the airport) was not taken into consideration, because this part is also served by the suburban railway (and these stations resemble more suburban rail stations and less metro stations). Proximity to tram (<500m) or bus stop (<50m) increases the transaction price; however, both variables are insignificant when estimating rents. Since the tram network is relatively new, future comparisons should also be made. In the case of Sheffield's supertram, the importance of proximity to a tram station changed from positive in 1988 to negative in 1993 and then ended to be insignificant in 1996 (Henneberry, 1998).

Dwelling prices and rents around the Athens International Airport (<7000m) are lower, because of the airplanes' noise during take off and landing, a finding that is consistent with many other studies (Nelson, 1980; Collins and Evans, 1996; Nelson, 2004; Cohen and Coughlin, 2008). Furthermore, houses within 1500 meters of the main four marinas of Attica are more expensive, while both prices and rents increase according to the distance from the ports of Piraeus and Rafina. Finally, regarding the impact of the Attica Tollway, three different classes corresponding to three distance ranges, were considered: <250m, 250-1500m and >1500. The results show that the proximity of a dwelling to the highway has a negative impact on its purchase price, while, rents of houses located in the middle-range (250-1500m), are lower.

Other spatial variables

Prices and rents are slightly negatively affected by the distance from CBD and the logarithmic distance from coastline, while the logarithmic distance from archaeological sites has a positive effect. Regions around archaeological sites (excluding the recently restored area around Acropolis), have been left to degrade. Moreover, the model verifies that dwellings located near the Athens Universities campuses (Zografou, Goudi, Ilisia, Kaisariani, Pagrati), have a strong geographical advantage compared to the others. Finally, houses located at the Northern Suburbs of Athens (an area with a higher concentration of green areas and home to a more affluent population) are more expensive, while dwellings located in a zone with high educational level (percentage of people with university degree) or low density (<500 persons per km^2) are also more expensive.

Model comparison

Aggregate tests of goodness of fit of the various models are summarized in Table 3.6. Because of the large number of observations, asymptotic analytical values –rather than the approximated– were used into the numerical Hessian in our SDM models. Otherwise, the regression coefficients would have been wrongly approximated leading in a wrong estimation of λ. However, a significant number of spatial variables (which are often omitted in similar model estimations) is included here. As a result, many estimated coefficients of SDM price and rent models seem to be insignificant. According to LeSage and Pace (2009), despite the fact that the SDM model may result in insignificant coefficient estimations when there are no omitted characteristics, it outperforms the SAC because it includes the spatial lags of the explanatory variables while SAC doesn't. When applying spatial econometric models in large datasets, efficiency of estimated coefficients is not always the case. Tables 3.7 and 3.8 present the approximate significance of smooth terms, the ranges of the estimated coefficients of the GWR models are presented in Tables 3.9 and 3.10, and Figures 3.2 to 3.5 visually present the spatial distribution of the predicted from the spatial models values. The colour indicates the predicted values, while the size of each point in proportional with the residual. The models' fit and "success" in overcoming spatial autocorrelation, is indicated by the AIC and Moran's I measurements respectively.

In spatial econometric models, ρ is the average coefficient of lagged variables and λ is the dependent vector error parameter. Both purchase and rent price SAR models have a ρ value close to 0.2, which indicates the existence of spatial autocorrelation. Spatial autocorrelation is positive when the data points with high absolute values are located at the same area and those with low are also located together, while negative is exhibited when data points with both high and low values are located together (Fotheringham et al., 1998). The global models ignore the possible differences of spatial autocorrelation – exhibited on different points of the same dataset– which can be either positive or negative over different regions of the study area. This differentiation of spatial autocorrelation at a local level is tested by the Moran's I (Fotheringham et al., 1998).

Unlike the linear models and GWR, SAC, SEM and SDM models deal with spatial autocorrelation successfully, resulting in low Moran' s I values. However, GWR gives the smallest AIC, which means that it fits the data better than the other spatial models with the same variables. There cannot be a direct comparison between the GAM models (with p-splines) and the others, because the use of p-splines has reduced the number of variables used, which results in lower AIC. The p-splines are piecewise-defined smooth polynomial functions that are applied to the distance variables in our models. The approximate significance of the p-splines is presented in Tables 3.2, 3.3, 3.4 and 3.5.

The results indicate that spatial econometric models, and especially the SEM, perform better than OLS in two important modelling issues: spatial autocorrelation and prediction. Taking into consideration the short estimation time required, and the relatively easy way of implementation, it is proposes the use of SEM in LUTI models.

Table 3.2: Linear model estimations - purchase price (Athens)

	HPModel 1		**HPModel 2**		**HPModel 3**	
	n=8066		n=8066		n=8066	
Variables						
Intercept	3.954	***	5.822	***	11.615	***
House attributes						
X_{sqm}	1.055	***	1.049	***	p-spline	***
$X_{\lesssim 51}$	0.419	***	0.597	***	0.419	***
X_{51}	reference category for year of construction					
X_{61}	-0.052	*	-0.078	*	-0.075	**
X_{71}	0.046	.	0.057		0.006	
X_{81}	0.155	***	0.216	***	0.122	***
X_{91}	0.264	***	0.370	***	0.232	***
X_{01}	0.354	***	0.501	***	0.368	***
$X_{\gtrsim 11}$	0.458	***	0.654	***	0.456	***
X_{base}	-0.179	***	-0.300	***	-0.177	***
X_{gf}	reference category for floor					
X_{f1}	0.116	***	0.180	***	0.084	***
X_{f2}	0.140	***	0.212	***	0.102	***
X_{f3}	0.161	***	0.250	***	0.122	***
X_{f4}	0.198	***	0.302	***	0.170	***
X_{f5}	0.216	***	0.334	***	0.208	***
$X_{\gtrsim f6}$	0.291	***	0.435	***	0.270	***
$X_{parking}$	0.109	***	0.157	***	0.086	***
X_{openP}	-0.078	***	-0.124	***	-0.036	*
$X_{storage}$	0.045	***	0.067	***	0.040	***
X_{firep}	0.064	***	0.090	***	0.032	***
X_{AH}	0.060	***	0.093	***	0.048	***
X_{AC}	0.047	***	0.064	***	0.033	***
X_{sfh}	0.186	***	0.256	***	0.241	***
$X_{seaview}$	0.119	***	0.092	***	0.080	***
X_{front}	0.044	**	0.058	**	0.036	**
Transport attributes						
X_{ring}	0.379	***	0.510	***	0.291	***
X_{bus}	0.028	**	0.049	**	p-spline	***
X_{metro}	0.101	***	0.151	***	p-spline	*
X_{isap}	-0.133	***	-0.197	***	p-spline	**
X_{natl}	-0.199	***	-0.298	***	p-spline	*
$X_{railway}$	-0.108	**	-0.137	**	p-spline	***
X_{sub}	-0.059	**	-0.298	***	p-spline	***
X_{tram}	0.041	**	0.084	***	p-spline	***
X_{air}	-0.149	***	-0.300	***	p-spline	***
X_{port}	0.242	***	0.249	***	p-spline	***
X_{mar}	0.209	***	0.170	***	p-spline	***
X_{rroad1}	reference category for distance from ring road					
X_{rroad2}	0.102	**	0.139	*		
X_{rroad3}	0.154	***	0.210	***	p-spline	***
Other spatial attributes						
X_{cbd}	-4.78E-06	***	-0.040	***	p-spline	***
X_{NS}	0.451	***	0.542	***	0.309	***
X_{uni}	0.162	***	0.230	***	-0.020	
X_{arch}	0.086	***	0.045	***	p-spline	**
X_{coast}	0.000	***	-0.019		p-spline	***
$X_{lowdens}$	0.057	***	0.102	***	0.051	***
$X_{highedu}$	0.016	*	0.030	***	0.031	***
R^2	0.861		0.864		0.888	

(·) for p $\lesssim 0.1$, (*) for p $\lesssim 0.05$, (**) for p $\lesssim 0.01$, (***) for p $\lesssim 0.001$

Table 3.3: Linear model estimations - rent price (Athens)

	HPModel 1		**HPModel 2**		**HPModel 3**	
	n=8400		n=8400		n=8400	
Variables						
Intercept	4.003	***	5.373	***	8.368	***
House attributes						
X_{sqm}	0.653	***	0.653	***	p-spline	***
$X_{\lesssim 51}$	0.301	***	0.438	***	0.270	***
X_{51}	reference category for year of construction					
X_{61}	0.032	.	0.044	.	0.008	*
X_{71}	0.054	***	0.080	***	0.032	***
X_{81}	0.079	***	0.117	***	0.063	***
X_{91}	0.176	***	0.259	***	0.138	***
X_{01}	0.249	***	0.359	***	0.224	***
$X_{\gtrsim 11}$	0.330	***	0.482	***	0.318	***
X_{base}, X_{gf}	reference category for floor					
X_{f1}	0.030	***	0.042	***	0.060	***
X_{f2}	0.033	***	0.048	***	0.067	***
X_{f3}	0.059	***	0.087	***	0.088	***
X_{f4}	0.077	***	0.113	***	0.110	***
X_{f5}	0.092	***	0.134	***	0.127	***
$X_{\gtrsim f6}$	0.117	***	0.170	***	0.150	***
$X_{parking}$	0.097	***	0.136	***	0.079	***
X_{openP}	0.097	***	-0.140	***	-0.051	**
$X_{storage}$	0.034	***	0.053	***	0.029	***
X_{firep}	0.098	***	0.147	***	0.063	***
X_{AH}	0.012	.	0.018	.	0.036	***
X_{AC}	0.084	***	0.119	***	0.056	***
X_{sfh}	0.078	***	0.110	***	0.043	***
$X_{seaview}$	0.119	***	0.110	***	0.052	***
X_{front}	0.025	***	0.034	**	-	
Transport attributes						
X_{ring}	0.188	***	0.233	***	0.119	***
X_{bus}	-		-		-	
X_{metro}	0.057	***	0.089	***	p-spline	**
X_{isap}	-0.023	***	-0.041	***	p-spline	***
X_{natl}	-0.126	***	-0.213	***	p-spline	***
$X_{railway}$	-0.054	*	-0.086	***	p-spline	***
X_{sub}	-0.085	***	-0.194	***	p-spline	***
X_{tram}	-		-		p-spline	***
X_{air}	-0.139	***	-0.259	***	p-spline	***
X_{port}	0.175	***	0.168	***	p-spline	***
X_{mar}	0.184	***	0.155	***	p-spline	***
X_{rroad1}, X_{rroad3}	reference category for distance from ring road					
X_{rroad2}	-0.052	***	-			
Other spatial attributes						
X_{cbd}	-3.94E-06	***	-0.035	***	p-spline	***
X_{NS}	0.381	***	0.458	***	0.257	***
X_{uni}	0.094	***	0.118	***	-	
X_{arch}	0.016	**	-0.011	*	-	
X_{coast}	0.000	***	-0.094	*	-	
$X_{lowdens}$	0.011	.	0.030	**	-	
$X_{highedu}$	-		-		0.021	
R^2	0.831		0.831		0.880	

(·) for $p \lesssim 0.1$, (*) for $p \lesssim 0.05$, (**) for $p \lesssim 0.01$, (***) for $p \lesssim 0.001$

Table 3.4: Spatial econometric model estimations - purchase price (Athens)

	SEM		SAR		SAC	
	n=8066		n=8066		n=8066	
Variables						
Intercept	4.108	***	2.740	***	3.603	***
House attributes						
m^2	1.022	***	1.005	***	1.027	***
$X_{\lesssim 51}$	0.424	***	0.382	***	0.419	***
X_{51}	reference category for year of construction					
X_{61}	-0.030		-0.043		-0.033	
X_{71}	0.027		0.036		0.027	
X_{81}	0.150	***	0.143	***	0.149	***
X_{91}	0.260	***	0.243	***	0.258	***
X_{01}	0.379	***	0.334	***	0.374	***
$X_{\gtrsim 11}$	0.466	***	0.424	***	0.462	***
X_{base}	-0.168	***	-0.187	***	-0.176	***
X_{gf}	reference category for floor					
X_{f1}	0.103	***	0.123	***	0.107	***
X_{f2}	0.125	***	0.142	***	0.129	***
X_{f3}	0.146	***	0.172	***	0.152	***
X_{f4}	0.201	***	0.216	***	0.207	***
X_{f5}	0.238	***	0.239	***	0.242	***
$X_{\gtrsim f6}$	0.298	***	0.303	***	0.302	***
$X_{parking}$	0.093	***	0.093	***	0.093	***
X_{openP}	-0.044	**	-0.061	***	-0.046	***
$X_{storage}$	0.044	***	0.042	***	0.044	***
X_{firep}	0.041	***	0.053	***	0.042	***
X_{AH}	0.052	***	0.050	***	0.052	***
X_{AC}	0.041	***	0.044	***	0.042	***
X_{sfh}	0.203	***	0.178	***	0.201	***
$X_{seaview}$	0.074	***	0.098	***	0.076	***
X_{front}	0.020	***	0.039	**	0.022	*
Transport attributes						
X_{ring}	0.348	***	0.304	***	0.333	***
X_{bus}	0.028	**	0.029	*	0.029	**
X_{metro}	0.065	***	0.089	***	0.071	***
X_{isap}	-0.131	***	-0.087	***	-0.116	***
X_{natl}	-0.167	***	-0.160	***	-0.161	***
$X_{railway}$	-0.068	.	-0.111	***	-0.077	***
X_{sub}	-0.051	.	-0.033	.	-0.045	.
X_{tram}	0.038	.	0.019		0.031	
X_{air}	-0.130	***	-0.100	***	-0.118	***
X_{port}	0.254	***	0.177	***	0.227	***
X_{mar}	0.208	***	0.171	***	0.195	***
X_{rroad1}	reference category for distance from ring road					
X_{rroad2}	0.033		0.091	**	0.042	
X_{rroad3}	0.033		0.129	***	0.084	*
Other spatial attributes						
X_{cbd}	-4.51E-06	***	-6.36E-06	***	-5.66E-06	***
X_{NS}	0.457	***	0.300	***	0.401	***
X_{uni}	0.087	*	0.191	***	0.119	**
X_{arch}	0.087	***	0.057	***	0.076	***
X_{coast}	0.000	***	-2.75E-05	***	0.000	***
$X_{lowdens}$	0.057	***	0.032	***	0.046	***
$X_{highedu}$	0.019	**	0.010	**	0.016	**
lambda	0.429	***	-	-	0.368	***
rho	-	-	0.189	***	0.066	***

(·) for p $\lesssim 0.1$, (*) for p $\lesssim 0.05$, (**) for p $\lesssim 0.01$, (***) for p $\lesssim 0.001$

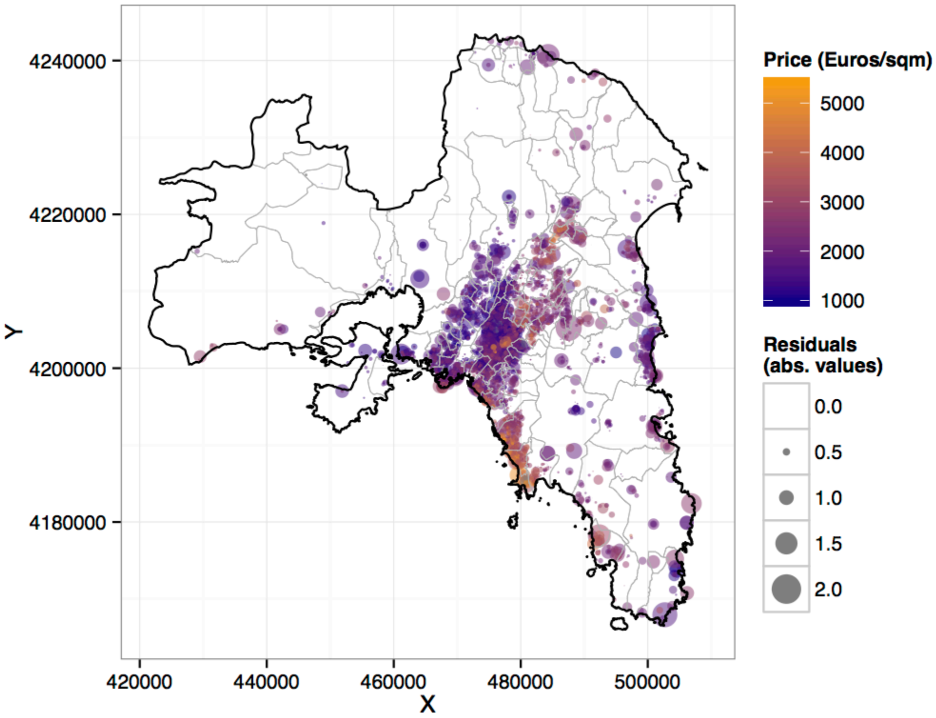
Table 3.5: Spatial econometric model estimations - rent price (Athens)

	SEM		SAR		SAC	
	n=8400		n=8400		n=8400	
Variables						
Intercept	4.020	***	2.977	***	3.559	***
House attributes						
X_{sqm}	0.627	***	0.621	***	0.633	***
$X_{\lesssim 51}$	0.307	**	0.296	**	0.307	**
X_{51}	reference category for year of construction					
X_{61}	0.035	***	0.041	***	0.038	***
X_{71}	0.062	***	0.065	***	0.065	***
X_{81}	0.087	***	0.085	***	0.089	***
X_{91}	0.186	***	0.175	***	0.186	***
X_{01}	0.266	***	0.254	***	0.267	***
$X_{\gtrsim 11}$	0.347	***	0.323	***	0.346	***
X_{base}, X_{gf}	reference category for floor					
X_{f1}	0.039	***	0.038	***	0.040	***
X_{f2}	0.046	***	0.046	***	0.048	***
X_{f3}	0.070	***	0.072	***	0.072	***
X_{f4}	0.091	***	0.094	***	0.094	***
X_{f5}	0.107	***	0.110	***	0.111	***
$X_{\gtrsim f6}$	0.143	***	0.128	***	0.143	***
$X_{parking}$	0.085	***	0.084	***	0.085	***
X_{openP}	-0.073	***	-0.077	***	-0.074	***
$X_{storage}$	0.033	***	0.032	***	0.033	***
X_{firep}	0.076	***	0.081	***	0.077	***
X_{AH}	0.020	***	0.010	***	0.018	***
X_{AC}	0.063	***	0.078	***	0.067	***
X_{sfh}	0.084	***	0.083	***	0.084	***
$X_{seaview}$	0.081	***	0.104	***	0.087	***
X_{front}	0.015	**	0.020	**	0.015	**
Transport attributes						
X_{ring}	0.187	***	0.143	***	0.160	***
X_{bus}	-		-		-	
X_{metro}	0.041	***	0.048	***	0.043	***
X_{isap}	-0.017		-0.014		-0.014	
X_{natl}	-0.100	***	-0.098	***	-0.100	***
$X_{railway}$	-0.054	**	-0.052	**	-0.054	**
X_{sub}	-0.069	***	-0.062	***	-0.064	***
X_{tram}	-		-		-	
X_{air}	-0.115	***	-0.088	***	-0.099	***
X_{port}	0.185	***	0.115	***	0.149	***
X_{mar}	0.179	***	0.145	***	0.167	***
X_{rroad1}, X_{rroad3}	reference category for distance form ring road					
X_{rroad2}	-0.067	***	-0.037	***	-0.054	***
Other spatial attributes						
X_{cbd}	-3.94E-06	***	-5.76E-06	***	-4.35E-06	***
X_{NS}	0.391	***	0.251	***	0.320	***
X_{uni}	0.074	***	0.097	***	0.085	***
X_{arch}	0.016	**	-	-	-	-
X_{coast}	0.000	**	0.000	**	0.000	**
$X_{lowdens}$	0.018	***	0.005	***	0.013	***
$X_{highedu}$	-		-		-	
lambda	0.392	***	-	-	0.290	***
rho	-	-	0.211	***	0.102	***

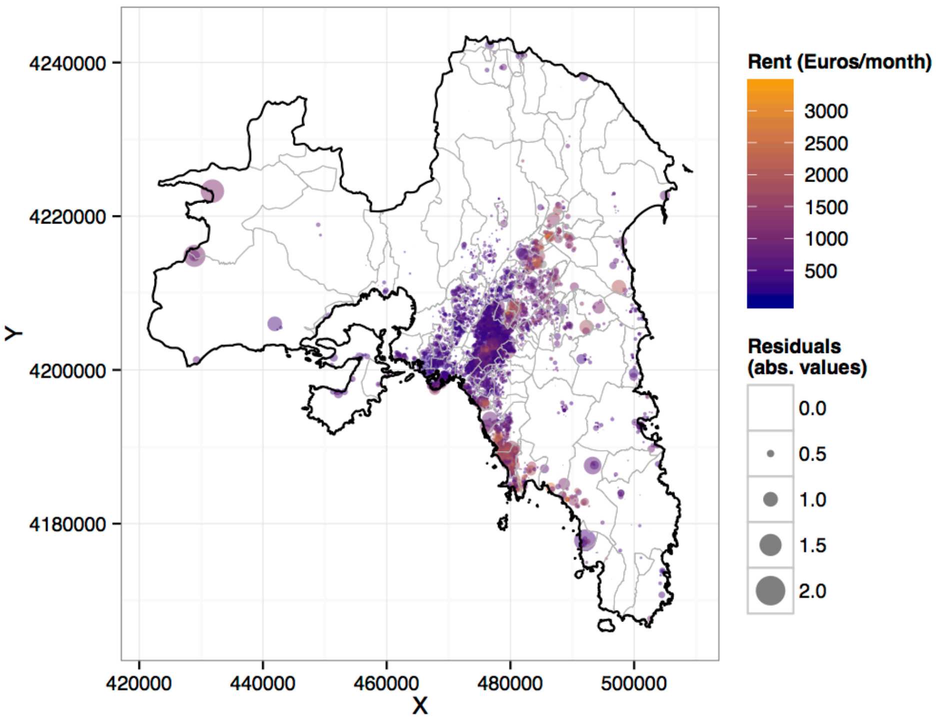
(·) for p $\lesssim 0.1$, (*) for p $\lesssim 0.05$, (**) for p $\lesssim 0.01$, (***) for p $\lesssim 0.001$

Table 3.6: Model estimations comparison – aggregate goodness of fit measures (Athens)

		semi-log	log-log	P-Splines	SARerr	SARlag	SAC	SARmix	GWR
Price	n	8066	8066	8066	8066	8066	8066	8066	8066
	R^2	0.861	0.864	0.888	-	-	-	-	0.902
	AIC	4740.702	4610.111	3093.475	3533.100	4012.200	3496.700	3284.900	2917.276
	Moran's I	0.280	0.274	0.150	0.001	0.180	0.003	-0.001	0.237
	λ	-	-	-	0.429	-	0.368	-	-
	ρ	-	-	-	-	0.189	0.066	0.405	-
Rent	n	8400	8400	8400	8400	8400	8400	8400	8400
	R^2	0.831	0.831	0.880	-	-	-	-	0.869
	AIC	-1685.943	-1678.807	-4466.307	-2682.200	-2475.400	-2750.100	-2940.900	-3631.150
	Moran's I	0.242	0.245	0.105	-0.001	0.129	0.005	0.003	0.239
	λ	-	-	-	0.392	-	0.290	-	-
	ρ	-	-	-	-	0.211	0.102	0.360	-

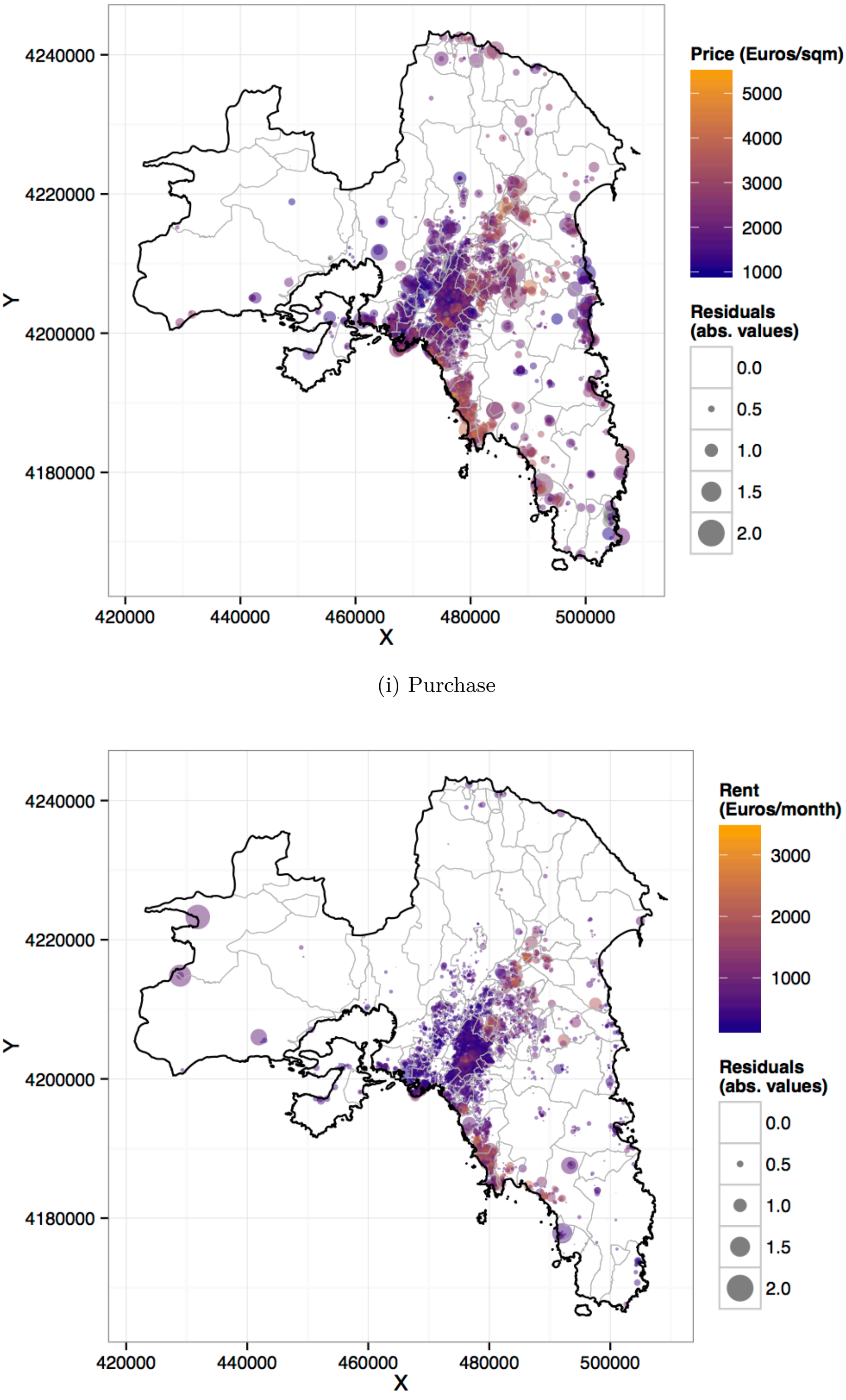


(i) Purchase



(ii) Rent

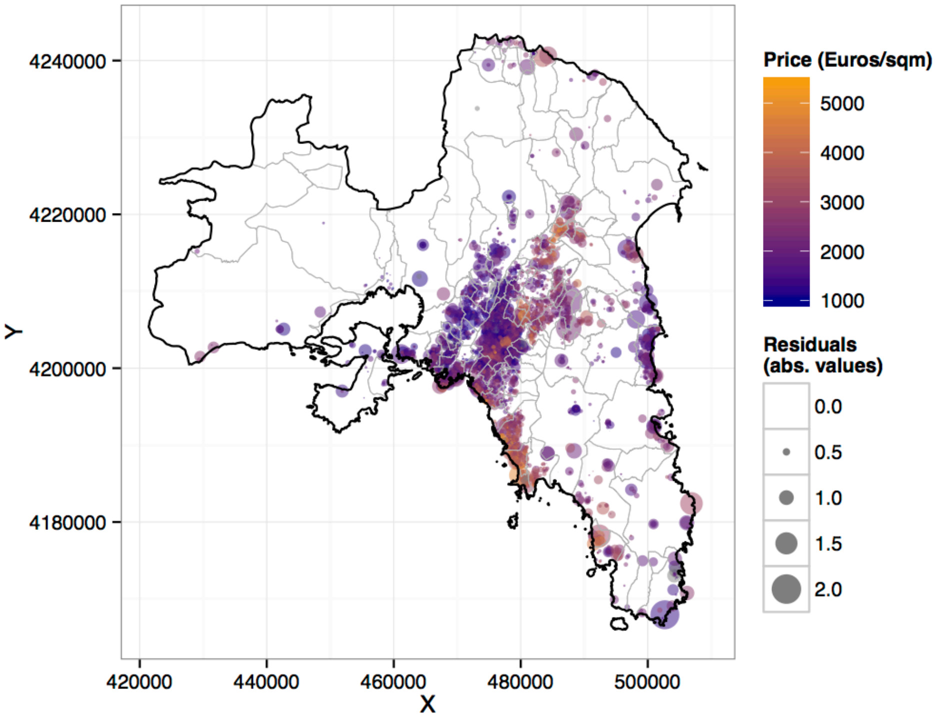
Figure 3.2: SEM results (Athens)



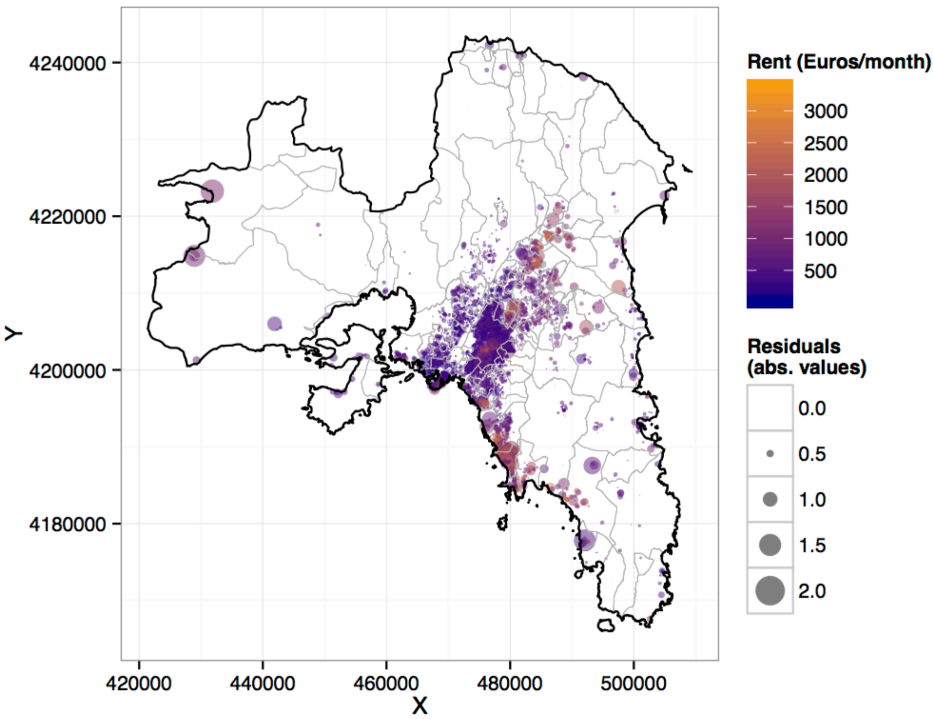
(i) Purchase

(ii) Rent

Figure 3.3: SAR results (Athens)

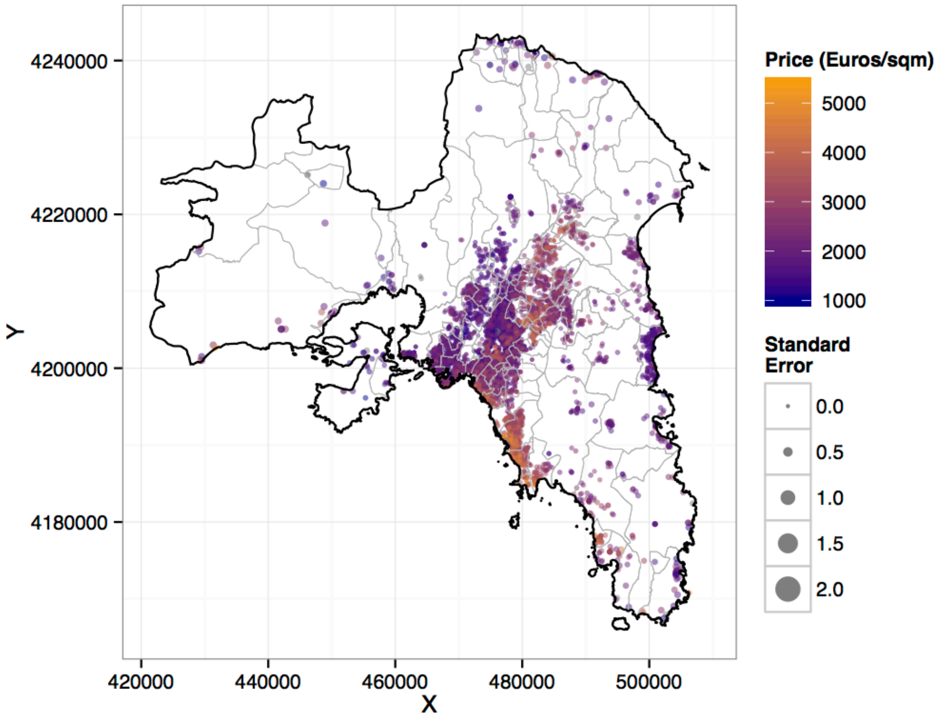


(i) Purchase

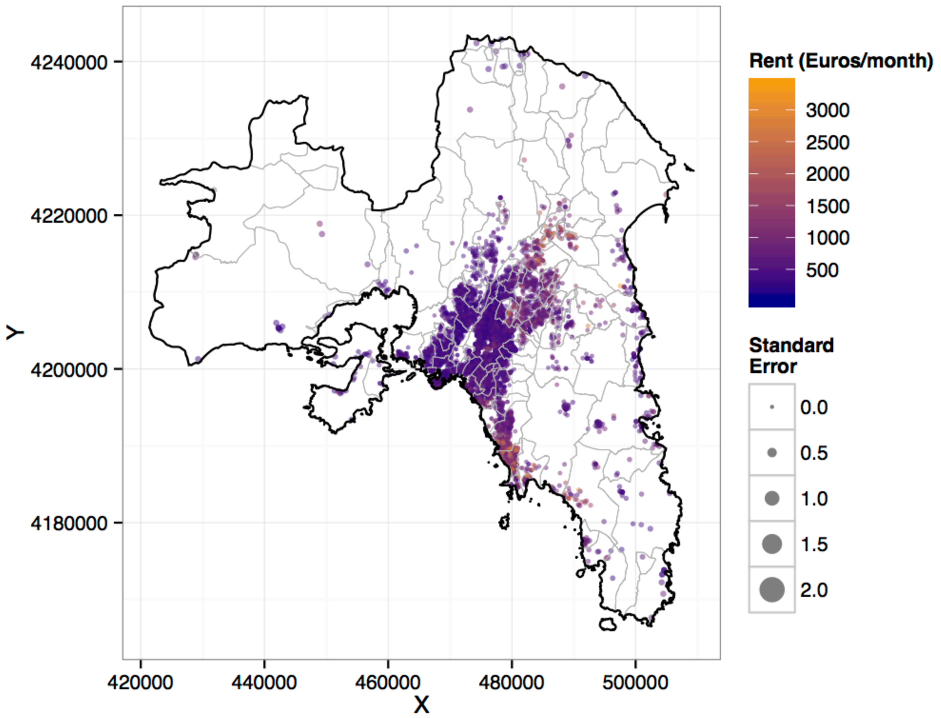


(ii) Rent

Figure 3.4: SDM results (Athens)



(i) Purchase



(ii) Rent

Figure 3.5: GWR model results (Athens)

In order to compute the actual impact –measured by the variables of spatial econometrics with lagged response variable (SAR and SAC)– on real estate prices, the equations had to be transformed to their "reduced" form. In this case, eq. 2.4 (SAR) becomes:

$$Y = [I - \rho W]^{-1} X\beta + [I - \rho W]^{-1} \varepsilon \tag{3.3}$$

and eq. 2.6 (SAC):

$$Y = [I - \rho W]^{-1} X\beta + [I - \rho W]^{-1} u \tag{3.4}$$

where $[I - \rho W]^{-1}$ represents a $n \times n$ inverse matrix.

If x (an element of X) is a continuous variable, then its marginal effect (computed by the *spatial multiplier*) is:

$$\left(\frac{1}{1-\rho}\right) b \tag{3.5}$$

Won Kim et al. (2003) and Steimetz (2010) clarify the use of spatial multipliers in regression analysis.

In this research, the elements of Y are natural logarithms of prices and X is a dummy variable. As a result, the percentage change in price is:

$$\left(e^{\frac{1}{1-\rho} b}\right) \cdot 100 \tag{3.6}$$

Table 3.11 outlines the minimum and maximum percentage impact of main house and spatial characteristics on purchase prices and rents. Houses located 500m around the metro stations have higher purchase prices on average between 6.74% and 11.66% and rental price from 4.20% to 6.21%, while those located 500m around ISAP stations have sale prices from -10.20% to -12.24% but the effect is significantly smaller on rental (from -1.55% to -1.73%). This shows that people are less willing to purchase a house –than to rent one– next to ISAP stations.

Table 3.7: Approximate significance of smooth terms - purchase (Athens)

Variable	**edf**	**Ref.df**	**F**	
X_{sqm}	8.076	8.396	2094.156	***
X_{ring}	7.847	8.333	12.303	***
X_{bus}	1.000	1.000	776.721	***
X_{metro}	5.748	6.289	2.255	*
X_{isap}	8.008	8.227	3.199	**
X_{natl}	3.934	4.615	2.833	*
$X_{railway}$	9.000	9.000	8.214	***
X_{sub}	8.210	8.589	8.986	***
X_{tram}	8.107	8.238	5.224	***
X_{air}	7.370	7.638	7.676	***
X_{port}	8.799	8.955	4.649	***
X_{mar}	5.861	6.722	4.208	***
$X_{rroad1,2,3}$	3.233	3.987	6.926	***
X_{cbd}	7.031	7.360	7.090	***
X_{arch}	5.292	6.021	3.076	**
X_{coast}	7.928	8.374	4.800	***
GCV score	0.085921			
Scale est.	0.084489			
Deviance explained	89			

edf: Estimated degrees of freedom for each model parameter
Ref.df: Reference degrees of freedom for each model parameter

(·) for p $\lesssim 0.1$, (*) for p $\lesssim 0.05$, (**) for p $\lesssim 0.01$, (***) for p $\lesssim 0.001$

Table 3.8: Approximate significance of smooth terms - rent (Athens)

Variable	edf	Ref.df	F	
X_{sqm}	8.968	8.000	2213.305	***
X_{cbd}	5.292	5.765	1.231	
X_{ring}	7.847	8.333	12.303	***
X_{metro}	2.889	3.549	4.081	**
X_{isap}	8.795	8.884	8.880	**
X_{natl}	3.133	3.727	5.328	*
$X_{railway}$	9.000	9.000	3.456	***
X_{sub}	8.970	8.994	11.653	***
X_{tram}	5.304	5.887	3.874	***
X_{air}	6.722	7.016	13.674	***
X_{port}	4.741	5.526	2.678	*
X_{mar}	7.676	7.987	2.678	***
$X_{rroad1,2,3}$	8.418	8.727	9.522	***
X_{arch}	8.249	8.419	8.199	***
X_{coast}	7.041	7.711	7.403	***
GCV score	0.034404			
Scale est.	0.033894			
Deviance explained	88.2			

edf: Estimated degrees of freedom for each model parameter
Ref.df: Reference degrees of freedom for each model parameter

(·) for p $\lesssim 0.1$, (*) for p $\lesssim 0.05$, (**) for p $\lesssim 0.01$, (***) for p $\lesssim 0.001$

Table 3.9: Ranges of estimated coefficients for GWR - purchase price (Athens)

Variables	Min.	1st Qu.	Median	3rd Qu.	Max
Intercept	-7.181	3.540	4.192	5.007	124.700
House attributes					
m^2	0.339	1.051	1.090	1.105	1.330
$X_{\lesssim 51}$	-1.124	0.396	0.444	0.465	2.435
X_{51}	reference category for year of construction				
X_{61}	-3.040	-0.084	-0.064	-0.038	2.037
X_{71}	-1.772	-0.021	0.009	0.059	1.367
X_{81}	-1.641	0.116	0.128	0.151	1.311
X_{91}	-0.857	0.217	0.244	0.268	2.787
X_{01}	-1.447	0.358	0.385	0.423	1.614
$X_{\gtrsim 11}$	-1.873	0.434	0.469	0.527	1.750
X_{base}	-1.212	-0.280	-0.259	-0.198	0.994
X_{gf}	reference category for floor				
X_{f1}	-0.707	0.078	0.108	0.135	0.291
X_{f2}	-0.934	0.107	0.133	0.162	0.333
X_{f3}	-4.201	0.110	0.157	0.191	0.765
X_{f4}	-2.934	0.163	0.201	0.225	0.943
X_{f5}	-2.298	0.184	0.232	0.269	0.836
$X_{\gtrsim f6}$	-5.137	0.244	0.308	0.355	1.504
$X_{parking}$	-0.501	0.076	0.092	0.106	1.004
X_{openP}	-1.171	-0.041	-0.026	-0.012	-0.200
$X_{storage}$	-0.463	0.009	0.017	0.027	0.848
X_{firep}	-0.137	0.038	0.048	0.054	0.609
X_{AH}	-0.242	0.048	0.059	0.066	0.350
X_{AC}	-0.251	0.030	0.044	0.048	0.649
X_{sfh}	-0.484	0.172	0.227	0.267	0.625
$X_{seaview}$	-0.922	0.085	0.110	0.187	0.455
X_{front}	-1.722	0.038	0.043	0.047	1.174
Transport attributes					
X_{ring}	-14.270	0.283	0.359	0.401	1.246
X_{bus}	-0.290	0.022	0.026	0.033	1.755
X_{metro}	-8.921	0.062	0.099	0.136	2.265
X_{isap}	-2.139	-0.135	-0.100	-0.051	1.802
X_{natl}	-1.522	-0.211	-0.195	-0.178	0.261
$X_{railway}$	-1.244	-0.104	-0.088	-0.069	0.329
X_{sub}	-4.270	-0.176	-0.101	-0.030	1.441
X_{tram}	-3.117	0.000	0.031	0.081	1.846
X_{air}	-12.560	-0.167	-0.028	0.094	0.818
X_{port}	-12.660	0.172	0.257	0.320	1.771
X_{mar}	-43.130	0.197	0.233	0.301	1.449
X_{rroad1}	reference category for distance from ring road				
X_{rroad2}	-1.053	0.063	0.121	0.191	8.861
X_{rroad3}	-0.432	0.106	0.136	0.178	13.800
Other spatial attributes					
X_{cbd}	-2.41E-04	-5.59E-06	6.85E-06	2.12E-05	5.97E-05
X_{NS}	-4.388	0.414	0.453	0.578	0.833
X_{uni}	-12.660	0.127	0.162	0.215	0.749
X_{arch}	-0.375	-0.011	0.003	0.033	0.812
X_{coast}	-4.65E-04	-4.12E-05	-3.42E-05	-4.06E-05	2.43E-04
$X_{lowdens}$	-2.184	0.002	0.037	0.087	0.548
$X_{highedu}$	-1.207	-0.005	0.033	0.051	0.407

Table 3.10: Ranges of estimated coefficients for GWR - rent price (Athens)

Variables	Min.	1st Qu.	Median	3rd Qu.	Max
Intercept	-4.009	4.237	4.479	4.710	41.510
House attributes					
m^2	-1.091	0.594	0.607	0.624	1.036
$X_{\lesssim 51}$	-1.539	0.254	0.265	0.290	10.950
X_{51}	reference category for year of construction				
X_{61}	-0.751	0.012	0.014	0.024	9.411
X_{71}	-0.689	0.031	0.035	0.047	10.480
X_{81}	-0.696	0.056	0.061	0.072	9.700
X_{91}	-0.432	0.139	0.147	0.156	8.505
X_{01}	-0.485	0.227	0.231	0.238	10.680
$X_{\gtrsim 11}$	-0.273	0.266	0.279	0.316	11.090
X_{base}, X_{gf}	reference category for floor				
X_{f1}	-0.621	0.046	0.056	0.058	0.887
X_{f2}	-0.453	0.053	0.065	0.067	0.291
X_{f3}	-0.538	0.073	0.084	0.087	0.964
X_{f4}	-3.915	0.095	0.103	0.106	0.365
X_{f5}	-3.633	0.101	0.122	0.135	1.013
$X_{\gtrsim f6}$	-2.409	0.128	0.142	0.153	0.763
$X_{parking}$	-0.118	0.086	0.092	0.097	0.756
X_{openP}	-1.699	-0.053	-0.042	-0.033	0.137
$X_{storage}$	-0.307	0.024	0.031	0.039	1.441
X_{firep}	-0.650	0.085	0.106	0.118	1.823
X_{AH}	-1.636	0.012	0.016	0.022	1.150
X_{AC}	-1.577	0.066	0.069	0.073	0.272
X_{sfh}	-0.767	0.068	0.103	0.124	1.723
$X_{seaview}$	-0.383	0.013	0.013	0.015	1.247
X_{front}	-0.380	0.038	0.043	0.047	1.174
Transport attributes					
X_{ring}	-11.010	0.174	0.183	0.193	1.360
X_{bus}	-	-	-	-	-
X_{metro}	-7.396	0.053	0.065	0.071	0.367
X_{isap}	-0.778	-0.036	-0.028	-0.014	0.682
X_{natl}	-1.682	-0.135	-0.127	-0.115	0.400
$X_{railway}$	-0.625	-0.067	-0.058	-0.040	0.987
X_{sub}	-2.987	-0.121	-0.094	-0.079	0.541
X_{tram}	-	-	-	-	-
X_{air}	-7.299	-0.035	0.013	0.061	3.081
X_{port}	-2.838	0.136	0.155	0.177	1.359
X_{mar}	-18.590	0.156	0.184	0.209	2.581
X_{rroad1}, X_{rroad3}	reference category for distance from ring road				
X_{rroad2}	-3.880	-0.057	-0.042	-0.013	0.258
Other spatial attributes					
X_{cbd}	-3.88E-04	-5.16E-06	4.99E-07	3.08E-06	4.73E-05
X_{NS}	-2.636	0.358	0.383	0.427	1.487
X_{uni}	-10.490	0.077	0.084	0.094	1.138
X_{arch}	-1.239	-0.012	-0.007	-0.003	0.228
X_{coast}	-4.77E-04	-4.16E-05	-1.96E-05	-1.088E-05	4.20E-05
$X_{lowdens}$	-1.778	-0.001	0.015	0.027	0.167
$X_{highedu}$	-	-	-	-	-

Table 3.11: Selected coefficient ranges (Athens)

Variable	Price				Rent			
	Minimum change		Maximum change		Minimum change		Maximum change	
	Change (%)	Model	Change (%)	Model	Change (%)	Model	Change (%)	Model
Constructed after 2010	59.39	SEM	68.76	SAR	41.46	SEM	50.51	SAR
6th floor or more	34.69	SEM	45.35	SAR	15.36	SEM	17.63	SAR
Single family house	22.55	SEM	24.49	SAR	8.73	SEM	11.11	SAR
Sea view	7.71	SEM	12.84	SAR	8.42	SEM	14.05	SAR
Inner ring	41.65	SEM	45.55	SAR	19.80	SAR	20.56	SEM
Bus station in 50m	2.87	SEM	3.65	SAR	-	-	-	-
Metro station in 500m	6.74	SEM	11.66	SAR	4.20	SEM	6.21	SAR
ISAP station in 500m	-10.20	SAR	-12.24	SEM	-1.55	SAC	-1.73	SEM
Rail station in 500m	-15.35	SEM	-17.94	SAR	-9.50	SEM	-11.70	SAR
Railway in 100m	-6.60	SEM	-12.80	SAR	-5.30	SEM	-6.37	SAR
Suburban rail station in 2000m	-3.97	SAR	-4.95	SEM	-6.69	SEM	-7.53	SAR
Tram station in 500m	2.38	SAR	3.82	SEM	-	-	-	-
Airport in 7000m	-11.57	SAR	-12.18	SEM	-10.47	SAC	-10.87	SEM
Marina in 1500m	23.07	SEM	23.45	SAR	19.65	SEM	20.43	SAC
Northern suburbs	44.71	SAR	57.94	SEM	37.47	SAR	47.84	SEM
University area	9.07	SEM	26.52	SAR	7.68	SEM	13.08	SAR

3.2.4 Conclusions

Transport infrastructure is a key determinant of land use evolution. Real-estate prices are a significant measure reflecting these changes. In this research, it is examined how transport infrastructure and policies affect the dwelling purchase prices and rents, using data from Athens, Greece. Different linear regression and spatial econometric models were estimated: 1) Semi logarithmic models, where the response variable is the natural logarithm of price -or rent- and the explanatory are either continuous, dummies or natural logarithms; 2) log-log models that use both explanatory and dependent variables as logarithms; 3) generalized additive models (GAM) with penalized least squares (p-splines); 4) spatial error model (SEM); 5) spatial lagged model (SAR); 6) spatial Durbin model (SDM); 7) spatial autocorrelation (SAC) (LeSage and Pace, 2009) and 8) geographically weighted regression (GWR) (Brunsdon et al., 1998).

An effort was made to examine exhaustively and extensively the various models suggested in the literature for house price modeling. The estimations suggest that prices and rents are –either positively or negatively– affected by transportation infrastructure, so this spatial information should always be taken into consideration when estimating price models. Similarly, when housing and urban development policies are planned to be implemented, they should be tightly connected to the transport infrastructure plans and strategies. This is not a novel finding per se, but this research contributes in quantifying these parameters. The obtained results are consistent with expectations and the literature. While these results cannot be assumed directly transferable to other locations and areas, they could be cautiously used as indications for planning applications (where dedicated models have not been developed); for instance, is cities with similar demographic and economic characteristics, public transport network coverage and demand.

The transferability of the findings in other areas forms an interesting direction for future research. In this research, several findings have been identified, including the relative performance of various models, ranges of parameters for these models and estimated values of model coefficients. Understanding which of these findings are robust to different areas, as well as over time, would be an interesting addition to the literature in the field.

Furthermore, the unstable financial situation and the continuous depreciation of the house prices and rents, in parallel with the need for residents to decrease their commuting costs –leading to the increase of the public transport demand and the expansion of the rail networks– creates fertile field for research in this area. Comparative studies could shed some light on the reaction of households about the choice of residential location and the elasticities of house prices and rents, during historic events such as a long–term recession and a default.

3.3 Real estate prices in Thessaloniki

3.3.1 Transportation infrastructure in Thessaloniki

Thessaloniki is located in Northern Greece and is the second biggest city of the country (also called co-capital of Greece). The population of the region is 819,770 inhabitants (Hellenic Statistical Authority, 2011). Currently, the only mode of public transport in the city is the buses, which are operated by the Urban Transport Organization of Thessaloniki (OASTH). The company owns 604 vehicles that serve 75 routes and move 180,000,000 passengers per year, around the metro area (OASTH, 2013). The construction of a metro line has began in 2006 and is planned to be completed in 2014; however, the project is being delayed and is not expected to be completed before a few more years, at least. The first line will have a length of 9.5km, connecting 13 stations and will serve 250,000 passengers per day around the metro area. The network is planned to be expanded to 3 lines until 2020 (AMETRO, 2013). Moreover, the transportation infrastructure of the city includes: 1) a suburban railway, that connects Thessaloniki with Larisa; 2) the Macedonia International Airport, serving national and international flights; 3) the port of Thessaloniki, which connects the city with the northern Aegean islands, and 4) a railway station of the Hellenic Railways Organization, that links the city with other Greek municipalities. Figure 3.6 presents a map of the study area, along with the main points of transportation infrastructure and the CBD.

The impact of the new metro line construction is a very timely subject in the Greek transportation research community. Until now, three research studies attempted to measure the effects on the commercial properties' prices: Roukouni et al. (2012), analyzed the results of a stated preference survey, and tried to quantify the expected demand of the Papafi metro station. Moreover, they surveyed people's perception on how they expect that the station will affect the market demand of the local enterprises. The results show that about 85% believe that the companies will benefit. Karanikolas and Anastasiadou (2012) examined the effect of the Venizelou metro station in 250 meters –or what can be reached within 5min with a 3km/h walk speed– around the station. The researchers begun their study by interviewing real estate agents, in order to gain a first, general image of the current situation. Then, they questioned 52 shop owners about the rents of the area, and their opinion on the effects of the metro station on their revenues. Their results show that the 83% of the owners believe that there will be positive effects from the new metro, while 73% of the respondents believe that they are in a difficult economic situation not only because of the current economic recession, but also because of the construction works that take place in the area. In another study, Xifilidou et al. (2012), examined the impact of three metro stations (Venizelou, Agia Sofia and Flemigk) on commercial prices, using the same dataset as in Karanikolas and Anastasiadou (2012), and they estimated a hedonic model with a sample of 40 measurements, using only one explanatory variable namely the "negotiation of rent".

The aforementioned studies give a first indication of the expected effects of the new metro line. However, they use a limited amount of data and focus on a small number of stations. The scope of the analysis presented in this research is different. A sound methodology for web-based data collection is first presented, while the use of other open data sources is also investigated. The models that are applied are complicated and computationally very demanding, giving useful results for the transportation research and policy communities.

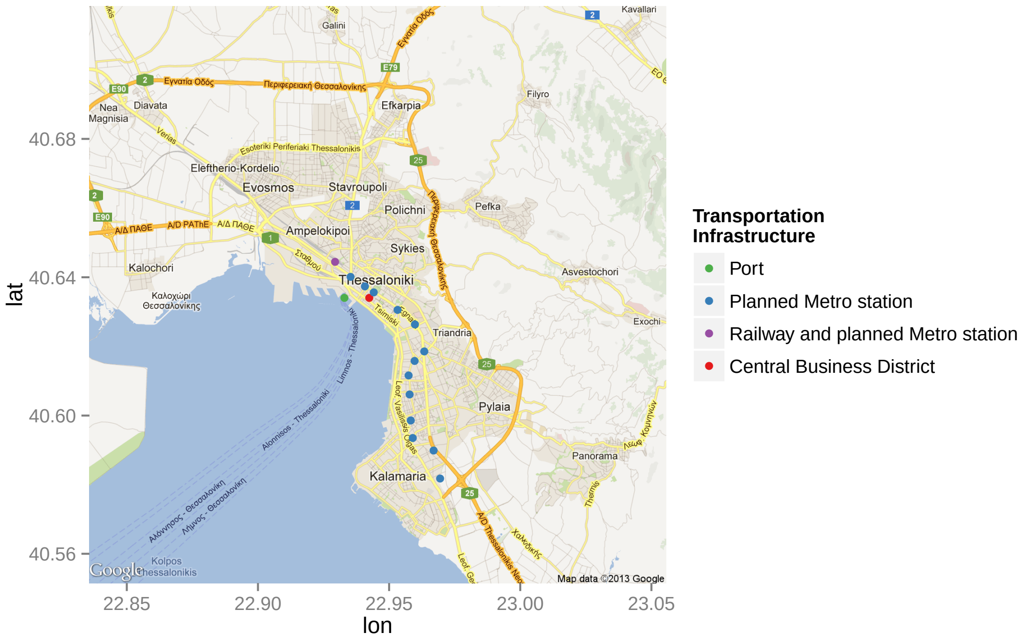


Figure 3.6: Central transport network in Thessaloniki (the metro stations are under construction)

3.3.2 Model development

Similar to Section 3.2, the aim is to model the effect of the transportation infrastructure on dwellings' prices and rents. The dataset used in this research is described in Chapter 6.3.1 of this dissertation. Figure 3.7 presents some exploratory statistics of the purchase and rent datasets.

In the context of this research, following the findings of Section 3.2, it was decided to focus on the spatial error model (SEM), spatial autoregressive model (SAR) and spatial autocorrelation model (SAC), due to their better performance. Moreover, a linear regression using a Box-Cox transformation Box and Cox (1946) of the dependent variable is estimated, which has been proved to stabilize the variance and result in better correlation between the response and explanatory variables (Brennan et al., 1984; Farooq et al., 2010).

The response variable in the developed models is the logarithm of price, and explanatory variables are structural attributes, distances to main transport infrastructure, environmental and demographic indicators (Table 3.12). Despite the fact that the main interest of this research is about the resulted coefficients of the transportation infrastructure related attributes, the inclusion of structural and demographic characteristics when modeling real state prices is important. It should be noted that at first, the potential effect of the –under construction– metro line in not investigated. The final model specifications for the purchase and rent models are presented next.

Before selecting the following specification, a model using all the available data as explanatory variables was initially estimated, and then statistically insignificant variables were eliminated gradually, or grouped into categories (e.g. the year of construction).

Purchase model:

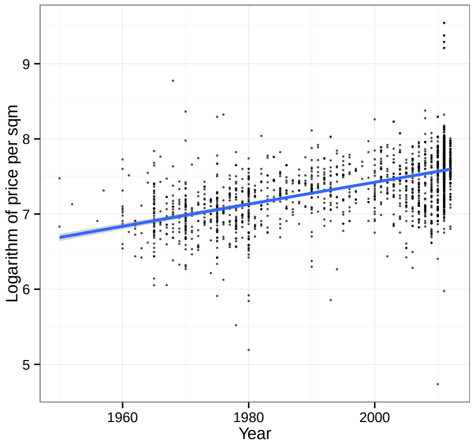
$$\begin{aligned} log(Y) = &\beta_{sqm} log(X_{sqm}) + \beta_{\lesssim 71} X_{\lesssim 71} + \beta_{71} X_{71} + \beta_{81} X_{81} + \beta_{01} X_{01} + \beta_{f1} X_{f1} + \beta_{f2} X_{f2} + \\ &\beta_{f3} X_{f3} + \beta_{f4} X_{f4} + \beta_{f5} X_{f5} + \beta_{f6} X_{f6} + \beta_{f7} X_{f7} + \beta_{\gtrsim f8} X_{\gtrsim f8} + \beta_{con_g} X_{con_g} + \\ &\beta_{con_{rn}} X_{con_{rn}} + \beta_{park} X_{park} + \beta_{safe} X_{safe} + \beta_{front} X_{front} + \beta_{AH} X_{AH} + \\ &\beta_{storage} X_{storage} + \beta_{air} log(X_{air}) + \beta_{rail} X_{rail} + \beta_{port} X_{port} + \\ &\beta_{coast} log(X_{coast}) + \beta_{agr} X_{agr} + \beta_{lowden} X_{lowden} + \epsilon_i \end{aligned} \tag{3.7}$$

Rent model:

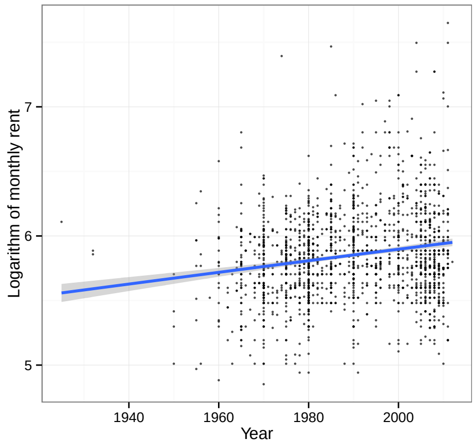
$$
\begin{aligned}
log(Y) = &\beta_{sqm} log(X_{sqm}) + \beta_{\lesssim 81} X_{\lesssim 81} + \beta_{91} X_{91} + \beta_{01} X_{01} + \beta_{f1} X_{f1} + \beta_{f2} X_{f2} + \beta_{f3} X_{f3} + \\
&\beta_{f4} X_{f4} + \beta_{f5} X_{f5} + \beta_{f6} X_{f6} + \beta_{f7} X_{f7} + \beta_{\gtrsim f8} X_{\gtrsim f8} + \beta_{con_g} X_{con_g} + \\
&\beta_{con_{rn}} X_{con_{rn}} + \beta_{park} X_{park} + \beta_{safe} X_{safe} + \beta_{front} X_{front} + \beta_{AH} X_{AH} + \\
&\beta_{storage} X_{storage} + \beta_{air} log(X_{air}) + \beta_{rail} X_{rail} + \beta_{port} X_{port} + \\
&\beta_{coast} log(X_{coast}) + \beta_{agr} X_{agr} + \beta_{for} X_{for} + \beta_{lowden} X_{lowden} + \epsilon_i
\end{aligned} \tag{3.8}
$$

Table 3.12: Specification of variables - real estate price models (Thessaloniki)

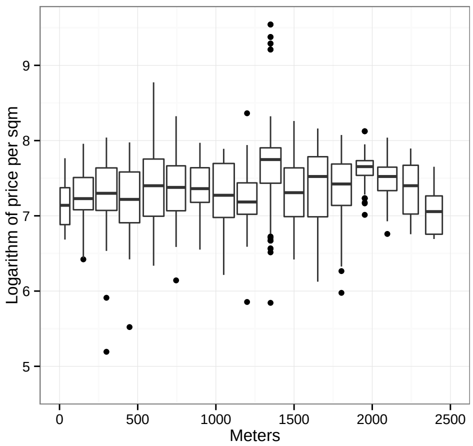
Variable	**Form**	**Name**
Dependent		
Price	logarithm	log(Y)
House attributes		
m^2	logarithm	X_{sqm}
Before 1971	dummy (1 if Yes, 0 if No)	$X_{\lesssim 71}$
Before 1981	dummy (1 if Yes, 0 if No)	$X_{\lesssim 81}$
1971-1980	dummy (1 if Yes, 0 if No)	X_{71}
1981-1990	dummy (1 if Yes, 0 if No)	X_{81}
1991-2000	dummy (1 if Yes, 0 if No)	X_{91}
After 2000	dummy (1 if Yes, 0 if No)	X_{01}
1st floor	dummy (1 if Yes, 0 if No)	X_{f1}
2nd floor	dummy (1 if Yes, 0 if No)	X_{f2}
3rd floor	dummy (1 if Yes, 0 if No)	X_{f3}
4th floor	dummy (1 if Yes, 0 if No)	X_{f4}
5th floor	dummy (1 if Yes, 0 if No)	X_{f5}
6th floor	dummy (1 if Yes, 0 if No)	X_{f6}
7th floor	dummy (1 if Yes, 0 if No)	X_{f7}
8+ floor	dummy (1 if Yes, 0 if No)	X_{f8}
Condition: good	dummy (1 if Yes, 0 if No)	X_{con_g}
Condition: renovation needed	dummy (1 if Yes, 0 if No)	$X_{con_{rn}}$
Parking	dummy (1 if Yes, 0 if No)	X_{park}
Safety door	dummy (1 if Yes, 0 if No)	X_{safe}
Orientation: sunny, corner and front	dummy (1 if Yes, 0 if No)	X_{front}
Auto heat	dummy (1 if Yes, 0 if No)	X_{AH}
Storage place	dummy (1 if Yes, 0 if No)	$X_{storage}$
Transport attributes		
Distance from airport	logarithm (meters)	X_{air}
Distance from railway station	dummy (1 if <500 meters, 0 o/w)	X_{rail}
Distance from port	dummy (1 if <1000 meters, 0 o/w)	X_{port}
Distance from metro	dummy (1 if <500 meters, 0 o/w)	X_{metro}
Other attributes		
Distance from coastline	logarithm	X_{coast}
Ratio of agriculture surface in zone	ratio	X_{agr}
Ratio of forests surface in zone	ratio	X_{for}
Low density	<500 people/km^2	X_{lowden}



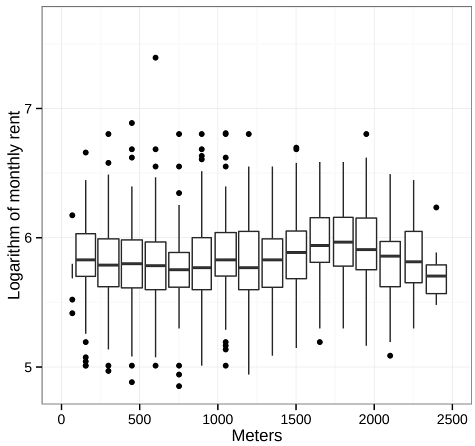
(i) Year of construction (Purchase)



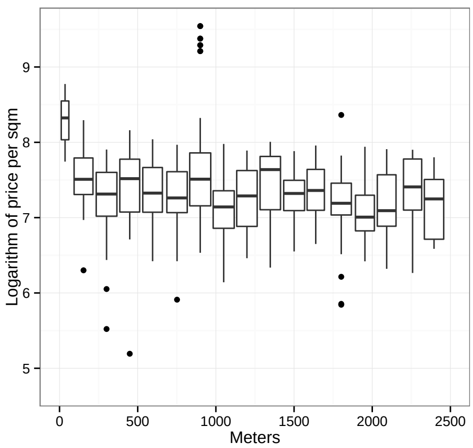
(ii) Year of construction (Rent)



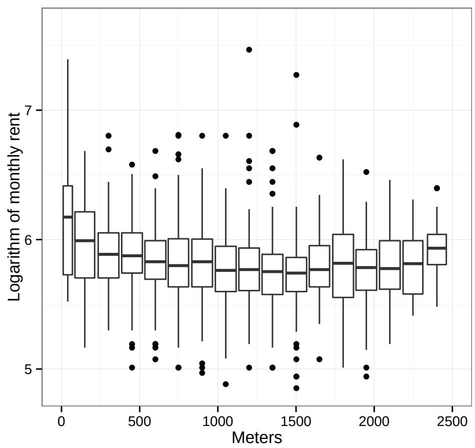
(iii) Distance from metro station (Purchase)



(iv) Distance from metro station (Rent)



(v) Distance from coastline (Purchase)



(vi) Distance from coastline (Rent)

Figure 3.7: Exploratory analysis (Thessaloniki)

3.3.3 Results and comparison

Tables 3.13 and 3.14 summarize the model estimation results of the “purchase” and “rent” model respectively. The results of the models show that the purchase and rent prices increase with the year of construction. It should be noted that in Section 3.2 it was proved that the dwellings constructed before 1950 are more expensive than those built between 1951 and 2010, while those constructed in 60s had the lowest prices. This is interpretable, since they are of architectural inters and their number is limited. However, there is no evidence of similar differentiation in price here, probably because of the weak sample of the database in these years. The purchase price decreases more when the dwellings is in bad condition that “requires renovation”, and less when the condition is “good”. Dwellings in “very good” or “unknown” condition were grouped, and *ceteris paribus* have higher values. The prices increase when a parking, safety door, auto heat or storage place are available. Sunny, corner and front dwellings have higher purchase prices than the others (such as interior). The response characteristic also tends to increase with the floor level. As expected, dwellings located in the ground floor or basement have the lowest prices.

Dwellings located within 500 meters around the National Railway station have lower prices, because of the externalities (noise, social problems, etc.). This finding is similar to Section 3.2, where the purchase prices are -14.8 to -15.4% , and the rents about -9.3 to -9.6% lower. The effect of the railway on dwellings located between 500 and 1000 meters was also investigated, but was found to be insignificant.

Thessaloniki is a coastal city, and the prices reduce logarithmically with the distance from the coastline, a finding similar with that found for the Athens metro area (Section 3.2). The main commercial way of the city (Tsimiski street), is parallel to the coastline, and this geometry creates a zone with high prices.

Finally, the higher the ratio of the “forest” area in the commune, the higher the price of the dwellings is. On the other hand, dwellings located in communes with high ratio of agricultural areas per total m^2, have lower prices. Furthermore, dwellings in communes with low population density (less than 500 people/km^2) are more expensive, and intuitive finding, reflecting preference towards suburban areas.

Despite the fact that the orientation, distance from railway station and distance from port are not significant in every model, they have been retained in all of them, to provide consistency and facilitate easier comparisons across models. Besides the individual parameter significance, a number of aggregated goodness of fit measures are also calculated.

In this research, AIC is closer to zero in the cases of SEM and SAC models for both prices and rents, revealing a better fit of these models (Tables 3.13 and 3.14). Moreover Moran’s I and APLE show that the SEM purchase model and the SEM and SAC rent models successfully address the spatial autocorrelation issue. As expected, the linear regression model (even with the Box-Cox transformation) is outperformed by the spatial models.

Table 3.13: Spatial econometric model estimations - purchase price (Thessaloniki)

	Box-Cox		**SEM**		**SAR**		**SAC**	
Variables	n=2130		n=2130		n=2130		n=2130	
Intercept	3.908	***	7.960	***	6.246	***	7.007	***
House attributes								
X_{sqm}	0.316	***	1.098	***	1.063	***	1.097	***
$X_{\lesssim 71}$	-0.059		-0.163	***	-0.171	***	-0.160	***
X_{71}	-0.046	***	-0.142	***	-0.146	***	-0.142	***
X_{81}	-0.015	.	-0.047		-0.056	*	-0.049	
X_{01}	0.050	***	0.165	***	0.144	***	0.161	***
X_{f1}	0.046	***	0.145	***	0.149	***	0.148	***
X_{f2}	0.061	***	0.202	***	0.200	***	0.203	***
X_{f3}	0.062	***	0.205	***	0.210	***	0.207	***
X_{f4}	0.060	***	0.200	***	0.196	***	0.201	***
X_{f5}	0.100	***	0.340	***	0.353	***	0.349	***
X_{f6}	0.091	***	0.276	***	0.305	***	0.282	***
X_{f7}	0.139	***	0.447	***	0.437	***	0.442	***
X_{f8}	0.125	***	0.362	***	0.402	***	0.366	***
X_{con_g}	-0.030	***	-0.086	***	-0.097	***	-0.087	***
$X_{con_{rn}}$	-0.077	***	-0.234	***	-0.248	***	-0.236	***
X_{park}	0.034	***	0.112	***	0.100	***	0.108	***
X_{safe}	0.020	***	0.077	***	0.071	***	0.077	***
X_{front}	0.156	.	0.623	***	0.478		0.600	*
X_{AH}	0.017	***	0.050	***	0.050	***	0.051	***
$X_{storage}$	0.029	***	0.073	***	0.080	***	0.072	***
Transport attributes								
X_{air}	-0.050	***	-0.111	***	-0.090	***	-0.083	***
X_{rail}	-0.050	*	-0.191		-0.142	*	-0.177	***
X_{port}	0.046	*	0.237	***	0.158	***	0.219	
Other spatial attributes								
X_{coast}	-0.010	***	-0.049	***	-0.047	***	-0.053	***
X_{agr}	-0.207	***	-0.688	***	-0.560	***	-0.633	***
X_{lowden}	0.101	***	0.340	***	0.229	***	0.288	***
λ	-	-	0.423	***	-	-	0.368	***
ρ	-	-	-	-	0.144	***	0.062	***
AIC	-4163	-	960	-	1038	-	955	-
Moran's I	0.185	***	0.020	***	0.126	***	0.121	***
APLE	0.475	***	0.079	***	0.357	***	0.080	***

(·) for $p \lesssim 0.1$, (*) for $p \lesssim 0.05$, (**) for $p \lesssim 0.01$, (***) for $p \lesssim 0.001$

Table 3.14: Spatial econometric model estimations - rent price (Thessaloniki)

Variables	Box-Cox		SEM		SAR		SAC	
	n=2483		n=2483		n=2483		n=2483	
Intercept	0.817	***	5.484	***	4.454	***	5.522	***
House attributes								
X_{sqm}	0.015	***	0.517	***	0.506	***	0.517	***
$X_{\lesssim 81}$	-0.001	*	-0.026	***	-0.024	***	-0.026	***
X_{91}	0.001	*	0.039	***	0.034	***	0.039	***
X_{01}	0.003	***	0.107	***	0.106	***	0.107	***
X_{f1}	0.002	***	0.074	***	0.066	***	0.073	***
X_{f2}	0.003	***	0.088	***	0.078	***	0.088	***
X_{f3}	0.002	***	0.065	***	0.058	***	0.065	***
X_{f4}	0.003	***	0.104	***	0.095	***	0.104	***
X_{f5}	0.004	***	0.147	***	0.133	***	0.147	***
X_{f6}	0.006	***	0.189	***	0.180	***	0.189	***
X_{f7}	0.007	***	0.253	***	0.245	***	0.253	***
X_{f8}	0.009	***	0.241	***	0.262	***	0.241	***
X_{con_g}	-0.002	***	-0.078	***	-0.077	***	-0.078	***
$X_{con_{rn}}$	-0.009	**	-0.307	***	-0.338	***	-0.307	***
X_{park}	0.001	***	0.053	***	0.054	***	0.053	***
X_{safe}	0.002	***	0.055	***	0.024	***	0.055	***
X_{AH}	0.001	***	0.023	***	0.024	***	0.023	***
$X_{storage}$	0.001		0.029	***	0.026	*	0.029	***
Transport attributes								
X_{air}	-0.005	***	-0.186	***	-0.151	***	-0.187	***
X_{rail}	-0.003	**	-0.074	*	-0.073	**	-0.074	*
X_{port}	0.002	**	0.067	***	0.067	***	0.067	***
Other spatial attributes								
X_{coast}	-0.001	*	-0.018	***	-0.019	***	-0.018	***
X_{agr}	-0.008	***	-0.288	***	-0.269	***	-0.288	***
X_{for}	0.001	***	0.016	***	0.014	***	0.016	***
X_{lowden}	0.003	***	0.114	***	0.088	***	0.115	***
λ	-	-	0.255	***	-	-	0.260	***
ρ	-	-	-	-	0.130	***	-0.004	***
AIC	-18565	-	-1042	-	-1001	-	-1040	-
Moran's I	0.143	***	0.013	***	0.084	***	0.013	***
APLE	0.275	***	0.029	***	0.175	***	0.029	***

(·) for p $\lesssim 0.1$, (*) for p $\lesssim 0.05$, (**) for p $\lesssim 0.01$, (***) for p $\lesssim 0.001$

3.3.4 The role of metro under construction

In this section, the role of the metro under construction is being investigated. The models of Section 3.3.2 are re-estimated, including one more explanatory variable, the proximity to the metro station ($\beta_{metro}X_{metro}$).

The results show that dwellings located within 500 meters of the expected location of the metro stations have a lower purchase price (Table 3.15), while the rental price is not affected by the proximity, since the variable is insignificant (the results are not presented here because of that reason). This finding leads to the following:

The construction of the new metro line affects negatively the purchase prices of nearby residencies, because of the negative externalities that are generated. This conclusion verifies the findings of Karanikolas and Anastasiadou (2012) that 73% of the shop owners established around the Venizelou station believe that they the construction works worsen their financial situation. Therefore, it is concluded that this negative impact is also reflected in the real estate purchase prices. On the other hand, the rents are not affected by the forthcoming metro stations, perhaps due to their temporary nature. This indicates that the candidate tenants are indifferent to the future implementation of the metro.

In Section 3.2 it was proved that dwellings located within 500 meters of stations of an already operating metro system have about 6.7% to 9.4% higher purchase prices and 4.3% to 4.9% higher rents. This suggests that people perceive quite differently the construction period (which is mostly associated with externalities), and the operation period (which is in general associated with increased accessibility). It should be noted, that the same company manages the construction and operation of both metros (namely Attiko Metro S.A.), which indicates that the quality of service in Thessaloniki is expected to be as high as in Athens.

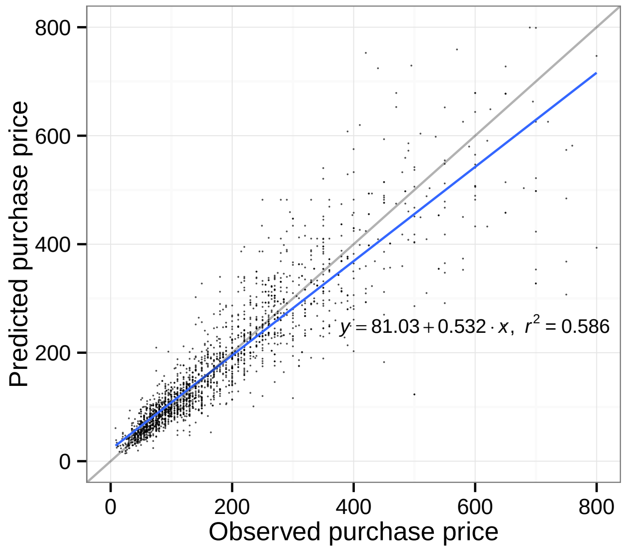
Table 3.15: Spatial econometric model estimations (with metro accessibility variable) - purchase price (Thessaloniki)

	Box-Cox		SEM		SAR		SAC	
Variables	n=2130		n=2130		n=2130		n=2130	
Intercept	3.958	***	8.145	***	6.357	***	7.125	***
House attributes								
X_{sqm}	0.317	***	1.098	***	1.069	***	1.097	***
$X_{\lesssim 71}$	-0.057		-0.161	***	-0.167	***	-0.159	***
X_{71}	-0.056	***	-0.142	***	-0.147	***	-0.142	***
X_{81}	-0.046	*	-0.052	***	-0.062	***	-0.053	***
X_{01}	0.045	***	0.160	***	0.140	***	0.157	***
X_{f1}	0.048	***	0.147	***	0.154	***	0.150	***
X_{f2}	0.062	***	0.204	***	0.205	***	0.205	***
X_{f3}	0.063	***	0.208	***	0.213	***	0.210	***
X_{f4}	0.062	***	0.203	***	0.205	***	0.205	***
X_{f5}	0.103	***	0.344	***	0.361	***	0.354	***
X_{f6}	0.094	***	0.283	***	0.311	***	0.289	***
X_{f7}	0.141	***	0.450	***	0.443	***	0.444	***
X_{f8}	0.135	***	0.382	***	0.429	***	0.388	***
X_{con_g}	-0.031	***	-0.087	***	-0.096	***	-0.088	***
$X_{con_{rn}}$	-0.077	***	-0.235	***	-0.246	***	-0.237	***
X_{park}	0.033	***	0.109	***	0.101	***	0.107	***
X_{safe}	0.020	***	0.077	***	0.069	***	0.077	***
X_{front}	0.150	.	0.610	***	0.469		0.586	*
X_{AH}	0.018	***	0.052	***	0.054	***	0.052	***
$X_{storage}$	0.027	***	0.072	***	0.073	***	0.070	***
Transport attributes								
X_{air}	-0.049	***	-0.108	***	-0.081	***	-0.078	***
X_{rail}	-0.044	*	-0.167		-0.125	*	-0.156	***
X_{port}	0.050	**	0.244	***	0.166	***	0.224	
X_{metro}	-0.022	***	-0.082	***	-0.072	***	-0.075	***
Other spatial attributes								
X_{coast}	-0.016	***	-0.066	***	-0.072	***	-0.072	***
X_{agr}	-0.186	***	-0.635	***	-0.454	***	-0.559	***
X_{for}	0.005	**	0.014	**	0.021	***	0.016	***
X_{lowden}	0.080	***	0.275	***	0.127	***	0.207	***
λ	-	-	0.413	***	-	-	0.351	***
ρ	-	-	-	-	0.149	***	0.068	***
AIC	-4177.8	-	954		1019	-	949	-
Moran's I	0.180	***	0.020	***	0.116	***	0.021	***
APLE	0.470	***	0.080	***	0.341	***	0.082	***

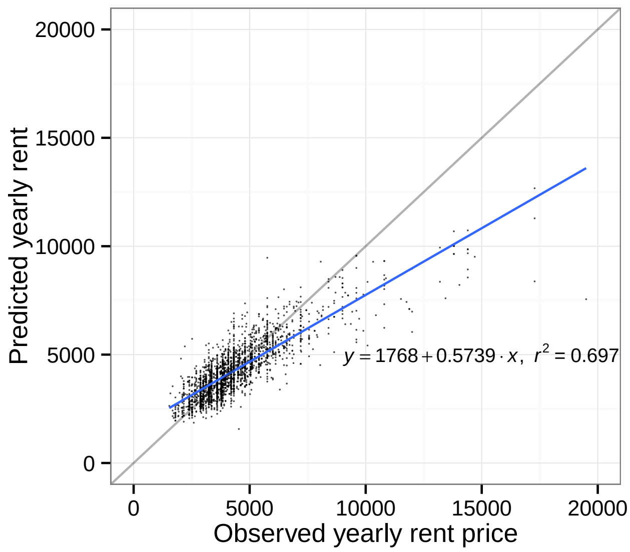
(·) for $p \lesssim 0.1$, (*) for $p \lesssim 0.05$, (**) for $p \lesssim 0.01$, (***) for $p \lesssim 0.001$

Figure 3.8 shows the relation between the observed and the predicted –by the SEM– values, for both purchase prices and rents. Concerning the purchase prices (Figure 3.8i) it seems that there is not a systematic over– or under-prediction and there is heteroscedasticity. In terms of rents (Figure 3.8ii), on the other hand, the model predicts accurately for rents up to 8000 euros per year, while it increasingly under-predicts as rent price increases. Anselin and Lozano-Garcia (2008) and Baltagi et al. (2008) have proposed methods to reduce this effect.

Figure 3.9 shows the predicted values of the SEM models and Figure 3.10 the purchase–to–rent ratio. In order to compute this ratio, the respective SEM models where applied on both datasets, so as to generate both prices for each data point. The spatial variation of the ratio reveals higher values in the Southern portion of the study area, which are also the most expensive. According to DiPasquale and Wheaton (1996), $Price = Rent/i$, where i is the interest rate. In this case, i ranges from about 2% in the central areas, to about 5% in the Northern and Southern suburbs. This ratio indicates whether an area is suitable for investment on property (low values), or rent instead of purchase (high values). The ratio takes higher values in the more prestigious areas, such as Kalamaria and Panorama. Panorama is characterized as one of the most privileged areas in Thessaloniki, because it is located in a forest area, with view on the Thermaikos gulf, and Kalamaria is a very good coastal area to live in.

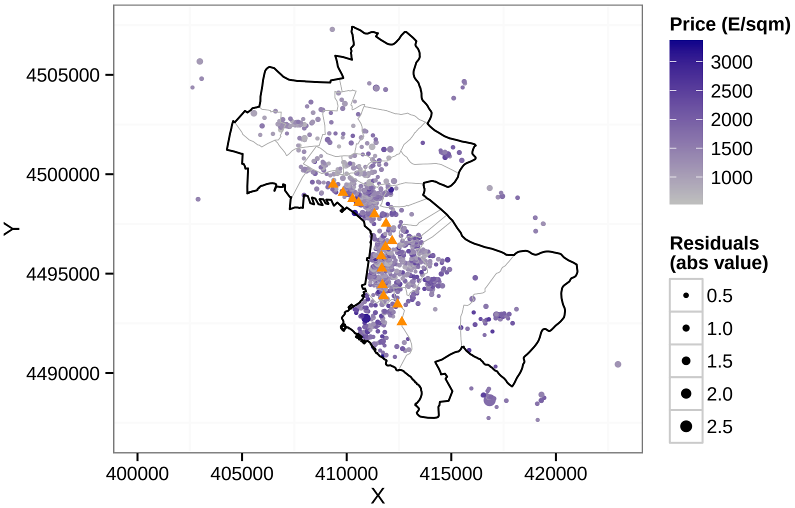


(i) Purchase price in thousands of euros

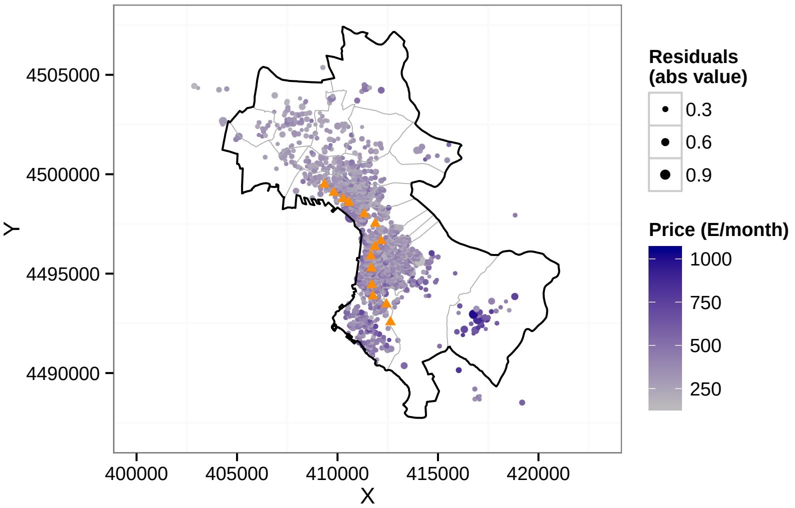


(ii) Rent in euros

Figure 3.8: Observed vs predicted prices (Thessaloniki)



(i) Predicted prices by SEM (Purchase)



(ii) Predicted prices by SEM (Rent)

Figure 3.9: Spatial distribution of predicted prices in Thessaloniki. The orange triangles show the locations of the metro stations

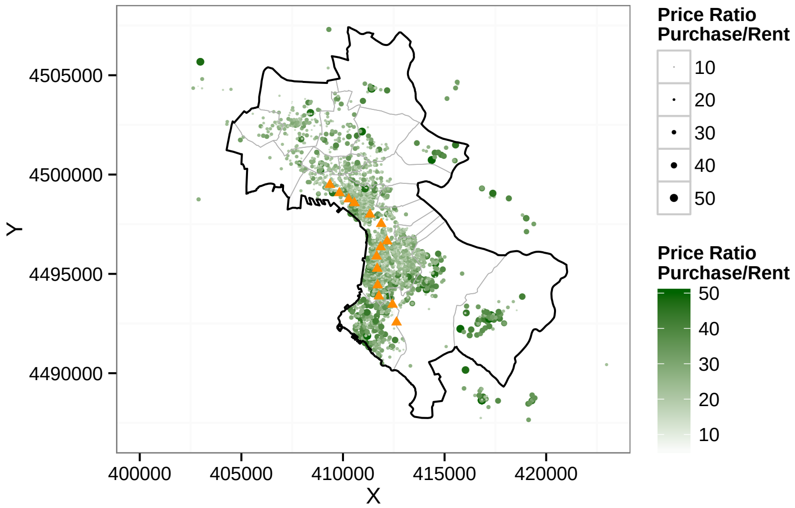


Figure 3.10: Predicted price-to-rent ratio in Thessaloniki. The orange triangles show the locations of the metro stations

3.3.5 Conclusions

The objective of this research is multifaceted: 1) to measure the effects of transportation infrastructure and land-use characteristics on the real estate prices and rents of Thessaloniki, Greece; 2) investigate the current impact of a planned metro line, and 3) underline the relative benefits of various spatial econometrics models in order to be integrated in Integrated Land Use and Transport (LUTI) models.

The results reveal a negative effect of the proximity to port and railway station, and a positive impact of the airport. The metro line, which is still under construction, has a negative impact on the purchase prices, while it does not affect the rents. In the first case, this is presumably because of the negative externalities generated by the construction, which is generally not expected to be be completed within the next 5 years, a fact that keeps the expected demand in low levels. This conclusion comes in accordance with other research works based on surveys in the area. In the second case, the variable is insignificant since the rent is usually short term which means that the future implementation of the metro does not affect the candidate tenants' decision as much.

Moreover, similar to Section 3.2 it is concluded that the spatial econometric models can effectively remove the spatial autocorrelation and the predicted values are closer to the real, comparing with the traditional linear regression. As a consequence, their implementation to the next generation LUTI models is recommended. Further steps and research directions include, measuring the prices again, when the metro and/or the ongoing debt crisis will have been finished; this will allow for the effects of these interventions to be better quantified. Moreover, data is planned be collected in regular basis (e.g. annually) for a period of e.g. 10-20 years, when the externalities –if any– will be clear. Finally, the developed real estate price model is planned to be incorporated it to a future LUTI model of Thessaloniki, along with all other models.

3.4 The price–to–rent ratio model

3.4.1 Real estate ratios

Real estate ratios (e.g. price–to–income, price–to–rent) are widely used for valuation purposes. The price–to–rent ratio is used as a comparison indicator of the average purchase and rent prices of an area. It is computed a by dividing the mean purchase price over the annual rent (DiPasquale and Wheaton, 1996):

$$\textit{Price-to-rent ratio} = \frac{Price}{Rent \cdot 12} \tag{3.9}$$

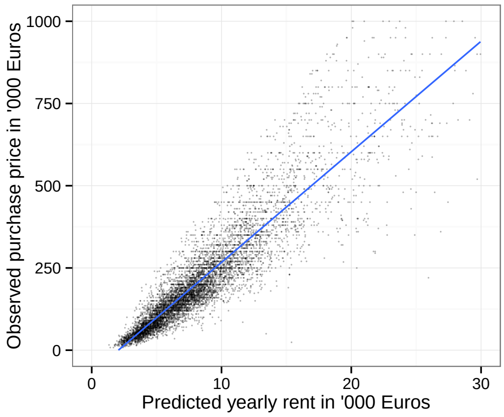
Trulia (2013) developed a methodology for the price–to–rent ratio computation in US, aiming to provide advice to its users (www.trulia.com) on which choice, between renting and buying a dwelling, is preferable. They found that in 2013 is US, it is more economical to buy a house than rent. In the proposed methodology, financial rates (e.g. mortgage and inflation) are taken into consideration for the computation of the net present value, while several assumptions are made. The following thresholds have been established: 1) for ratio values from 1 to 15, it is much better to buy than rent; 2) for values between 16 and 20, it is typically better to rent than buy; 3) for values 21 or more, it is much better to rent than buy.

Moreover, another application of the price–to–rent ratio –computed at aggregate spatial level (city or country)– is as indicator for the detection of real estate bubbles. Kraier and Chishen (2004) used it to predict the US real estate bubble, by revealing the its sharp increase in time. Another study in Korea (Park et al., 2010) showed that the ratio increased by 2.66 times from 1999 to 2006 in Kangnam, while in the whole nation the increase was 1.73 times.

3.4.2 Exploratory analysis

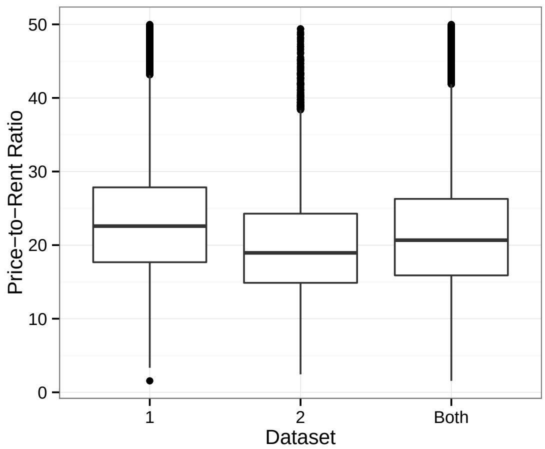
The aim of this research is to develop a price–to–rent ratio model using disaggregate data, in order to examine how it is affected by transport infrastructure developments.

Using the spatial autoregressive models developed in Section 3.2, the predicted prices and rents for all data points were estimated, wherever they were not available. While this imputation technique might transfer errors in the data (e.g. there is no residual variance in the estimated values, which fit perfectly along the regression line), it is an effective way to compute both price and rent of each dwelling in our datasets. The resulting dataset contains data of 16139 dwellings located in Attica, for which both the purchase price and rent (or an estimate of them) are known. An exploratory analysis is first performed to the two sub-datasets, in order to reveal potential systematic trends in the observed –and respectively predicted– prices and rents. Figures 3.11i and 3.11ii make a comparison between the observed and predicted prices of the two datasets. The slope of the price-rent line in dataset 1 is steeper that in dataset 2, indicating higher price-to-ratio values in this dataset, which is more clear in Figure 3.11iii. Figures 3.12i and 3.12ii show the variance of the price–to–rent ratio per year of construction and floor level for the joined dataset; dwellings constructed in 60s have the lowest ratio, and its value increases for newer and older (similar behavior to price [Section 3.2]). Dwellings located less than 500 meters distance from ISAP and national railway stations (Figures 3.12iii and 3.12iv) have lower ratio values.



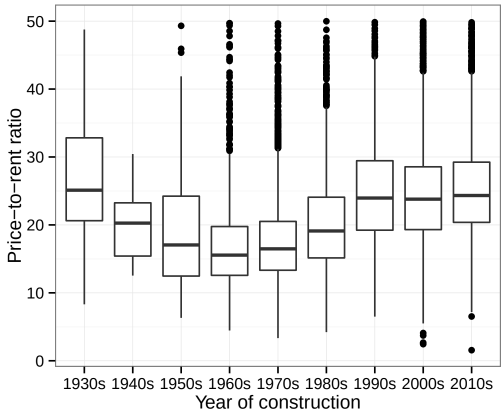
(i) Dataset 1

(ii) Dataset2

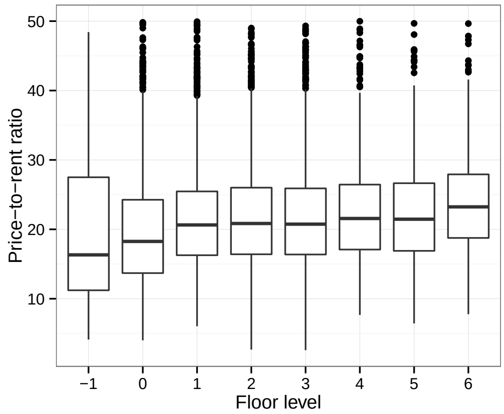


(iii) Ratio comparison between purchase and rent sub-datasets

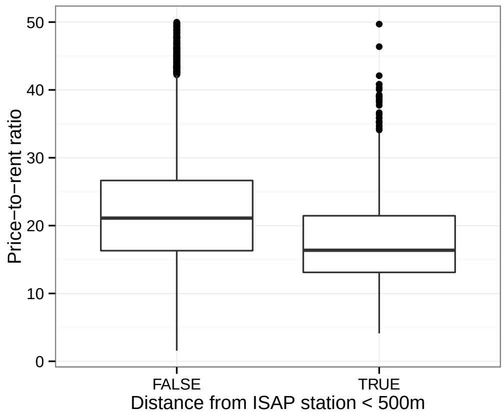
Figure 3.11: Purchase vs rent price



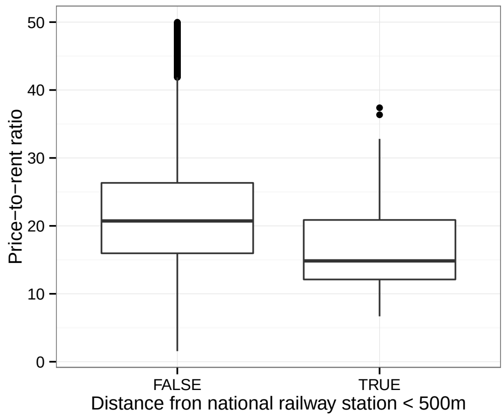
(i) Dataset 1



(ii) Dataset2



(iii) Ratio comparison between purchase and rent sub-datasets



(iv) Ratio comparison between purchase and rent sub-datasets

Figure 3.12: Purchase vs rent price of price–to–rent ratio

3.4.3 Model development and results

Dependent variable of the suggested model is the natural logarithm of the price–to–rent ratio, and explanatory are building attributes, transport accessibility indicators, demographics and other spatial attributes. In order to capture the error between the two ex–sub–datasets, an extra term α is included, taking the value 1 when the regression point belongs to the sub-dataset 1 (purchase) and 0 when it belongs to sub-dataset 2 (rent).

The generalized form of the suggested linear model, is:

$$ln(Y) = \beta_{house}X_{house} + \beta_{transport}X_{transport} + \beta_{other}X_{other} + \alpha + \varepsilon \tag{3.10}$$

where Y is the price-to-rent ratio, X_{house} represents the house attributes, $X_{transport}$ the transport accessibility characteristics, X_{other} the demographic and the other spatial attributes, β are the coefficients, α the error term between the two datasets and ε the error term of the unobserved characteristics. Table 3.16 describes the variables of the models in detail.

Table 3.16: Specification of variables used in price–to–rent models

Variable	Specification	Name
Dependent		
Price	logarithm	
House Attributes		
m^2	logarithm	X_{sqm}
Before 1951	dummy, 1 if TRUE	$X_{\lesssim 51}$
1951-1960	dummy, 1 if TRUE	X_{51}
1961-1970	dummy, 1 if TRUE	X_{61}
1971-1980	dummy, 1 if TRUE	X_{71}
1981-1990	dummy, 1 if TRUE	X_{81}
1991-2000	dummy, 1 if TRUE	X_{91}
2001-2010	dummy, 1 if TRUE	X_{01}
>=2011	dummy, 1 if TRUE	$X_{\gtrsim 11}$
Basement	dummy, 1 if TRUE	X_{base}
Ground floor	dummy, 1 if TRUE	X_{gf}
1st floor	dummy, 1 if TRUE	X_{f1}
2nd floor	dummy, 1 if TRUE	X_{f2}
3rd floor	dummy, 1 if TRUE	X_{f3}
4th floor	dummy, 1 if TRUE	X_{f4}
5th floor	dummy, 1 if TRUE	X_{f5}
6th floor or more	dummy, 1 if TRUE	$X_{\gtrsim f6}$
Parking	dummy, 1 if TRUE	$X_{parking}$
Auto heat	dummy, 1 if TRUE	X_{AH}
A/C	dummy, 1 if TRUE	X_{AC}
Single family house	dummy, 1 if TRUE	X_{sfh}
Orientation: front	dummy, 1 if TRUE	X_{front}
Transport attributes		
In the inner Ring	dummy, 1 if inside	X_{ring}
Distance from bus station	dummy, 1 if < 50m	X_{bus}
Distance from metro station	dummy, 1 if < 500m	X_{metro}
Distance from ISAP station	dummy, 1 if < 500m	X_{isap}
Distance from Rail station	dummy, 1 if < 500m	X_{natl}
Distance from railway	dummy, 1 if < 100m	$X_{railway}$
Distance from suburban rail station	dummy, 1 if < 2000m	X_{sub}
Distance from tram station	dummy, 1 if < 500m	X_{tram}
Distance from airport	dummy, 1 if < 5000m	X_{air}
Distance from port (Pireaus/Rafina)	logarithm	X_{port}
Distance from marina	dummy, 1 if < 1500m	X_{mar}
Distance from ring-road (1)	dummy, 1 if 0 to 250m	X_{rroad1}
Distance from ring-road (2)	dummy, 1 if 250m to 1500m	X_{rroad2}
Distance from ring-road (3)	dummy, 1 if >1500m	X_{rroad3}
Other spatial attributes		
Northern suburbs	dummy, 1 if TRUE	X_{NS}
University area	dummy, 1 if TRUE	X_{uni}
Distance from archaeological sites	logarithm	X_{arch}
Distance from coastline	meters	X_{coast}
Low population density	dummy, 1 if <500/km^2, zone	$X_{lowdens}$
High education level	log of % of population with university degree or higher, zone	$X_{highedu}$
Dataset 1	dummy, 1 if TRUE	X_{d1}

A linear regression with OLS, a spatial error model (SEM), a spatial autocorrelation model with lagged dependent variable (SAR) and a spatial autocorrelation model (SAC) have been estimated and the results are presented in Table 3.17. The following conclusions can be made.

Building attributes

Similar to the purchase and rent models presented in Section 3.2, the dependent variable does not increase linearly with the year of construction. Dwellings built before 1951 have higher price–to–rent ratio, those constructed between 1961 and 1971 have the smallest, and from 1971 to 2011 its value increases. The ratio increases continuously with the floor level, while the growth is more evident for dwellings located at 4th floor and more. The availability of parking or auto heat, being single family or front houses result in higher ratio values.

Transport attributes

Concerning the impact of transportation infrastructure, dwellings located within 50m distance from bus station, 500m from metro or tram station, or 2000m from sub-urban rail station and 1500m from marinas, have higher price–to–rent ratio. Moreover, its value increases with the distance from port. On the other hand, dwellings located within 500m from ISAP or national railway stations, and those within 100m distance from railway, have lower ratio values. The negative coefficient of the variables for small (<250m) or medium (250m-1500m) distance from the Attica tollway, indicates that renting is preferable than buying. Table 3.18 shows the elasticities of transportation variables. The price–to–rent ratio of dwellings located within 500m is lower by -7.94% (SAR) to -8.24% (OLS and SEM), while of those located within 250m from Attica tollway is lower by -9.97% (SEM) to -10.86% (SAR).

Other attributes

Dwellings located in the Northern Suburbs (prestigious area in Athens), in the University area next to Zografou Campus, (municipalities: Zografou, Goudi, Kaisariani, Pagrati), in low density areas and municipalities with high education level, have higher ratio. Finally, the variable used to capture the error between the two datasets is negative, which indicates that the points of the first dataset (purchase) have lower ratio values than the second (rent), although this variable is not significant in the SEM and SAC models.

3.4.4 Model comparison

The results indicate that the ratio values predicted by the SAC model are closer to the real (lower AIC), while the SEM model solves more the spatial autocorrelation (lower Moran's I and APLE). Figure 3.13i shows the spatial allocation of the predicted by the SEM model ratio, and Figure 3.13ii compares the observed and predicted values, revealing the presence of heteroscedasticity.

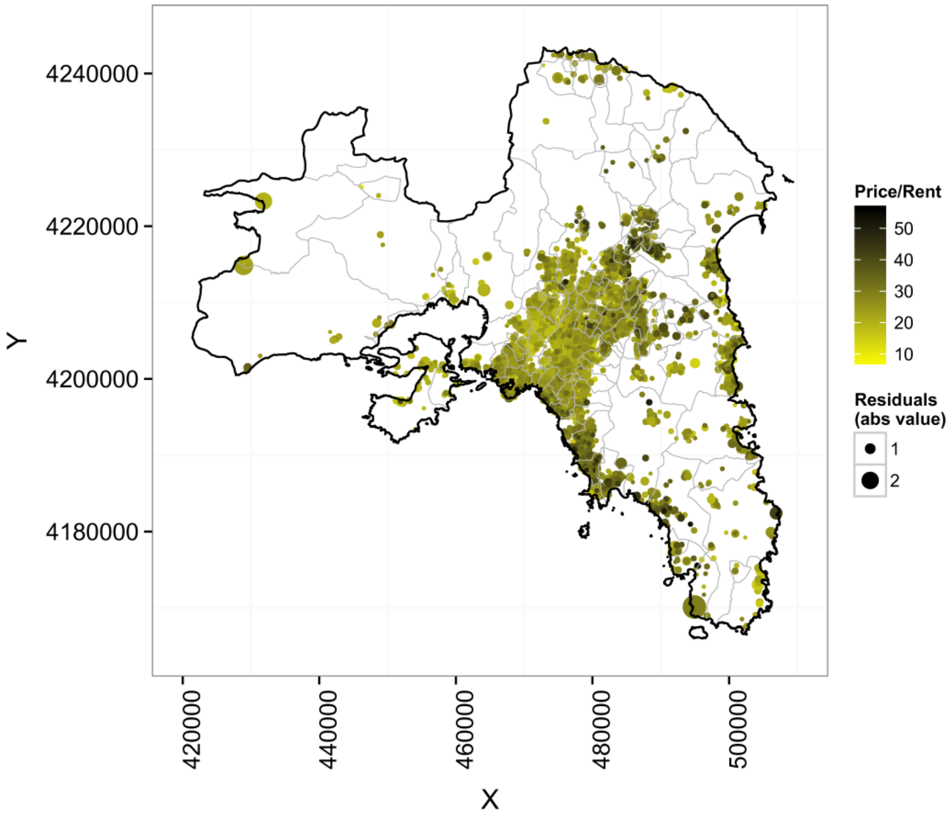
Table 3.17: Price–to–rent model estimations

	OLS		SEM		SAR		SAC	
	n=16139		n=16139		n=16139		n=16139	
Variables	**Coefficient**	**t-test**	**Coefficient**	**t-test**	**Coefficient**	**t-test**	**Coefficient**	**t-test**
Intercept	0.195	(3.885)	0.207	(3.592)	0.100	(1.558)	0.152	(2.530)
House attributes								
X_{sqm}	0.357	(79.926)	0.354	(77.648)	0.346	78.014	0.351	(75.535)
$X_{\lesssim 51}$	0.110	(7.084)	0.110	(7.107)	0.100	(5.632)	0.106	(6.957)
X_{51}	reference category for year of construction							
X_{61}	-0.086	(-6.096)	-0.086	(-6.188)	-0.083	(-4.945)	-0.085	(-6.185)
X_{71}	-0.025	(-1.848)	-0.028	(-2.083)	-0.026	(-1.512)	-0.028	(-2.046)
X_{81}	0.062	(4.293)	0.058	(3.997)	0.058	(3.308)	0.058	(4.040)
X_{91}	0.075	(4.931)	0.069	(4.542)	0.068	(3.836)	0.068	(4.588)
X_{01}	0.097	(6.743)	0.096	(6.668)	0.091	(5.290)	0.094	(6.708)
$X_{\gtrsim 11}$	0.122	(7.729)	0.120	(7.525)	0.114	(6.205)	0.117	(7.472)
X_{base}	-0.180	(-7.738)	-0.178	(-7.717)	-0.180	(-7.862)	-0.180	(-7.801)
X_{gf}	reference category for floor							
X_{f1}	0.101	(17.076)	0.096	(16.187)	0.101	(17.142)	0.098	(16.465)
X_{f2}	0.112	(17.841)	0.107	(17.127)	0.112	(18.054)	0.109	(17.315)
X_{f3}	0.113	(16.013)	0.106	(15.115)	0.113	(16.203)	0.109	(15.303)
X_{f4}	0.135	(17.314)	0.132	(16.904)	0.137	(17.716)	0.135	(17.142)
X_{f5}	0.140	(14.899)	0.139	(14.844)	0.143	(15.373)	0.141	(14.868)
$X_{\gtrsim f6}$	0.180	(15.294)	0.178	(15.154)	0.182	(15.695)	0.181	(15.374)
$X_{parking}$	0.018	(3.129)	0.016	(2.852)	0.014	(2.254)	0.015	(2.483)
X_{AH}	0.043	(7.790)	0.041	(7.437)	0.040	(7.257)	0.040	(7.421)
X_{AC}	-0.036	(-7.396)	-0.033	(-6.868)	-0.035	(-7.319)	-0.034	(-7.020)
X_{sfh}	0.045	(4.187)	0.047	(4.385)	0.043	(4.111)	0.045	(4.190)
X_{front}	0.015	(2.256)	0.013	(1.967)	0.015	(2.298)	0.013	(2.063)
Transport attributes								
X_{ring}	0.173	(17.525)	0.171	(14.941)	0.148	(14.749)	0.159	(14.383)
X_{bus}	0.020	(3.448)	0.020	(3.274)	0.019	(3.388)	0.020	(3.338)
X_{metro}	0.038	(5.224)	0.038	(4.510)	0.035	(4.740)	0.037	(4.571)
X_{isap}	-0.086	(-12.030)	-0.086	(-10.403)	-0.072	(-10.176)	-0.079	(-9.926)
X_{natl}	-0.073	(-3.972)	-0.075	(-3.562)	-0.067	(-3.612)	-0.071	(-3.573)
$X_{railway}$	-0.061	(-3.181)	-0.056	(-2.647)	-0.059	(-3.097)	-0.057	(-2.836)
X_{sub}	0.040	(3.426)	0.040	(2.981)	0.039	(3.634)	0.039	(3.232)
X_{tram}	0.030	(3.667)	0.029	(3.001)	0.024	(3.013)	0.026	(2.865)
X_{air}	-0.069	(-2.043)	-0.064*	(-1.646)	-0.063	(-1.803)	-0.064	(-1.712)
X_{port}	0.068	(16.046)	0.068	(13.975)	0.053	(11.108)	0.060	(11.724)
X_{mar}	0.026	(2.760)	0.026	(2.389)	0.023	(2.433)	0.025	(2.381)
X_{rroad1}	-0.109	(4.892)	-0.105	(-4.239)	-0.100	(-4.610)	-0.102	(-4.403)
X_{rroad2}	-0.023	(-2.112)	-0.022	(-1.795)	-0.021	(-2.113)	-0.022	(-1.950)
X_{rroad3}	reference category for distance from ring-road							
Other spatial attributes								
X_{NS}	0.057	(6.132)	0.057	(5.277)	0.031	(3.210)	0.042	(4.178)
X_{uni}	0.088	(6.798)	0.084	(5.566)	0.089	(6.963)	0.087	(6.230)
X_{arch}	0.073	(18.952)	0.073	(16.408)	0.056	(14.175)	0.064	(13.103)
X_{coast}	-0.00001	(-12.327)	-0.00001	(-10.663)	0.00001	(-9.200)	-0.00001	(-9.148)
$X_{lowdens}$	0.035	(6.502)	0.035	(5.521)	0.027	(5.118)	0.031	(5.098)
$X_{highedu}$	0.020	(5.968)	0.020	(5.195)	0.015	(4.500)	0.017	(4.535)
X_{d1}	-0.008	(-1.748)	-0.006*	(-1.385)	-0.007	(-1.900)	-0.007*	(-1.165)
λ	-	-	0.163	-	-	-	0.096	-
ρ	-	-	-	-	0.130	-	0.066	-
R^2	0.546	-	-	-	-	-	-	-
AIC	1084.2	-	810.6	-	818.6	-	795.9	-
Moran's I	0.127	-	0.020	-	0.040	-	0.022	-
APLE	0.201	-	0.036	-	0.069	-	0.038	-

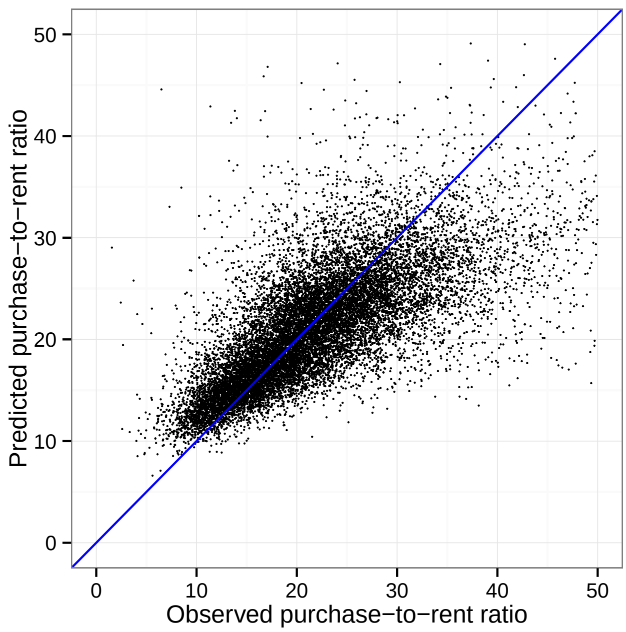
*Not significant at 0.05 level

Table 3.18: Price–to–rent model elasticities

Variable	**OLS**	**SEM**	**SAR**	**SAC**
X_{ring}	18.89	18.65	18.54	18.46
X_{bus}	2.02	2.02	2.21	2.16
X_{metro}	3.87	3.87	4.11	4.04
X_{isap}	-8.24	-8.24	-7.94	-8.11
X_{natl}	-7.04	-7.23	-7.41	-7.32
$X_{railway}$	-5.92	-5.45	-6.56	-5.92
X_{sub}	4.08	4.08	4.58	4.26
X_{tram}	3.05	2.94	2.80	2.82
X_{air}	-6.67	-6.20	-6.99	-6.62
X_{port}	7.04	7.04	6.28	6.63
X_{mar}	2.63	2.63	2.68	2.71
X_{rroad1}	-10.33	-9.97	-10.86	-10.35
X_{rroad2}	-2.27	-2.19	-2.38	-2.33



(i) Predicted ratio



(ii) Predicted vs Observed

Figure 3.13: SEM results of price–to–rent ratio

3.4.5 Conclusions

The contribution of this research to the state–of–the–art of real estate modeling, is the development of a price–to–rent ratio model. The aim is to measure the impact of large-scale transportation infrastructure location on price–to–rent ratio, and investigate the factors that determine which dwellings are appropriate for investments.

The results indicate that bus, metro, tram, suburban railway and marinas have positive impact on the price–to–rent ratio of dwellings located within particular distances from their stations, while on the other hand, ISAP, national railway airport and Attica tollway, have negative impact. Moreover, dwellings located in the inner ring have, by far, higher ratio than the others.

The increased price–to–rent ratio around some transportation infrastructure, indicates that prices are significantly higher than rents in specific locations. This could be interpreted as the result of an increase in demand for property purchase before –or after– the implementation of the policy, which has led to higher bids. As a result, the purchase prices are maintained in high levels, while the rents have reached equilibrium in lower.

At the same time, dwellings around transportation infrastructure with negative impact on the price–to–rent ratio, are highly affected by negative externalities. In case of rail transport infrastructure locations, it could be concluded that the level of ratio is related to the time of operation.

The benefits of modelling the price–to–rent ratio in urban planning research, are multiple. First, it may influence the location choice of utility-maximizing residents. This indicates that agents (households) who desire to purchase housing will be directed to dwellings with lower price–to–rent ratio, while those who plan to rent, to dwellings with higher. The probability of the latter to be relocated is higher, which should be taken into consideration in urban planning. Moreover, modeling the price–to–rent ratio creates new perspectives in maximizing the economic benefits of transport infrastructure investments, by directly relating them with investments in property. Furthermore, it could be used to estimate the value capture, so as to implement fairer taxation, aiming to assist in financing transportation investments, as well as to deal with equity issues often associated with such policies, that lead to displacement of lower income households (i.e. to compensate losers) or to extend public transit to serve lower income neighbourhoods. In addition, complex socioeconomic and demographic factors, and policy incentives that affect the rental and homeownership markets (for example in the US the tax expenditures for mortgage interest deduction dwarfs the relatively small tax incentives and direct investments in affordable rental housing), and these in turn profoundly influence "agent-based" decisions in these markets.

In conclusion, the use of price–to–rent ratio in LUTI models should be further investigated.

3.5 Impact of recession on real estate prices

3.5.1 Economic crisis

Greece has been highly affected by the financial crisis of the last five years. As in every market, its impact is apparent on real estate. According to Bank of Greece (2012b), real estate purchase prices in Athens have declined since 2008 (-4.5% in 2009, -3.2% in 2010, -6.4% in 2011, -11.8% in 2012 and -12.7% the first quarter of 2013). The real estate agents in Greece believe that despite the relative stabilization noted the last two quarters (Bank of Greece, 2012a), the market conditions will deteriorate further. Brokers assume that prices have declined by 11% to 24% since the latest quarter of 2012, while they note that the interest for smaller, second-hand residencies in increasing. On average, dwellings are sold by 22% lower than the initial listed price, after being advertised 10 months.

The pre-mentioned studies indicate the change of the real estate prices based on aggregated data. In this paper, we present a methodology that using disaggregate information, measures the impact of the crisis in two levels: holistic (total reduction between two time periods) and local (impact of transportation infrastructure).

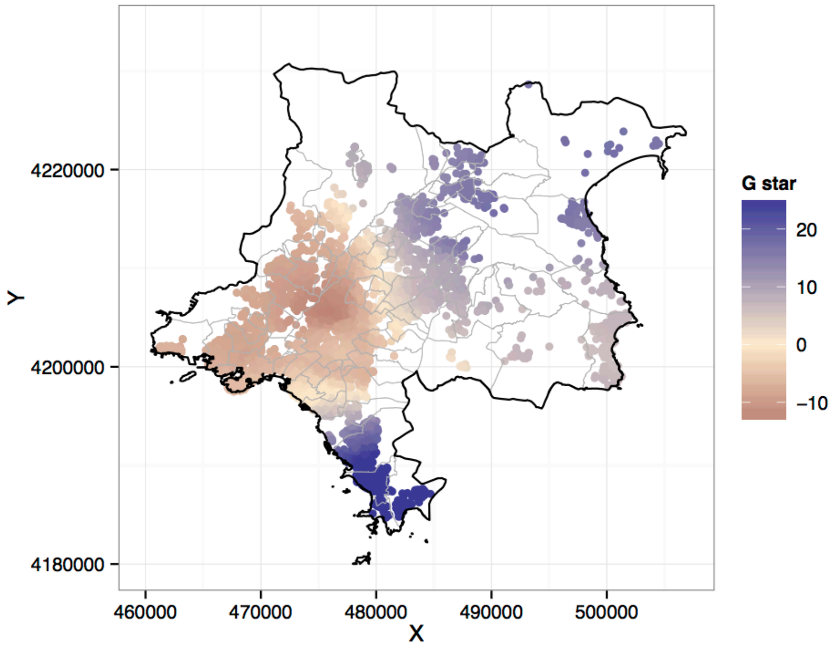
3.5.2 Spatial cluster analysis

Aiming to identify the presence of local spatial price clusters, the Gstar statistic is being used. The four maps presented in Figures 3.14 and 3.15, one for each dataset (purchase prices and rents 2011 and 2013), depict the output of the Gstar –the Z-values– that can be used for diagnostic purposes. High Z-values indicate the potential of a cluster with high prices to be present and similarly, low negative Z-values indicate a cluster of low prices. The closer the Z-value is to zero, the less possible is the presence of a cluster. The weights matrices required for the test had been structured using a 10% of the neighbor data points for each dataset.

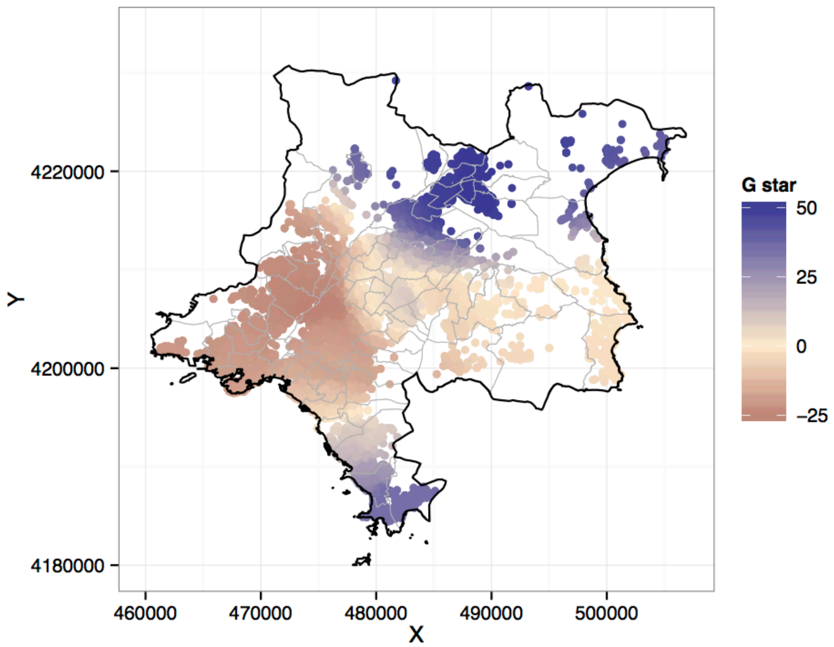
The results verify the presence of spatial price clusters in all the examined periods (2011 and 2013) and prices (purchase and rent). In all cases, clusters with higher prices (blue color) are formed far from the city center, in the Northern Suburbs and South Coastline, where people of higher incomes and education level usually reside. On the other hand, it is observed the presence of a lower price cluster on the west (red color).

Interesting enough is the fact that the city center (Municipality of Athens) does not form a single cluster. Dwellings located in the northern city center are co-clustered at low prices, while those located in the south result in Z-values closer to zero, indicating the absence of clear clustering. This is easier to interpret for an Athens' citizen, since the south city center is a conglomeration of "prestigious" areas (eg. Kolonaki, Plaka) and "degraded" (eg. Metaxourgeio) that share common boundaries. The differentiation of clustering in the city center is less evident in 2013, when the "prestigious" areas seem to be partially incorporated in the lower-price cluster of the west, indicating the decrease of the prices as an effect of the crisis.

Another remarkable difference between 2011 and 2013 is that the “unclustered” area (ivory color) between the city center and the Northern Suburbs increases, with parallel limiting of the Northern Suburbs cluster extend, both in purchase prices and rents. Similar effect is noticed at the area between the city center and the South Coastline. This indicates that the “elite” (extremes) areas resist on the crisis and maintain the prices high, while the prices of the southern Northern Suburbs (previously wealthier) decline. Similarly, the housing values of “poor” areas decline, creating a more apparent cluster of low prices. In other words, the range of Z-values in 2013 is higher than in 2011 (-26 to 51 instead of -12 to 25 for purchase prices, and -19 to 31 instead of -12 to 28 for rents). The strengthen of the extreme and weakness of the middle prices could be interpreted as the equivalent “middle-class squeeze” as direct impact of the recession.

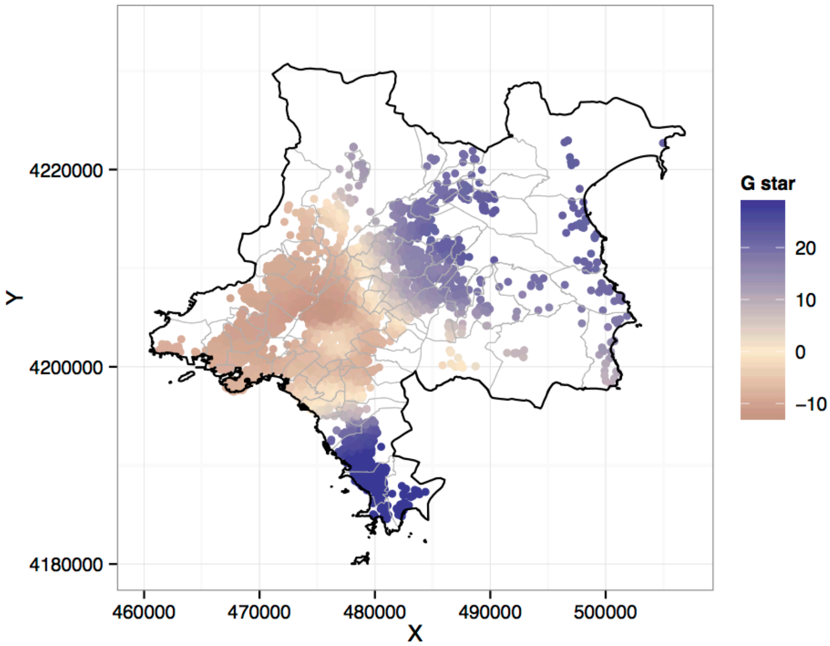


(i) Purchase prices 2011

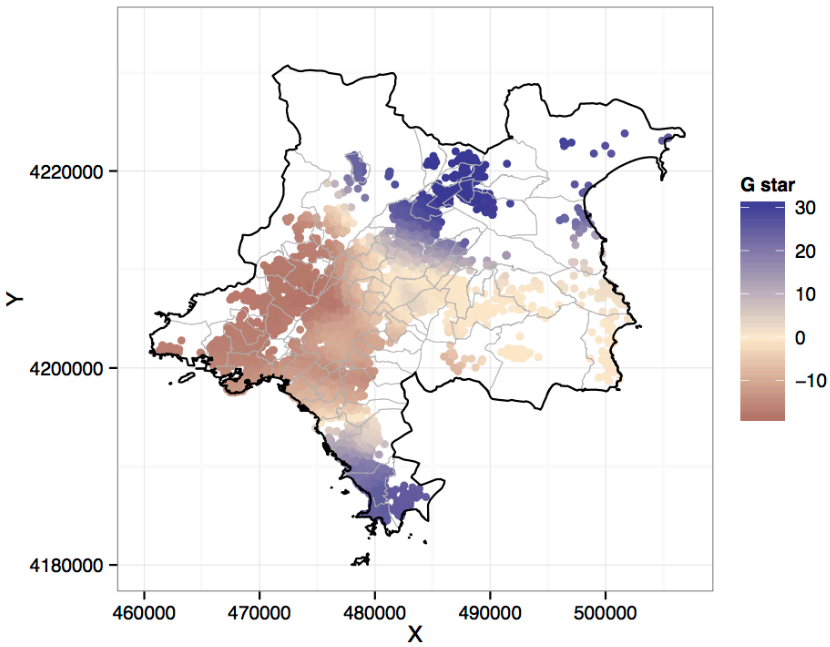


(ii) Purchase prices 2013

Figure 3.14: G^* spatial clustering - purchase prices



(i) Rent prices 2011



(ii) Rent prices 2013R

Figure 3.15: G^* spatial clustering - rent prices

3.5.3 Price model

Aiming to measure the impact of the crisis on real estate prices, spatial error models are estimated for purchases prices and rents. For this purpose, two real estate datasets collected in 2011 and 2013 (Section 6.3.1) were merged and used jointly. Table 3.19 shows the specification of the variables used in the presented models. Figure 3.16 depicts the distribution of the variables: distance to metro $< 500m$, inner-ring, and distance to tollway $< 1500m$, over the natural logarithm of purchase and rent prices in the two examined years. In both cases, the mean price around transportation infrastructure decreases, while it is observed an increase of the range between the 1st and 3rd quantile of rents, probably due to the larger amount of data collected in 2013.

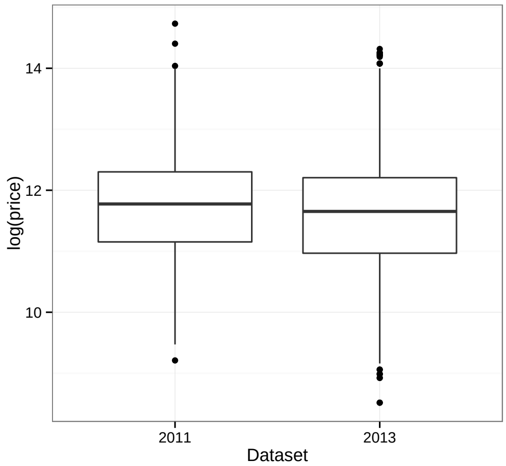
Prices are modeled in two ways: "holistically" or "globally", and "locally". Two SEM models, one for purchase prices and one for rents, are first developed using as explanatory variables the structural characteristics of the dwellings, accessibility indicators (distances to main transportation infrastructure and policy locations) and a dummy variable that takes value 1 when the observation belongs to the new dataset (2013) and 0 in the other case (2011), to capture the difference of mean of the error term; however, the difference between the variance of the data remains. Table 3.20 shows the model estimations. The results indicate that the newer the dwelling, the higher the price is, while those constructed before 1950 are more expensive. Moreover, the higher the floor the more expensive the dwelling is, while the availability of any extra facilities such as parking, garden, storage room, fire place, auto heat and air-conditioning has an added value. Dwellings with sea view and those closer to the coastline are more expensive. Concerning the transportation attributes, dwellings closer to the CBD, within 500 meters from a metro or tram station, within 1500 from a marina, and in the Inner-ring, are more expensive. On the other hand, dwellings in 500m distance from and ISAP or national railway station, 7000m from the airport and 1500 from the Attica tollway have lower prices. These findings come in accordance with Section 3.2.

What is more interesting in this research, is the scale of the crisis-related variable coefficient: it takes value -0.201 at the purchase model, which indicates that purchase prices have been reduced by -18.2% since 2011 (1-$e^{-0.201}$), and -0.165 at the rent model, indicating -15.2% reduction (1-$e^{-0.165}$). The variance between the datasets remains, which might have affected the estimated reduction. That figure of price decline is lower than the reported in Bank of Greece (2012a,b) and probably does not represent the actual reduction. However, this outcome was expected since the listed prices are lower than the final (Bank of Greece, 2012a). The objective is not to measure the absolute reduction, but get a first, general indication, and the propose a way for its computation based on disaggregate data. The significance of the spatial error variable in both models and the low AIC and Moran's I, verify the appropriateness of using the SEM model.

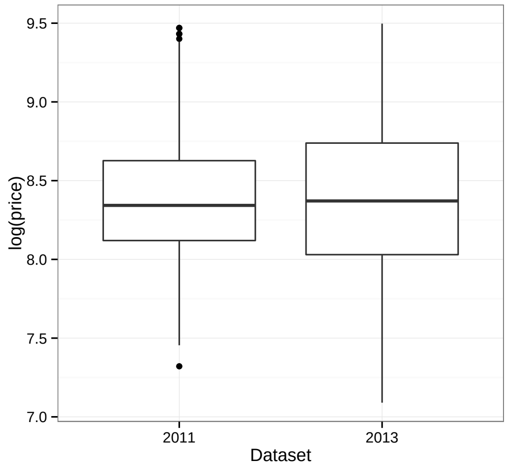
Having measured the reduction "globally" using disaggregate data, the impact of the proximity to transportation infrastructure locations is quantified. The aim is to measure the ability –or not– of transportation infrastructure to maintain the dwellings' prices at higher levels. In order to facilitate the comparison, the models using both datasets, instead of estimating one for 2013 and then compare with Section 3.2.

Using interaction variables for each transportation infrastructure location, two SEM models were estimated: one for purchase prices and one for rents. The results are presented in Table 3.21. The estimated coefficients of the structural characteristics of buildings are similar to the joint model (Table 3.20). Concerning the proximity to transportation infrastructure location, the results show that the impact differs between the two periods.

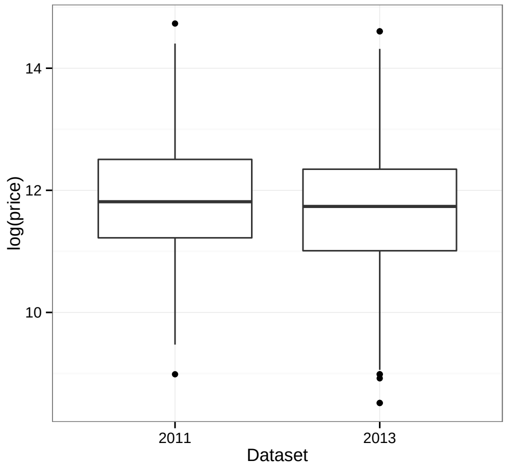
Table 3.22 shows the percentage impact of transportation infrastructure and policies on purchase prices and rents, and its change (increase/decrease) between the two examined periods. It is observed that overall, the positive impact has been reduced, with exception the location of tram within 500m of dwellings for rent. Dwellings located within 500m from a metro station have 4.3% higher purchase prices than the rest, when the respective percentage in 2011 was 7.5%, resulting in 42.5% decrease of the metro's impact. Similarly, dwellings located within 500m euclidian distance from ISAP stations are by -10.2% less expensive, while in 2011 they were -6.7% (53.5% increase of the reduction). Moreover, proximity to CBD is by 18.8% less important for the purchase prices than in 2011. The highest reduction is observed at the impact of bus station locations, the proximity to which turns from positive to negative. However, this huge change could probably be biased, due to the small radius considered (50m). Concerning the rents, dwellings located within 500m from a metro station are 3% more expensive than the others in 2013, while in 2011 they were 8.1%, which indicates a -62.5% decrease of the impact. The negative effect of the proximity to the airport increases by 36.7% in the rent model, significantly higher than the 2.4% in the price model. Finally, dwellings located within 1500m radius from the Attica tollway are -14% less expensive than the others in 2013, which is interpreted as 20.2% increase of the negative impact from 2011.



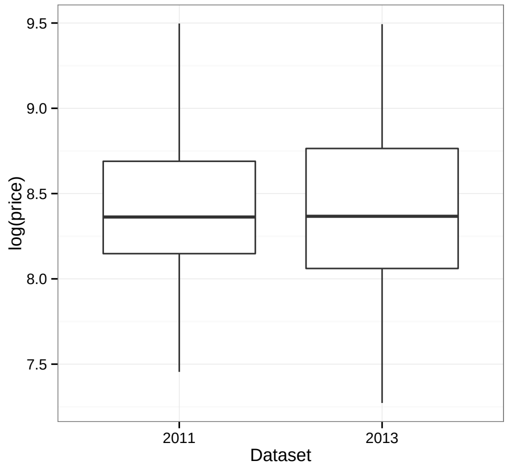
(i) Distance form metro < 500m - purchase



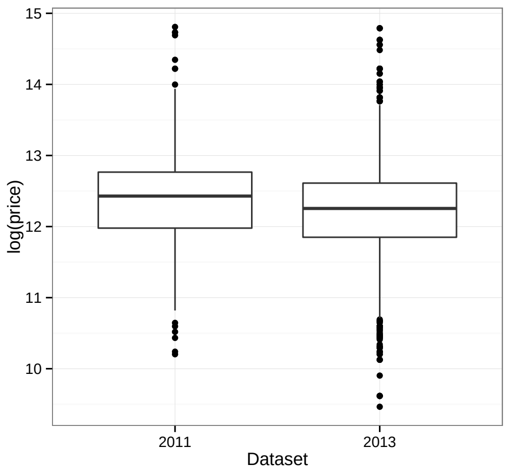
(ii) Distance form metro < 500m - rent



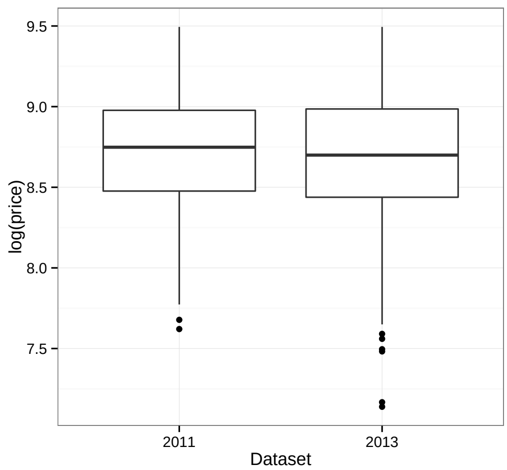
(iii) Ring road - purchase



(iv) Ring road - rent



(v) Distance form tollway <1500m - purchase



(vi) Distance form tollway <1500m - rent

Figure 3.16: Exploratory analysis of 2013 dataset

Table 3.19: Specification of variables - crisis model

Variable	**Variable specification**	**Name**
Dependent		
Price	logarithm	
House Attributes		
m^2	logarithm	X_{sqm}
Before 1950	dummy, 1 if TRUE	$X_{\lesssim 50}$
1950-1960	dummy, 1 if TRUE	X_{50}
1960-1970	dummy, 1 if TRUE	X_{60}
1970-1980	dummy, 1 if TRUE	X_{70}
1980-1990	dummy, 1 if TRUE	X_{80}
1990-2000	dummy, 1 if TRUE	X_{90}
2000-2010	dummy, 1 if TRUE	X_{00}
>=2010	dummy, 1 if TRUE	$X_{\gtrsim 10}$
Basement	dummy, 1 if TRUE	X_{base}
Ground floor	dummy, 1 if TRUE	X_{gf}
1st floor	dummy, 1 if TRUE	X_{f1}
2nd floor	dummy, 1 if TRUE	X_{f2}
3rd floor	dummy, 1 if TRUE	X_{f3}
4th floor	dummy, 1 if TRUE	X_{f4}
5th floor	dummy, 1 if TRUE	X_{f5}
6th floor or more	dummy, 1 if TRUE	$X_{\gtrsim f6}$
Parking	dummy, 1 if TRUE	$X_{parking}$
Open parking	dummy, 1 if TRUE	X_{openP}
Storage place	dummy, 1 if TRUE	$X_{storage}$
Fireplace	dummy, 1 if TRUE	X_{firep}
Auto heat	dummy, 1 if TRUE	X_{AH}
A/C	dummy, 1 if TRUE	X_{AC}
Single family house	dummy, 1 if TRUE	X_{sfh}
View: Sea	dummy, 1 if TRUE	$X_{seaview}$
Orientation: front	dummy, 1 if TRUE	X_{front}
Northern suburbs	dummy, 1 if TRUE	X_{NS}
Distance from coastline	logarithm	X_{coast}
Transport attributes		
Distance from CBD	logarithm	X_{cbd}
In the inner Ring	dummy, 1 if inside	X_{ring}
Distance from bus station	dummy, 1 if < 50m	X_{bus}
Distance from metro station	dummy, 1 if < 500m	X_{metro}
Distance from ISAP station	dummy, 1 if < 500m	X_{isap}
Distance from Rail station	dummy, 1 if < 500m	X_{natl}
Distance from railway	dummy, 1 if < 100m	$X_{railway}$
Distance from suburban rail station	dummy, 1 if < 2000	X_{sub}
Distance from tram station	dummy, 1 if < 500m	X_{tram}
Distance from airport	dummy, 1 if < 7000m	X_{air}
Distance from port (Pireaus/Rafina)	logarithm	X_{port}
Distance from marina	dummy, 1 if < 1500m	X_{mar}
Distance from ring-road	dummy, 1 if < 1500m	X_{rroad}

Table 3.20: Price model estimations - global

	Purchase1		**Rent1**	
	n=19703		n=18311	
	k=8		k=6	
Variables	**Coefficient**	**t-test**	**Coefficient**	**t-test**
Intercept	6.655	(46.57)	5.527	(65.03)
House attributes				
X_{sqm}	1.009	(204.89)	0.709	(171.08)
$X_{\lesssim 50}$	0.235	(6.69)	0.123	(3.37)
X_{50-80}	reference category for year of construction			
X_{80}	0.151	(19.18)	0.038	(6.98)
X_{90}	0.277	(28.25)	0.120	(17.81)
X_{00}	0.433	(52.91)	0.258	(41.18)
$X_{\gtrsim 10}$	0.516	(60.47)	0.369	(43.81)
X_{base}	-0.299	(-23.71)	0.133	(10.19)
X_{gf}	reference category for floor			
X_{f1}	0.076	(11.20)	0.019	(3.69)
X_{f2}	0.105	(14.76)	0.023	(4.13)
X_{f3}	0.143	(18.72)	0.063	(9.79)
X_{f4}	0.194	(23.32)	0.068	(8.88)
X_{f5}	0.238	(23.11)	0.102	(10.26)
$X_{\gtrsim f6}$	0.290	(23.23)	0.133	(10.19)
X_{NS}	0.341	(15.77)	0.286	(22.03)
$X_{parking}$	0.063	(9.82)	0.053	(10.29)
X_{garden}	0.021	(3.14)	0.024	(4.33)
$X_{storage}$	0.045	(8.03)	0.013	(2.74)
X_{firep}	0.034	(6.03)	0.075	(13.88)
X_{AH}	0.038	(6.42)	0.010	(2.11)
X_{AC}	0.026	(4.97)	0.076	(17.91)
$X_{seaview}$	0.022	(2.23)	0.090	(9.62)
X_{coast}	-0.063	(-9.53)	-0.044	(-9.99)
Transport attributes				
X_{cbd}	0.096	(7.60)	0.024	(3.50)
X_{metro}	0.051	(2.59)	0.054	(4.21)
X_{isap}	-0.037	(-1.86)	-0.028	(-2.16)
X_{tram}	0.082	(4.18)	0.083	(6.30)
X_{natl}	-0.098	(-2.20)	-	-
X_{rail}	-	-	-0.071	(-2.72)
X_{mar}	0.068	(2.87)	0.043	(2.67)
X_{air}	-0.159	(-3.48)	-0.159	(-4.53)
X_{ring}	0.234	(6.93)	0.214	(12.15)
X_{rroad}	-0.061	(-2.74)	-0.139	(-10.19)
$X_{database=2}$	-0.201	(-45.91)	-0.165	(-44.54)
λ	0.744	***	0.535	***
R^2	0.91		0.87	
AIC	7078	(OLS: 16131)	460	(OLS: 4183)
Moran's I	0.001	***	-0.001	***

(·) for $p \lesssim 0.1$, (*) for $p \lesssim 0.05$, (**) for $p \lesssim 0.01$, (***) for $p \lesssim 0.001$

Table 3.21: Price Model Estimations - local

	Purchase2		Rent2	
	n=19703 k=4		n=18311 k=4	
Variables				
Intercept	6.215	(51.96)	5.457	(64.00)
House attributes				
X^1_{sqm}	1.019	(161.15)	0.730	(172.13)
X^2_{sqm}	1.013	(152.92)	0.696	(166.78)
$X_{\lesssim 50}$	0.255	(7.03)	0.125	(3.39)
X_{50-80}	reference category for year of construction			
X_{90}	0.279	(27.51)	0.120	(17.68)
X_{00}	0.422	(50.03)	0.258	(40.93)
$X_{\gtrsim 10}$	0.507	(57.19)	0.369	(43.46)
X_{base}	-0.296	(-22.78)	-0.081	(-7.38)
X_{gf}	reference category for floor			
X_{f1}	0.080	(11.47)	0.018	(3.36)
X_{f2}	0.107	(14.71)	0.022	(3.89)
X_{f3}	0.146	(18.59)	0.062	(9.45)
X_{f4}	0.195	(22.89)	0.066	(8.58)
X_{f5}	0.228	(21.72)	0.101	(10.04)
$X_{\gtrsim f6}$	0.280	(22.30)	0.131	(9.96)
X_{NS}	0.336	(18.15)	0.287	(22.08)
$X_{parking}$	0.065	(9.93)	0.053	(10.26)
X_{garden}	0.030	(4.34)	0.024	(4.14)
$X_{storage}$	0.045	(7.76)	0.013	(2.75)
X_{firep}	0.042	(7.27)	0.076	(13.90)
X_{AH}	0.041	(6.80)	0.010	(1.99)
X_{AC}	0.032	(5.91)	0.076	(17.80)
$X_{seaview}$	0.023	(2.35)	0.090	(9.64)
X_{coast}	-0.048	(-8.36)	-0.044	(-10.00)
Transport attributes				
X_{cbd}	-	-	0.022	(3.07)
X^1_{cbd}	0.129	(12.28)	-	-
X^2_{cbd}	0.106	(12.28)	-	-
X^1_{metro}	0.072	(3.59)	0.078	(5.40)
X^2_{metro}	0.042	(1.90)	0.030	(2.01)
X_{isap}	-	-	-0.029	(-2.20)
X^1_{isap}	-0.069	(-3.39)	-	-
X^2_{isap}	-0.108	(-5.18)	-	-
X^1_{tram}	0.147	(7.22)	0.075	(4.74)
X^2_{tram}	0.126	(6.65)	0.083	(5.87)
X_{rail}	-	-	-0.074	(5.87)
X^1_{bus}	0.029	(2.95)	-	-
X^2_{bus}	-0.041	(-3.94)	-	-
X_{mar}	-	-	0.040	(2.51)
X^1_{mar}	0.087	(3.97)	-	-
X^2_{mar}	0.070	(3.42)	-	-
X^1_{air}	-0.191	(-4.53)	-0.127	(-3.01)
X^2_{air}	-0.196	(-4.65)	-0.178	(-4.67)
X^1_{ring}	0.338	(11.49)	0.228	(11.91)
X^2_{ring}	0.247	(7.45)	0.199	(10.12)
X^1_{rroad}	-0.093	(-4.35)	-0.124	(-7.97)
X^2_{rroad}	-0.100	(-4.95)	-0.151	(-10.37)
λ	0.651	***	0.533	***
R^2	0.90		0.87	
AIC	8028	(OLS: 16175)	676	(OLS: 4336)
Moran's I	0.020	***	-0.001	***

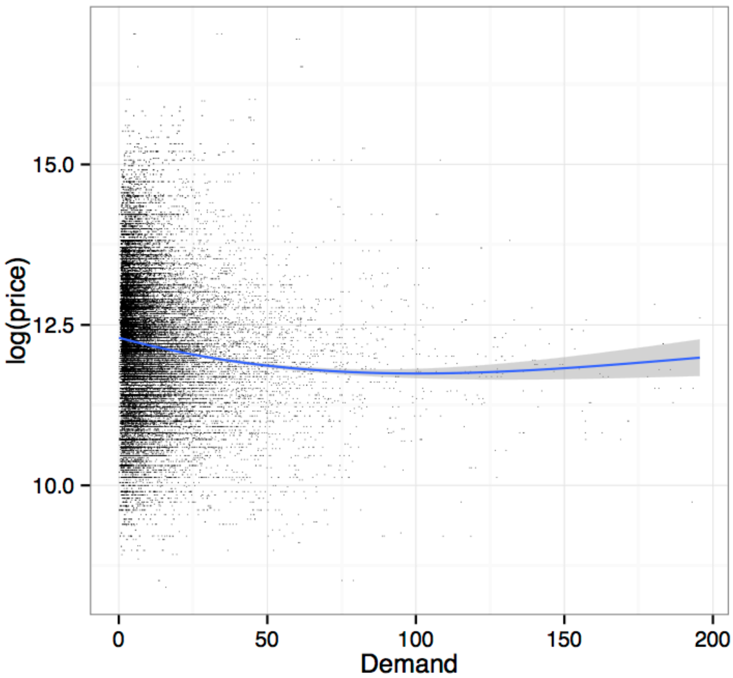
(·) for $p \lesssim 0.1$, (*) for $p \lesssim 0.05$, (**) for $p \lesssim 0.01$, (***) for $p \lesssim 0.001$

Table 3.22: Percentage impact of transportation infrastructure location

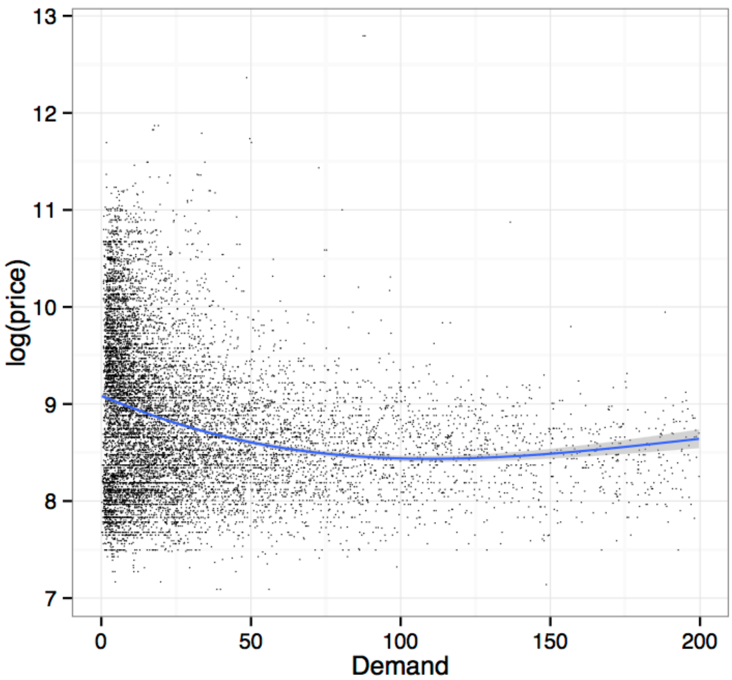
	Effect on prices			Effect on rents		
	2011 %	2013 %	Difference %	2011 %	2013 %	Difference %
X_{cbd}	+13.8	+11.2	-18.8	-	-	-
X_{metro}	+7.5	+4.3	-42.5	+8.1	+3.0	-62.5
X_{isap}	-6.7	-10.2	-53.5	-	-	-
X_{tram}	+15.8	+13.4	-15.2	+7.8	+8.7	+11.1
X_{bus}	+2.9	-4.0	-236.5	-	-	-
X_{mar}	+9.1	+7.3	-20.2	-	-	-
X_{air}	-17.4	-17.8	-2.4	-11.9	-16.3	-36.71
X_{ring}	+40.2	+28.0	-30.3	+25.6	+12.6	-50.7
X_{rroad}	-8.9	-9.5	-7.2	-11.7	-14.0	-20.2

3.5.4 Demand model

In this step of the analysis, the “demand” for on-line listed dwellings is modelled. As a measure of “demand” is considered the average number of on-line visits that an advertised dwelling has per week. This weight certainly does not represent the actual demand, but it could be considered as a proxy for the purpose of this research. For the development of these models, the 2013 dataset is used. Considering that there are dwellings with higher visibility than others –e.g. due to premium membership of the advertisers– observations with extreme demand values (more than 200 visits per week) were eliminated. Figure 3.17 illustrates the relationship between the “demand” and the logarithm of the purchase (Figure 3.17i) and rent (Figure 3.17ii) prices. The fitted line is an spline with 2 breaking points, and the shaded band is a 95% confidence interval on the fitted values. Dwellings at about €160,000 purchase price and €4,500 yearly rent have higher visibility than the others, and probably more possibilities to be sold/rent. However, this is just a hypothesis and can only be verified if the exact transaction price is available.



(i) Purchase



(ii) Rent

Figure 3.17: Demand vs price

The spatial error model formulation is used for the parametrisation of the demand. As dependent variable is used the natural logarithm of the number of visits per week, and as explanatory the structural characteristics of the dwelling (house attributes) and accessibility variables (transport attributes). Unlike price models, the effect of the metro, ISAP and national railway stations locations around 1000m (instead of 500m) is examined in this case. The reason behind using a higher buffer is that here the demand is determined by the user and is limited by the available on-line search criteria (e.g. when searching geolocated dwellings by map, it is difficult to select dwellings precisely within 500m from a metro station).

Due to the large number of available data, a sample of random 10,000 dwellings for sale has been used, in order to reduce the required computational time. As a result, two models were estimated, one for purchase (n=10,000) where 4 neighbor points were used for the construction of the spatial weights matrix, and one for rents (n=11,663) where 10 neighbor points were used. The ideal number of neighbor data points was selected taking into consideration the criteria of Moran's I minimization.

The results, presented in Table 3.23, show that demand increases with the logarithm of the dwelling's surface and decreases with price. The newer the dwelling the higher the demand, while residencies for rent constructed before 1950s have higher demand than those constructed between 50s and 80s. The higher the floor level and the structural quality, the higher the demand.

Concerning the proximity to transportation infrastructure locations, it is quite interesting that demand of dwellings for sale increases with the logarithmic distance from CBD, while for rent decreases. This indicates the tendency of citizens willing to purchase a house to move in the suburbs, while on the other hand those aiming to rent try to minimize their transportation costs by moving in the city center. Dwellings located 1000m around the metro stations have higher demand for both purchase prices and rents. Residences for sale located 1000m around ISAP stations have higher demand, while on the other hand those for rent have lower. This is the effect of the degradation of the areas around ISAP stations (Section 3.2) and indicates that people looking for a temporal residence are not willing to be located there, while those who aim to maximize their benefit by taking advantage of the economic and degraded situation tend to invest on these under-valuated residences. Dwellings for rent located within 1500m from tram stations have higher demand, while the proximity to national railway station has negative effect for the demand of both purchase and rent. Furthermore, the suburban rail station locations have negative impact on the demand for purchase, and positive for rent. Moreover, the demand for purchase dwellings located 1500m from marinas is higher, while the respective weight for rental dwellings is negative. Finally, proximity to port has positive impact on the demand.

The autoregressive coefficient in significant in both models, while the AIC of SEM is lower than the OLS, indicating that the predicted by this model values are closer to the real. The Moran's I test shows that the models partially solve spatial autocorrelation. However, despite the effort to developed as more detail models as possible, both result in low R^2. Taking this into consideration, it was decided the exploration of an alternative modeling methodology, the spatial filtering (SF). In one of the few spatial filtering applications in transportation research, Wang et al. (2013) showed that SF results in better model fit and lower spatial autocorrelation than the SAR lag model. While there might be possibility for endogeneity in these models, it has't been investigated at this phase of the research.

Applying spatial filtering in such big dataset was challenging. It should be noted that due to the computational time required for the generation of the eigentvectors, the rent model needed 72 hours to be estimated (107 vectors), and the purchase about 6 days (243 vectors). The results of spatial filtering are presented in Table 3.24. Indeed, spatial filtering models result in better fit (higher R^2 and lower AIC), while Moran's I worsens. Due to the large number of eigenvectors, the first six from each model –which are those with the highest Moran's I– are presented in Table 3.25 and Figures 3.18 and 3.19.

Table 3.23: Demand model estimations - SEM

	Purchase3		Rent3	
	n=10000		n=11663	
	k=4		k=10	
Variables				
Intercept	5.539	(14.38)	12.127	(31.99)
House attributes				
X_{sqm}	0.632	(16.31)	0.801	(22.14)
X_{price}	-0.567	(-19.30)	-1.305	(-36.86)
$X_{\lesssim 50}$	-	-	0.504	(2.44)
X_{80}	0.235	(6.67)	0.165	(5.11)
X_{90}	0.402	(9.25)	0.403	(10.52)
X_{00}	0.288	(8.00)	0.624	(16.19)
$X_{\gtrsim 10}$	0.305	(7.03)	0.822	(15.42)
X_{apart}	-	-	-0.656	(-13.83)
X_{single}	-	-	-0.476	(-9.88)
X_{base}	-0.212	(-3.63)	-0.278	(-4.60)
X_{gf}	reference category for floor			
X_{f1}	0.204	(6.68)	0.198	(6.47)
X_{f2}	0.296	(9.25)	0.290	(8.68)
X_{f3}	0.242	(7.07)	0.341	(9.05)
X_{f4}	0.340	(9.14)	0.544	(12.22)
X_{f5}	0.402	(8.06)	0.551	(9.33)
$X_{\gtrsim f6}$	0.418	(7.07)	0.612	(8.08)
X_{NS}	0.111	(2.13)	0.426	(9.58)
X_{ah}	-	-	0.177	(6.65)
X_{firep}	-	-	-0.11	(-4.09)
X_{ng}	-	-	0.097	(2.88)
X_{safety}	-	-	0.119	(4.83)
X_{new}	0.194	(6.22)	0.265	(7.36)
$X_{excellent}$	0.162	(6.75)	0.095	(4.29)
X_{sea}	-	-	0.099	(3.54)
X_{corner}	-	-	0.160	(7.61)
Transport attributes				
X_{cbd}	0.091	(2.77)	-0.117	(-4.08)
$X_{metro1000}$	0.094	(1.67)	0.082	(1.67)
$X_{isap1000}$	0.094	(1.95)	-0.080	(-1.72)
$X_{tram1500}$	-	-	0.127	(2.72)
$X_{natl1000}$	-0.273	(-3.17)	-0.560	(-6.32)
$X_{suburb2000}$	-0.376	(-5.90)	0.122	(2.01)
X_{bus50}	-0.134	(-4.10)	-	-
$X_{mar1500}$	0.256	(4.90)	-0.228	(-3.65)
$X_{air7000}$	-0.397	(-3.17)	-	-
X_{port}	-0.113	(-4.44)	-0.082	(-2.74)
λ	0.522	***	0.379	***
R^2	0.27		0.26	
AIC	25432	(OLS: 27530)	34277	(OLS: 34792)
Moran's I	0.078	***	0.009	***

(·) for $p \lesssim 0.1$, (*) for $p \lesssim 0.05$, (**) for $p \lesssim 0.01$, (***) for $p \lesssim 0.001$

Table 3.24: Demand model Estimations - spatial filtering

	Purchase3		**Rent3**	
	n=10000		n=11663	
	k=4		k=10	
Variables				
Intercept	4.650	(20.53)	11.926	(44.59)
House attributes				
X_{sqm}	0.531	(16.43)	0.753	(22.03)
X_{price}	-0.428	(-19.21)	-1.246	(-39.89)
$X_{\lesssim 50}$	-	-	0.773	(3.86)
X_{80}	0.203	(6.38)	0.207	(6.67)
X_{90}	0.275	(7.11)	0.409	(11.20)
X_{00}	0.096	(3.12)	0.588	(16.13)
$X_{\gtrsim 10}$	0.095	(2.55)	0.781	(15.30)
X_{apart}	-0.396	-9.154	-0.634	(-13.92)
X_{single}	-0.266	-6.40	-0.466	(-9.99)
X_{base}	-0.116	(-2.09)	-0.249	(-4.21)
X_{gf}	reference category for floor			
X_{f1}	0.202	(7.08)	0.202	(6.77)
X_{f2}	0.327	(10.96)	0.292	(8.97)
X_{f3}	0.245	(7.68)	0.336	(9.15)
X_{f4}	0.347	(9.93)	0.518	(11.98)
X_{f5}	0.422	(9.04)	0.533	(9.27)
$X_{\gtrsim f6}$	0.463	(8.49)	0.594	(8.09)
X_{NS}	0.063	(2.25)	0.439	(15.06)
X_{ah}	-	-	0.170	(6.59)
X_{firep}	-	-	-0.137	(-5.07)
X_{ng}	-	-	0.132	(4.07)
X_{safety}	-	-	0.122	(5.12)
X_{new}	0.249	(9.26)	0.288	(8.25)
$X_{excellent}$	0.210	(9.64)	0.101	(4.68)
X_{sea}	-	-	0.090	(3.34)
X_{corner}	-	-	0.180	(8.90)
Transport attributes				
X_{cbd}	0.078	(4.27)	-0.111	(-5.90)
$X_{metro1000}$	0.095	(3.15)	0.08	(2.28)
$X_{isap1000}$	0.065	(2.50)	-0.116	(-3.81)
$X_{tram1500}$	-	-	0.149	(4.93)
$X_{natl1000}$	-0.203	(-4.32)	-0.587	(-9.85)
$X_{suburb2000}$	-2.295	(-8.63)	0.152	(3.79)
X_{bus50}	-0.142	(-5.77)	-	-
$X_{mar1500}$	0.138	(5.06)	-0.257	(-6.43)
$X_{air7000}$	-0.381	(-5.80)	-	-
X_{port}	-0.135	(-10.13)	-0.100	(-5.21)
eigenvectors	1.000	(n=243)	1.000	(n=107)
R^2	0.34		0.30	
AIC	24804	(OLS: 27530)	33775	(OLS: 34792)
Moran's I	0.087	***	0.016	***

(·) for $p \lesssim 0.1$, (*) for $p \lesssim 0.05$, (**) for $p \lesssim 0.01$, (***) for $p \lesssim 0.001$

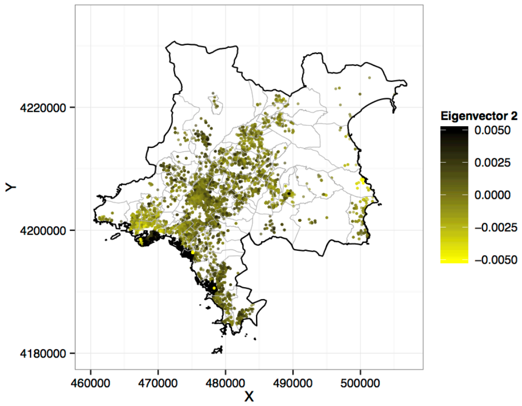
Table 3.25: Top 6 eigenvectors

	Step	Evector	Eval	Moran's I	ZMinMi	Pr(ZI)	R^2	γ
	0	0	0.000	0.435	56.135	0.000	0.094	0.000
	1	2	2.945	0.397	49.894	0.000	0.108	11.655
	2	5	2.414	0.376	46.400	0.000	0.117	9.690
Purchase	3	3	2.704	0.364	44.573	0.000	0.121	6.624
	4	6	2.308	0.358	43.574	0.000	0.124	5.408
	5	18	1.703	0.352	42.710	0.000	0.128	5.945
	6	7	2.209	0.347	41.889	0.000	0.134	6.132
	0	0	0.000	0.128	29.132	0.00	0.223	0.000
	1	1	2.782	0.110	24.479	0.000	0.229	-9.570
	2	4	1.439	0.107	23.781	0.000	0.230	5.334
Rent	3	169	1.030	0.105	23.203	0.000	0.232	-5.759
	4	21	1.152	0.103	22.649	0.000	0.234	5.317
	5	175	1.028	0.101	22.137	0.000	0.236	-5.418
	6	10	1.233	0.099	21.639	0.000	0.237	4.868

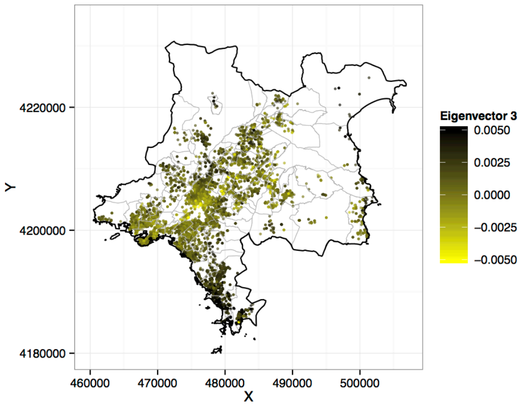
Evectoris the number of selected eigenvector, Eval stands for the eigenvalue, ZMimMi is the standardized Moran's I value, Pr(ZI) is the probability of the standardized deviate for the given value of the alternative argument, and gamma stands for the regression coefficient for the Evector in fit

In the purchase price model, eigenvector 2 (Figure 3.18i) identifies clustering in south-west (yellow color), where the Piraeus port is located, and at the marinas (black color); eigenvector 3 (Figure 3.18ii) in the city center and South Coastline; eigenvector 5 (Figure 3.18iii) in the Northern Suburbs; eigenvector 6 (Figure 3.18iv) successfully identifies clustering around the Attica-tollway (black color); eigenvector 7 (Figure 3.18v) in the northern city center and Northern Suburbs and eigenvector 18 (Figure 3.18vi) is more difficult to be interpreted.

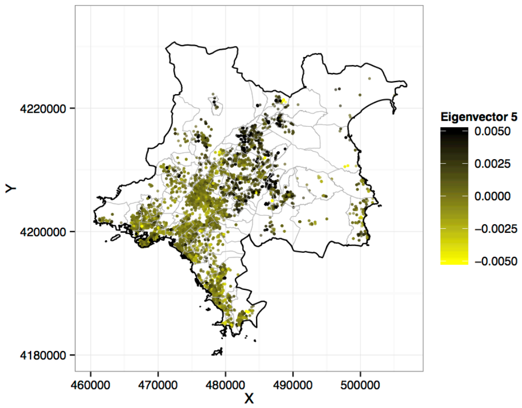
Concerning the rent model, similar to purchase, eigenvector 1 (Figure 3.19i) identifies clustering around the Piraeus port (black color) and marinas (yellow color); eigenvector 4 (Figure 3.19ii) around the Syngrou Avenue, the street that connects the city center with the coastline; eigenvector 10 (Figure 3.19iii) identifies a compact positive cluster in Athens (the city center) and smaller in Northern Suburbs and Southern; eigenvectors 21, 169 and 175 (Figures 3.19iv,3.19v and 3.19vi) are more difficult to be interpreted, however, they could be related with environmental characteristics of transport accessibility.



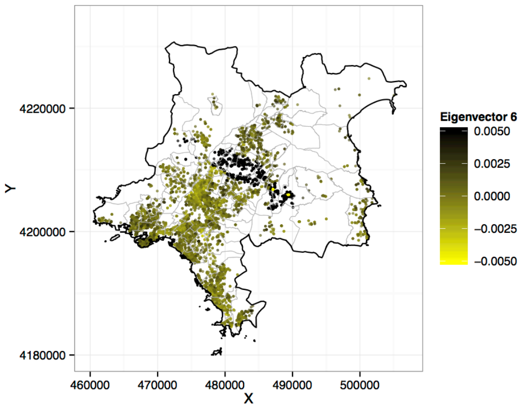
(i) Eigenvector 2



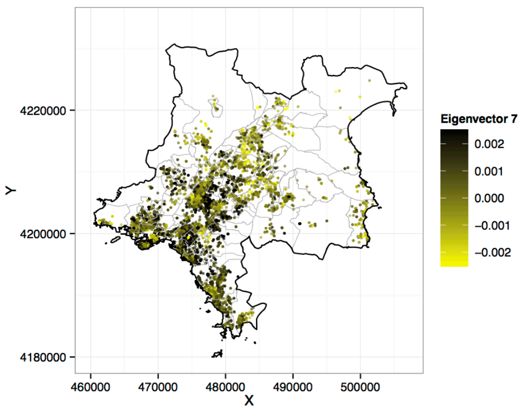
(ii) Eigenvector 3



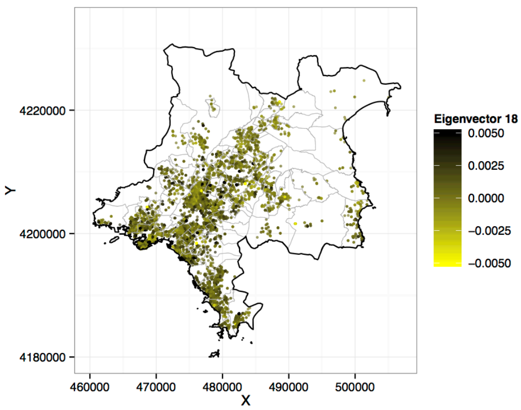
(iii) Eigenvector 5



(iv) Eigenvector 6

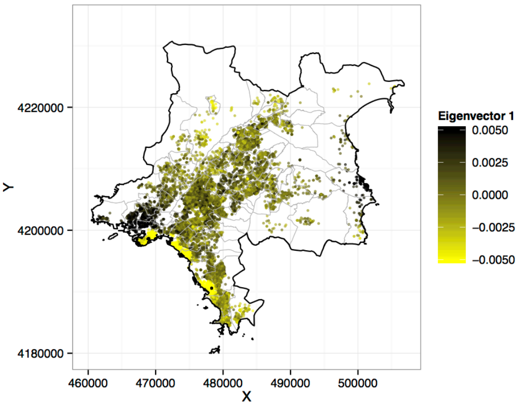


(v) Eigenvector 7

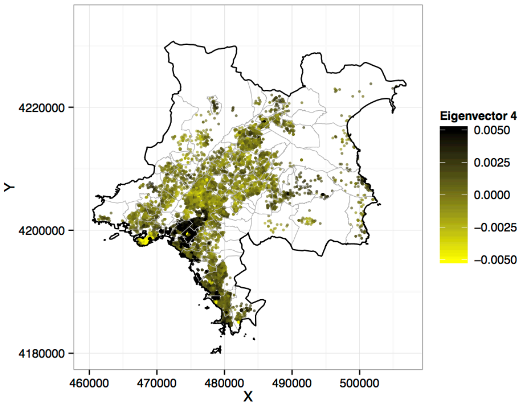


(vi) Eigenvector 18

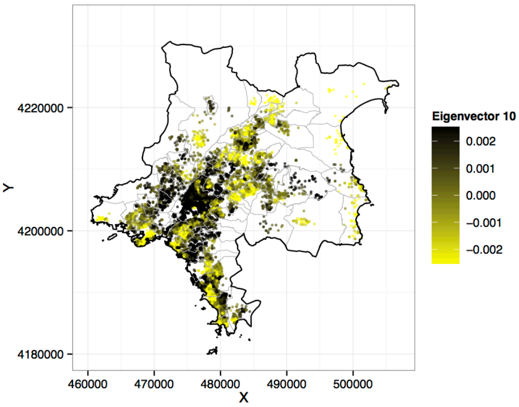
Figure 3.18: Eigenvectors with higher MI - purchase



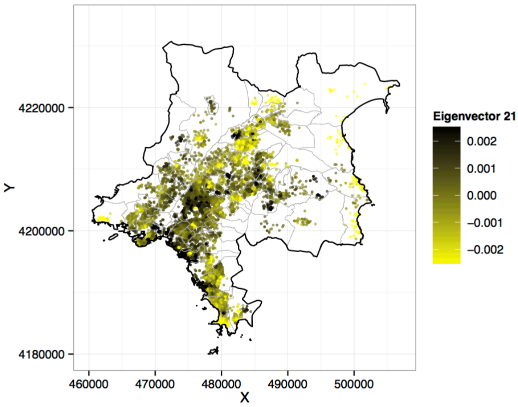
(i) Eigenvector 1



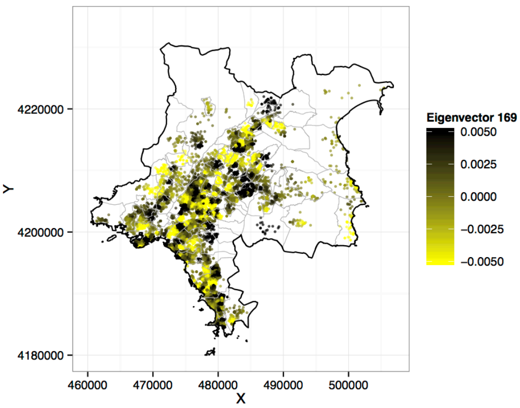
(ii) Eigenvector 4



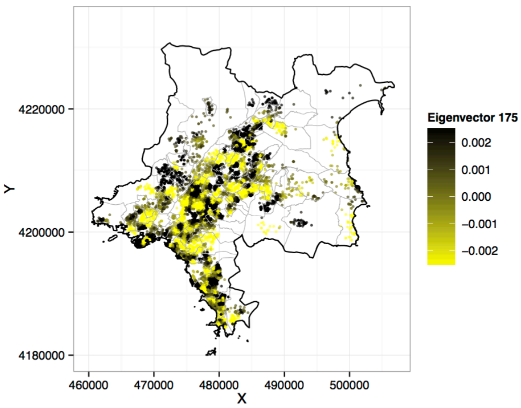
(iii) Eigenvector 10



(iv) Eigenvector 21



(v) Eigenvector 169



(vi) Eigenvector 175

Figure 3.19: Eigenvectors with higher MI - rent

3.5.5 Conclusions

Real estate prices have been highly affected by the current financial crisis that we are experiencing. The impact of transportation infrastructure locations on purchase prices and rents of the dwellings, is a subject that has been widely investigated during the last few years. But how does the crisis affect the values of properties and can transportation infrastructure maintain them at higher levels? In this research, we attempt to give answers to these questions, among others.

The results indicate that overall, from 2011 to 2013, the purchase prices of dwellings in Athens have decreased by -18.2% and rents by -15.2%. On–line real estate data have been used for the analysis, which makes the results somewhat biased, since the real/-transaction prices are likely lower than the advertised/listed. However, the reduction is not expected to be homogeneous and may be a function of the conditions in the area.

An interesting consequence of the crisis is demonstrated by the results of a spatial cluster analysis: The equivalent ‘middle-class squeeze’ effect (i.e. the inflation increases at a higher rate than the wages, impacting the middle–class earners) that occurs in times of recessions, is depicted on the real estate purchase and rent values. More specifically, the more expensive (‘elite’) areas in Athens (Northern Suburbs and South Coastline) maintain a uniformity at high price levels, while those in southern Northern Suburbs, that were expensive before, decline.

The impact of transportation infrastructure locations on real estate prices and rents differs between 2011 and 2013. For example, the impact of metro station (<500m) on purchase prices has declined by -42.5% and on rents by -62.5%; the effect of ISAP (light rail) station (<500m) on purchase prices has been reduced by -53.5%; the impact of airport proximity (<7000m) on purchase prices is lower by -30.3% and -50.7% for rents. However, despite the reduction, the impact has maintained its positive or negative sign, similar to Efthymiou and Antoniou (2013a), depending on the transportation system.

Housing demand, based on on–line ‘search’ data per property, has also been measured. Dwellings located within 1000m from metro stations have higher purchase and renting demand than the rest. Residencies for sale located within 1000m of ISAP stations have higher demand, indicating that people take advantage of the crisis and the degraded conditions around them (due to the externalities) to invest. On the other hand, the demand for dwellings for rent is lower.

The findings demonstrate the direct and indirect effects of the current financial crisis, on real estate prices in Athens, Greece; in particular, the impact variation of transportation infrastructure locations on dwelling values, before and at the peak of the crisis, is explored. While the impact of transportation infrastructure and policies on property values has been extensively investigated before, the research is usually based on static metrics, while the dynamic and spatial element is absent.

The results of this paper verify that the sensitivity of property values to economic conditions varies spatially. As expected, transport facilities –usually being strong determinants of real estate prices– lose part of their impact. It is imperative to exploit the increasingly available spatial and real estate data over time, in transport policy research. Living the reality of an extreme case, the on–going crisis scenario, should be turned into an opportunity for productive research.

3.6 Effects of large scale transportation infrastructure on land-use/cover

In many respects, urbanization is usually seen as a vitality sign of local economies and must be carefully planned respecting natural resources leading to welfare and sustainable development. Attica, the metropolitan area of Athens, is a special case of evolution distinguishing among the other capitals of South Europe. Based on Chorianopoulos et al. (2010), many districts of Attica exhibited ad-hoc patterns of urban expansion in less than 20 years, leading to an urban concentration of over one third of the Greek population. This phenomenon was exacerbated by external political and historical factors, such as the accession of Greece to the European Union and the 2004 Athens Olympic Games, in practice invalidating any sort of initial development planning.

Development of major transportation infrastructure in Attica was not necessarily guided by urban development, but often to serve disparate objectives, such as Olympic venues. Attica Tollway, a half ring-road around Athens, which was completed ahead of the Olympic Games, is probably the most significant highway infrastructure investment in the region. To the East, Attica Tollway connects Athens and other areas to the west, with a formerly less developed part of Attica (Messogia), which has recently seen rapid development, to a large degree because of the new Athens International Airport which was built there (it operated in 2001). Of course, as the move was well-known in advance, the development of Messogia started much earlier. Formerly mostly vacation houses, many residences in the region were quickly becoming primary residences.

It is essential for transportation researchers and planners to uncover the dynamics underlying urban systems (Frazier and Kockelman, 2005). Parker et al. (2003) analyzed a wide range of land use/cover change (LUCC) models and pointed out that owning to the complexity of the systems encompassing land-use/cover, no single existing model dominates and further investigation is needed in order to define the strengths and weaknesses of competing theoretical and empirical models. The first spatial models for urban areas were based on panel data derived from satellite imagery data. Special attention was given to relationships among geographic variables, land use/cover variables and demographic attributes. In many cases, apart from the formation of land use/cover models, demographic models about population and vehicle ownerships were also produced (Frazier and Kockelman, 2005).

The objective of this research is to investigate the impact of the rapid transportation infrastructure development that was held in Athens in recent years, on the urban structure. For the purpose of this research, spatial econometric models that link the impact of transportation infrastructure onto the evolution of land-use, using panel-data (multiple time points), are developed. The land-uses in Athens are modeled for the first time at such disaggregate level; the aim is to explore the potential of developing a LUTI model using spatial econometrics instead of traditional discrete choice, in the future.

3.6.1 Landscape detection through remote sensing techniques

Remote sensing has significantly contributed to reveal land use/cover evolution. The disposal of multi-date remotely sensed images, either of moderate, medium or high spatial resolution, has rendered them the main source of primary data to detect, identify, map and monitor ecosystem changes (Coppin et al., 2004). Classification procedures and constantly developing change detection techniques enhance to more accurate and less time and cost consuming results. The appropriate change detection procedure is highly dependent on the purpose of the study and no single method is suitable and applicable for all cases (Lu et al., 2004).

In general, change detection can be characterized as the process of identifying differences in the state of an object or phenomenon, such as urbanization or vegetation, when observed at different times (Singh, 1989). Supervised per-pixel classification along with the Maximum Likelihood as classifier is the method most extensively used as classification procedure for multi-spectral imagery. Frazier and Kockelman (2005) used thematic maps of Austin extracted through this classifier, as input panel-data for further spatial analysis, where only two out of nine categories were relevant to urban areas. Karathanassi et al. (2000) investigated part of the study area, trying to figure out urban sprawl and urban expansion using high resolution imagery and pixel-based analysis through textural statistical methods which outperformed ML classifier. Later, Chorianopoulos et al. (2010) focused on the evolution of urban areas in Messogeia through two temporal Landsat data –of 1987 and 2003– performed individual supervised classification giving overall accuracies both grater that 90%. Based on this research, the urbanized area was over 12% of the study area, where in general undergoes urban sprawl.

Considering that the training data requirement of supervised procedures is a limiting factor, an advanced pixel-based methodology for change detection has been introduced by Baraldi et al. (2006) free of any training data. Appropriate fuzzy spectral rules were joined for each desirable class, to automate the extraction process of preliminary maps for geometrically and radio metrically corrected Landsat TM and ETM imagery. Nowadays, the difficulty to distinguish urban areas using only spectral criteria coming from training data is commonly accepted, since the heterogeneity of urban areas and the context information are not taken into account (Zhang et al., 2002). Another important perspective is that pixel-based analysis does not represent true nature surface entities which vary in color, shape and texture (Ouma et al., 2008). Thus, per pixel analysis techniques tend to be less effective and in many cases are replaced by Object-Based Image Analysis (OBIA) techniques.

The way in which Object-Based Image Analysis operates actually approaches human brain functions. OBIA, since 2001, has totally changed the classification potential. OBIA has as spatial analysis unit an image object instead of a single pixel and exploits apart from objects spectral information, spatial relations (context) between image objects of the same or different scale, objects texture and additional texture indices, their shape and site and thus, it is more advantageous for identifying urban areas. Based on Baatz and Schape (2000), through OBIA, low-level image analysis techniques, such as segmentation procedure and algorithms are combined with high-level techniques, such as Artificial Intelligence (expert systems, knowledge-based systems using fuzzy logic), and with methods of pattern recognition as well. Benz et al. (2004), had mentioned that low-level image analysis techniques create the primary image objects and by implementing higher level techniques, primitives are converted into semantic objects. Owojori and Xie (2005) used also OBIA on Landsat TM data at Texas city to quantify the urban change. Training data were used to classify four generalized categories (agriculture areas, forests, water bodies, impervious areas) and then supplemental rules based on shape, adjacency and other spatial criteria were used. Confusion matrix for each image was created with remarkable accuracies achieved. Impervious areas increased by 33% to the detriment of forests and agricultural areas.

It is essential that through the classification processes of this study, land cover and not land use data is obtained, because land use data is coming from information about how surface areas are truly used. Both terms are acceptable and are used by derogation with the same meaning.

3.6.2 Data collection

The data used in this research were collected by Siora (2011). The reader should refer to this thesis for a more detailed description of the data collection methodology. Two Landsat 4/5 satellite images of the Attica region, for the years 1984 and 2010, were individually analyzed through remote-sensing techniques for the extraction of land cover data (Figure 3.21). These images provide medium resolution analysis data (30 m x 30 m) and thus they were more useful for this task than higher resolution images. The selected region (Figure 3.22) consisted of 1353x1426 pixels, representing a surface area of 40.59 km x 42.78 km. All imagery data were derived from the U.S. Geological Survey data base.

By converting individual multi-date images to thematic maps through OBIA, a pixel by pixel post-classification comparison can be performed providing contingency matrices of change information. What is more, based on Coppin et al. (2004), it is accepted that the post-classification comparison accuracy is approximately as the product of the classification accuracy of each image involved. Additionally, the comparison after individual classification procedures minimises the impacts of atmospheric, sensor and environmental differences between multi-temporal images rendering the radiometric normalisation of digital spectral units among imagery data a redundant process.

The workflow implemented for each of the multi-temporal data is presented in Figure 3.20. Apart from thematic maps, other output products were statistic results and digital input data at GIS software. To the next subsections, a concise description of the OBIA procedure followed using the Definiens Professional 5 software (Definiens, 2006) is made.

Prior to the development of the models, contingency matrices were extracted offering the information necessary about the transition from the one category to the other. The numbers in Tables 3.26 and 3.27 express pixels and the two possible cases are “urban” or developed, and “not urban” or undeveloped. The diagonal numbers express the pixels that they did not change. The main cases of interest are, of course, those converted from “not urban” to “urban” areas. Table 3.26 shows the results for the entire study area, while Table 3.27 presents the results for the sub area.

For the entire area, the pixels falling from not urban into urban are more than four times greater than the opposite transition (approx. 20000 hectares vs. 5000 hectares) and the new urban areas correspond to the 21% of the study area dry land. While a portion of the transition from urban to not urban may be due to misclassification or mismatching (between the two images), there are legitimate reasons for this to be observed (e.g. from urban areas to parks). Further analysis into this issue revealed that the majority of these points were due to the fuzzy zones for the sea between the two images, producing false transitions to the coastal urban areas. Misclassifications are also produced due to Landsat imagery weaknesses such as the mixed-pixel phenomenon etc. At the sub area (as presented in Table 3.27), the noise is considerably reduced and the transition of interest is really significant and impressive. More specifically, throughout 26 years, at eastern Attica an area of over 10500 hectares has been converted to new urban areas. This is over a 24% of the sub-image dry land area.

It is worth mentioning that 886997 pixels or approximately 79800 hectares of the initial imagery data and 92286 pixels or over 8300 hectares of the sub-area are covered with water. These areas are automatically excluded from the dataset, reducing considerably the number of observations in the models, as well as noise.

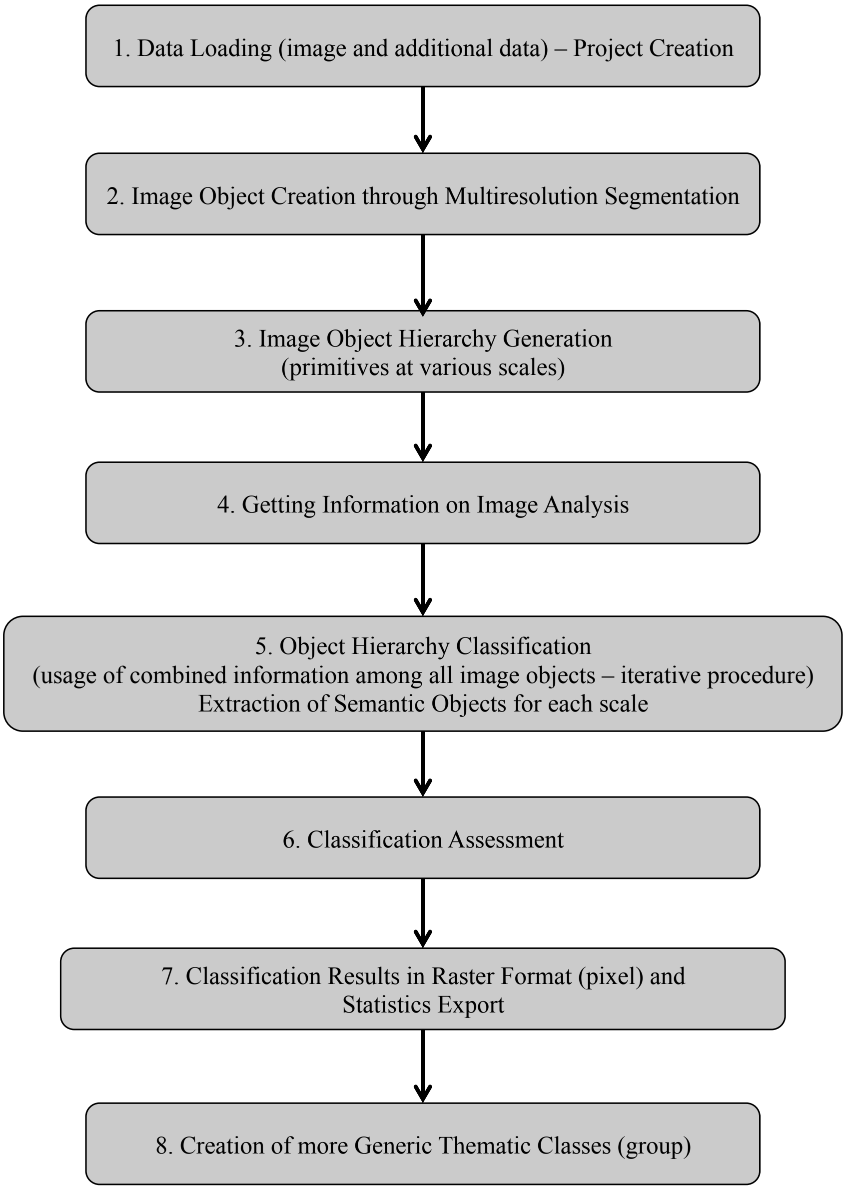
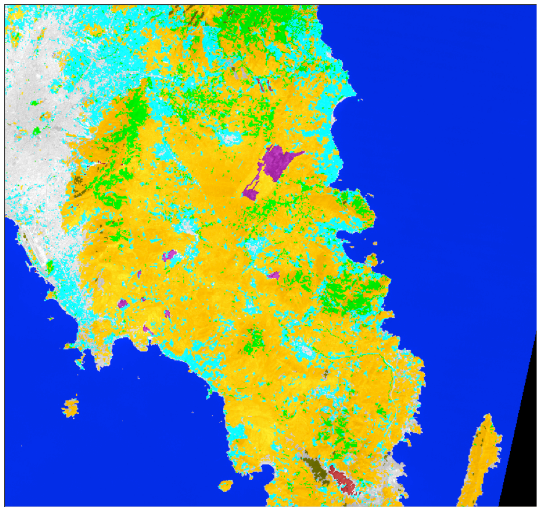


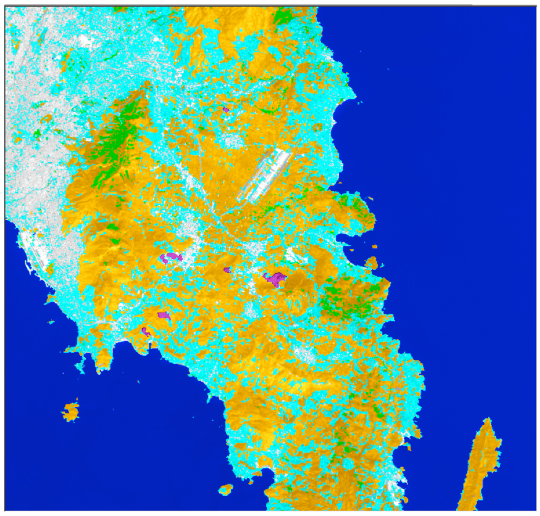
Figure 3.20: OBIA classification workflow followed for each remote sensing image. Adapted from Siora (2011)

(i) Area of Attica on Landsat 1984 imagery data



(ii) Land cover map of Attica for 1984

(iii) Area of Attica on Landsat 2010 imagery data



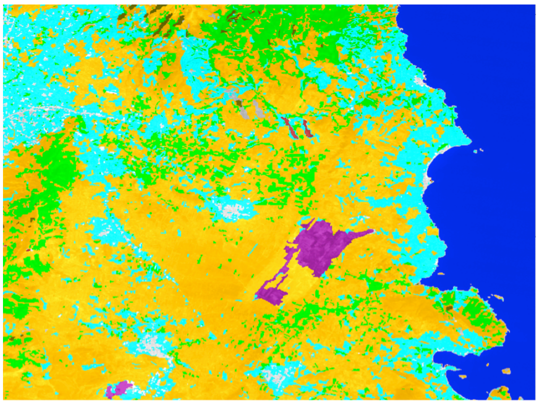
(iv) Land cover map of Attica for 2010

Figure 3.21: Remote sensing images - Attica. Adapted from Siora (2011)

Table 3.26: Contingency Matrix for Developed and Undeveloped Categories in pixels for the Area of Attica. Adapted from Siora (2011)

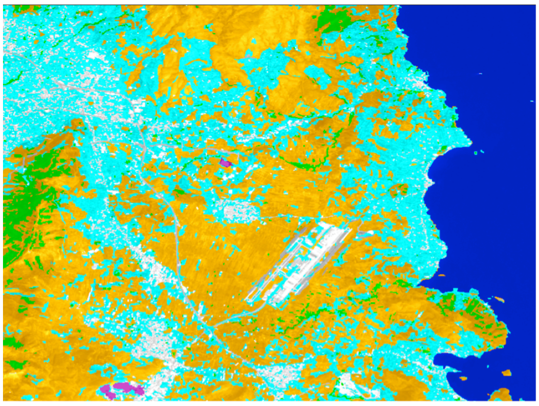
Year		**1984**		
		Not urban	**Urban**	**Sum**
2010	**Not urban**	341091	17337	358428
	Urban	116542	111383	227925
	Sum	457633	128720	586353

(i) ‘Messogia’ Sub-area of Attica (1984)



(ii) Land-cover map for ‘Messogia’ area in 1984

(iii) ‘Messogia’ Sub-area of Attica (2010)



(iv) Land-cover map for ‘Messogia’ area in 2010

Figure 3.22: Remote sensing images - area of interest. Adapted from Siora (2011)

Table 3.27: Contingency matrix for developed and undeveloped categories in pixels for the examined subarea. Adapted from Siora (2011)

Year		**1984**		
		Not urban	**Urban**	**Sum**
2010	**Not urban**	1347395	48641	1396036
	Urban	220844	312498	533342
	Sum	1568239	361139	1929378

3.6.3 Model development

In this research, binary response logistic regression and spatial probit models using panel data were specified and estimated. The objective of these models is to help interpret the urbanization of Attica and mostly of a sub-area in eastern Attica, over the 26 years that are covered by the available data, with an emphasis on the interaction between the development of transportation infrastructure and land-cover change. The models were specified and estimated using the R system for statistical computing (R Development Core Team, 2014).

A logistic regression model for the entire sub-area was first estimated, using all the pixels centroids. For the Generalized Regression Model, the response variable is defined as a binary 0/1 dummy, taking the value 1 if the urbanization level of the corresponding point has changed (presumably moving from undeveloped to developed) and 0 otherwise.

Due to the computation power demanded for the spatial probit model estimation, a sample of 7000 pixel centroids is used for its estimation. The spatial probit models are estimated using three different estimation procedures: the Maximum Likelihood (ML), Generalized Moments Method (GMM) and Bayesian Markov Chain Monte Carlo (MCMC) simulation. Table 3.28 shows the summary statistics of the whole and the sample points of the sub-area, proving that the sample is quite representative. The estimated coefficients, along with the significance level, are presented in Table 3.29.

For each spatial probit model, the weight matrix was constructed by 3 nearest neighbor points to each regression point. Moreover, for the GMM and maximum likelihood spatial probit models, sub-matrices were used per Municipality of Attica. Functions included in the packages “McSpatial” (McMillen, 2014) and “spatialprobit” (Wilhem and Matos, 2013) of R were used for the estimation of the models. The estimation of the Bayesian MCMC spatial probit model run on a server with 64 GB RAM memory, using 1 core, for 3.5 hours.

- Distance from motorways (one Km): binary 0/1 dummy variable, taking the value 1 when the Euclidean distance of the point from the nearest motorway (in 2010) is less than 1 km, and 0 otherwise. This variable captures the effect of proximity to the motorways in the propensity of a location to be further developed.

- Construction of new motorways (roads difference): this variable captures the impact of the development of new motorways (mostly the Attica Tollway). The variable is constructed from the difference of the Euclidean distance from motorways between the two time-points that are considered. The variable is expressed in km.

- Proximity to conurbations (banks): this variable captures the distance of a specific point from conurbations and municipalities, using the existence of banks (in 2010) as a proxy of financial activity. The variable is defined as the Euclidean distance from the closest bank. Distance is also expressed in km.

- Population change (1984-2010): this variable reflects the percent change in population from 1984 to 2010. The population of the municipality in which the modeled location is included is considered.

Table 3.28: Summary statistics for the Area of Attica

	Min	1st Q.	Median	Mean	3rd Q.	Max.
Response	0	0	0	0.238	0	0
Distance to suburban rail stations	11.29	2677.72	4387.40	4772.70	6607.37	12756.19
Distance to highway	0.007	1050.683	2506.359	3030.075	4623.494	10271.011
Distance to points of interest	2.223	317.219	680.220	858.967	1275.295	3483.467
Distance to CBD	5264	12206	16644	16758	21376	31149
Difference of population (ratio) (2010 - 1984)	-20.66	67.69	95.75	127.00	132.83	529.77
Difference of distance from highway (2010 1984)	-11257	-7334	-4443	-4419	-1211	0

				Subset (N=7000)		
	Min	**1st Q.**	**Median**	**Mean**	**3rd Q.**	**Max.**
Response	0	0	0	0.239	0	0
Distance to suburban rail stations	45.08	2696.46	4376.23	4768.99	6559.78	12358.25
Distance to highway	0.092	1063.003	2514.075	3023.973	4574.424	9922.693
Distance to points of interest	4.863	319.510	675.264	856.374	1263.724	3435.137
Distance to CBD	5265	12419	16751	16871	21537	30522
Difference of population (ratio) (2010 - 1984)	-13.08	67.68	95.75	128.40	132.83	529.74
Difference of distance from highway (2010 1984)	-11252	-7341	-4536	-4439	-1237	0

3.6.4 Results

Based on Table 3.26, the results show how the construction of new transportation infrastructure affects the urbanization of the rural Attica. Although, a variable directly related with the proximity to the airport is not included in the models, the three transportation infrastructures are interrelated. The railway and the toll-way were constructed to link the Northern Suburbs with the center of Athens, and both with the airport. Moreover, despite that the Suburban Railway lies within the two directions of the Attica tollway, the impact of the logarithmic distance from the toll-way (and also the railway) differs from the logarithmic distance from the Suburban Rail stations, while both have a negative coefficient (the more the distance, the less the urbanization).

The log-likelihood of the spatial ML probit model is less than the value of the simple generalized linear regression, which shows a better fit. A more direct comparison of the spatial probit models results can be achieved by the computation of the elasticities for each variable. Considering the dependent variable of the spatial probit models as the probability, we attempt to measure the marginal effect of the transportation infrastructure on the urbanization within the examined years. The derivative of the dependent variable, with respect to each explanatory, is its marginal implicit probability of change (Kim et al., 2003)).

Table 3.30 shows the estimated elasticities of the spatial models. It is observed, that they differ depending on the estimation procedure followed. The increase of the population has the highest impact on the probability of land-use change, which is even higher when the model is estimated using the Generalized Moments Method. Moreover, the longer the distance from the CBD, the more possible it is that the land-use has changed from rural to urban. This variable has the second highest impact on the probability to land-use change, meaning that it has the second higher absolute elasticity. The construction of new highways and the suburban railway leads the areas more proximal to their location to become urban. The higher the distance from the points of interest, the less possible it is that the land-use has changed, however, these points have the lowest effect comparing to the other attributes (lower absolute elasticity). Finally, the higher the absolute difference of the distance of the area from highway, the highest the probability of urbanization. Figures 3.23 to 3.26 show the traces, the ACF and the posterior distribution of the MCMC SAR model variables.

Table 3.29: Land-cover change model estimation - regression models

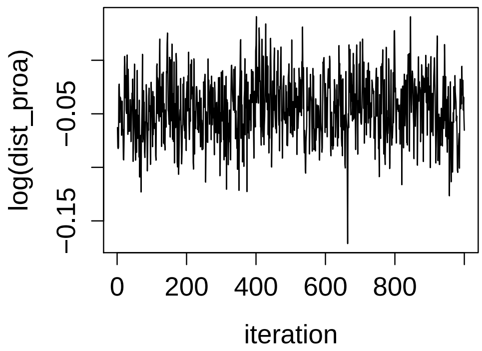
		GLM (whole area) N=473130		GLM (sample area) N=7000	
	Specification	β	**t-test**	β	**t-test**
Distance to suburban rail stations	logarithm	-0.164	-41.33	-0.086	-2.635
Distance to highway	logarithm	-0.205	-95.04	-0.238	13.347
Distance to points of interest	kilometers	-0.179	-48.45	-0.124	4.154
Distance to CBD	logarithm	0.589	116.70	0.512	12.646
Difference of population (ratio) (2010 - 1984)	ratio	0.001	52.91	0.084	5.054
Difference of distance from highway (2010 1984)	kilometers	-0.026	-37.42	-0.0328	-5.718
Log-likelihood		-239494.6		-3579.5	

Table 3.29. Land-cover change model estimation - spatial models (continue)

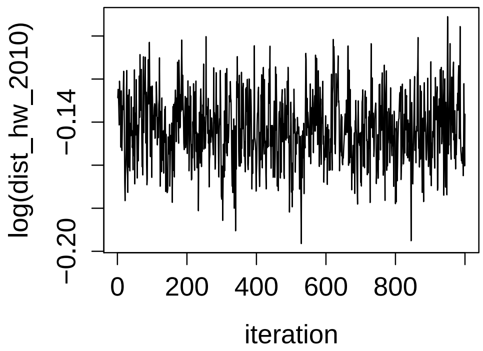
		Spatial probit (sample area) N=7000		Spatial probit (sample area) N=7000		Spatial probit (sample area) N=7000	
	Specification	β	**t-test**	β	**t-test**	β	**t-test**
Distance to suburban rail stations	logarithm	-0.055	-2.029	-0.045	-1.531	-0.049	-1.848
Distance to highway	logarithm	0.178	-6.480	-0.143	-8.889	-0.168	-4.196
Distance to points of interest	kilometers	-0.0001	-3.496	-0.00001	-3.400	-0.0001	-2.927
Distance to CBD	logarithm	0.372	5.751	0.315	9.285	0.343	4.000
Difference of population (ratio) (2010 - 1984)	ratio	0.062	4.045	0.045	2.993	0.057	3.253
Difference of distance from highway (2010 - 1984)	kilometers	-0.023	-4.243	-0.019	-4.034	-0.025	-3.380
ρ		0.321	2.201	0.564	36.500	0.377	1.720
Log-likelihood		-3577.2					

Table 3.30: Land-cover change - elasticities

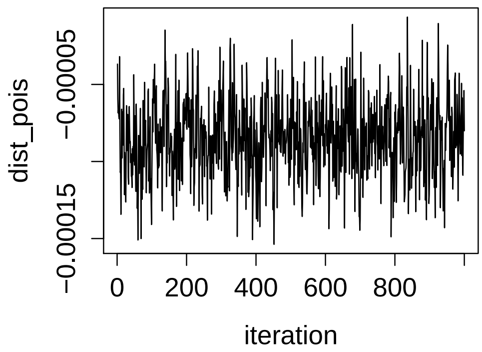
	ML	**GMM**	**Bayesian**
Distance to suburban rail stations	-0.30	-0.38	-0.29
Distance to highway	-0.91	-1.14	-0.94
Distance to points of interest	-0.13	-0.02	-0.01
Distance to CBD	2.32	3.05	2.33
Difference of population (ratio) (2010 - 1984)	11.72	13.25	11.75
Difference of distance from highway (2010 - 1984)	0.15	0.19	0.18



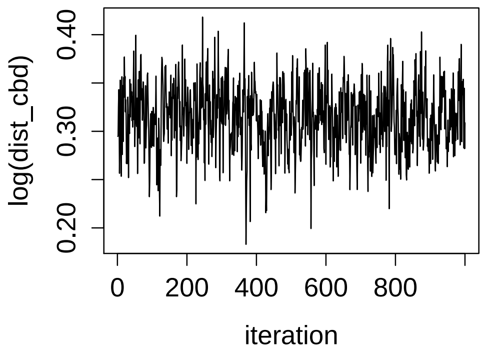
(i) Traces of distance to suburban rail stations



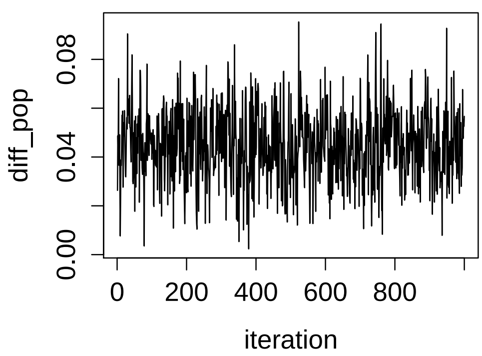
(ii) Traces of distance to highway



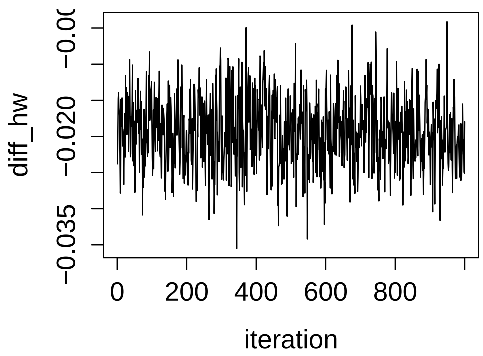
(iii) Traces of distance to points of interest



(iv) Traces of distance to CBD

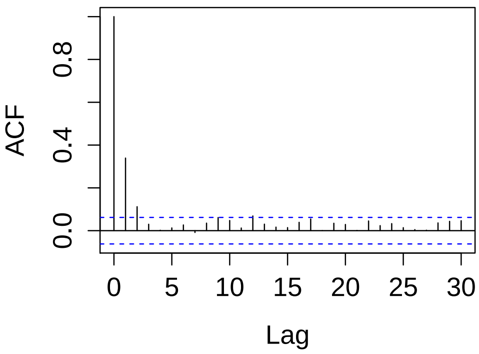


(v) Traces of difference of population

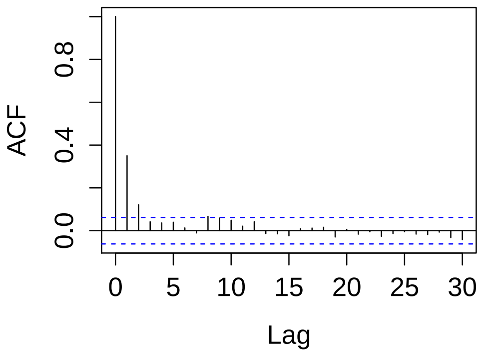


(vi) Traces of difference of distance from highway

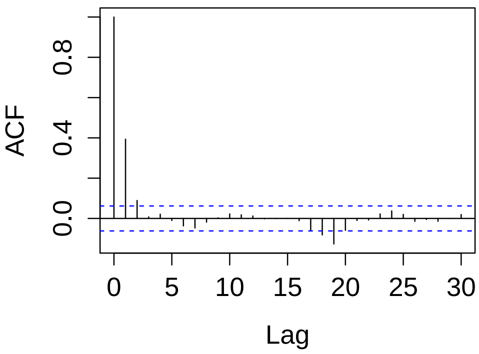
Figure 3.23: Traces of MCMC SAR model variable



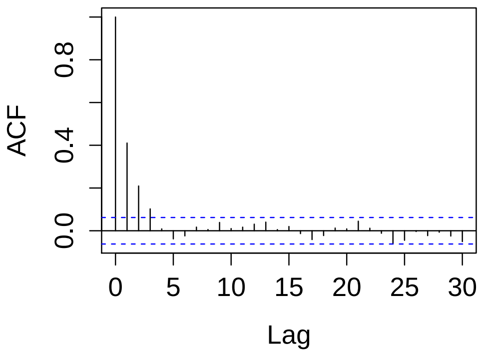
(i) ACF of distance to suburban rail stations



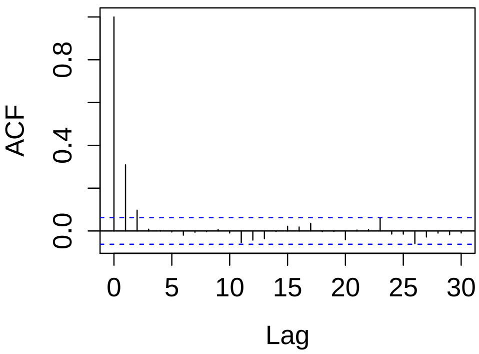
(ii) ACF of distance to highway



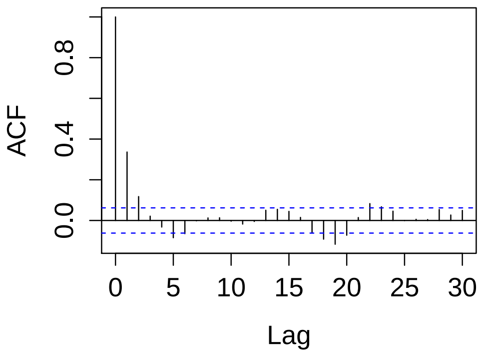
(iii) ACF of distance to points of interest



(iv) ACF of distance to CBD

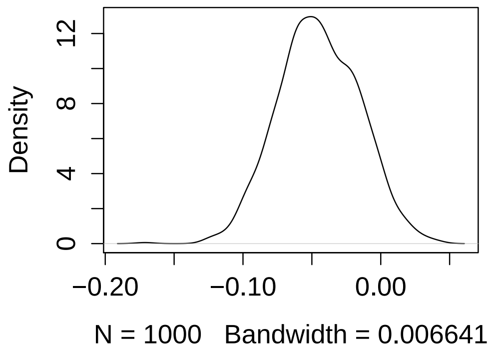


(v) ACF of difference of population

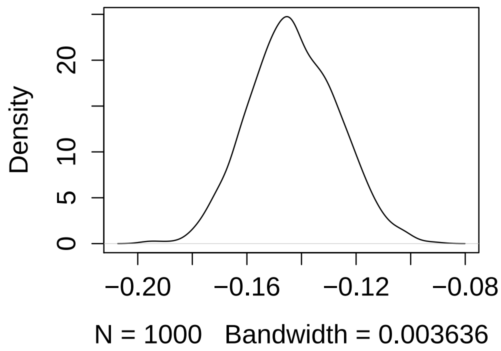


(vi) ACF of difference of distance from highway

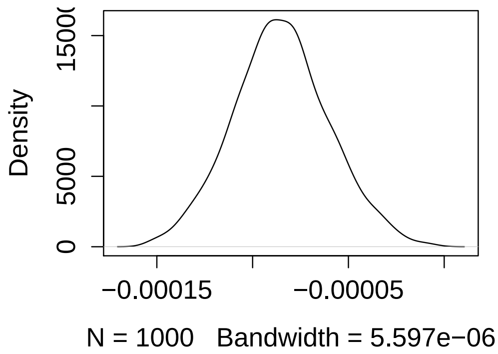
Figure 3.24: ACF of MCMC SAR model variables



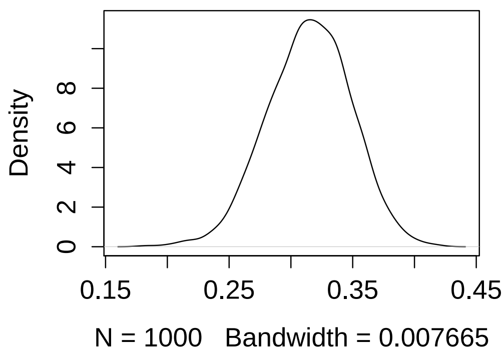
(i) Posterior distribution of distance to suburban rail stations



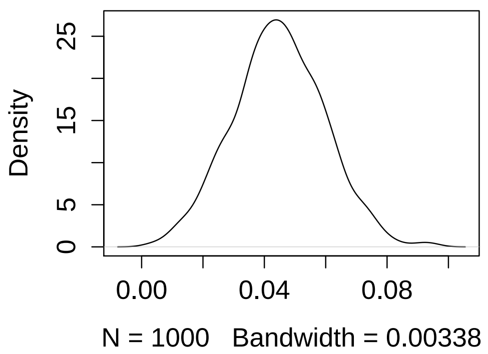
(ii) Posterior distribution of distance to highway



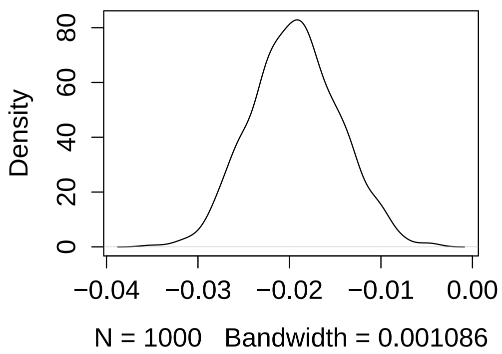
(iii) Posterior distribution of distance to points of interest



(iv) Posterior distribution of distance to CBD

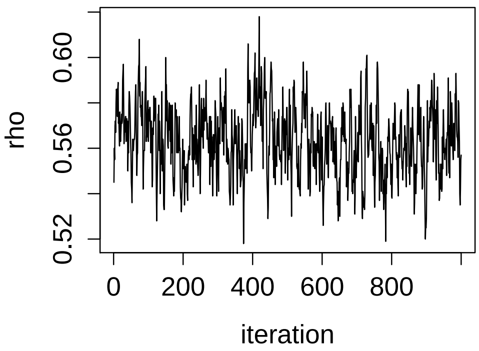


(v) Posterior distribution of difference of population

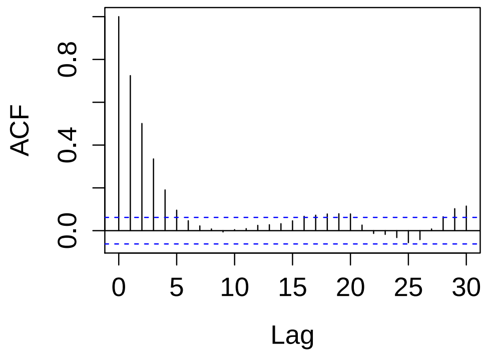


(vi) Posterior distribution of difference of distance from highway

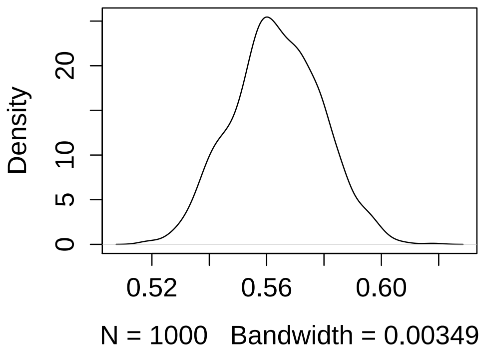
Figure 3.25: Posterior distribution of MCMC SAR model variables



(i) Traces of rho



(ii) ACF of rho



(iii) Posterior distribution of rho

Figure 3.26: Traces, ACF and posterior distribution of ρ

3.6.5 Conclusions

Following the accession of Greece to the Euro-Zone in 2001 and the 2004 Athens Olympic Games, the development of new large-scale transportation infrastructure has shaped urban development. This research presents the development of spatial econometric models relating urbanization trends in the densely populated area of Attica, using panel data from multiple data sources: land use/land cover data derived from Landsat satellite imagery data, census data and further spatial information such as Euclidean distances from transportation infrastructure and points of interest.

Taking advantage of advanced remote sensing methods and more specifically of the Object-Based Image Analysis (OBIA), credible land cover information was extracted for 1984 and 2010 affirming qualitatively and quantitatively the significant landscape change of the area. The Haralick texture Index of Homogeneity had substantially contributed to the extraction of urban areas. It can be concluded that OBIA is highly recommended from now on, for similar research objectives using Landsat imagery instead of conventional per-pixel methods providing very high results to all assessment criteria with no use of any training data to the knowledge base developed.

The estimated global and local econometric models contribute in the effort to interpret the correlation between major road infrastructure projects and urbanization trends. In addition, socioeconomic factors such as population and distances from points of interest entered into the models, in an effort to control for these factors.

The results have shown how the construction of new transportation infrastructure affects the urbanization of the rural Attica. It can be concluded that for all spatial econometric models the longer the Euclidean distance of the pixel from the CBD, the higher the probability of land-use change. Moreover, the construction of the Attica tollway and the suburban railway has led to the increase of the expansion of the urban areas closer to this infrastructure. The computation of the elasticities using the spatial multipliers, allows a more direct comparison of each attribute s impact. The results of the spatial probit models are compared with the Generalized Logistic Regression. The log-likelihood of the spatial ML probit model is less than the value of the simple GLM which shows a better fit.

Ultimately, the results are promising as they provide preliminary answers to the main questions set out in the objectives of the study. Of course there is room for improvement through the inclusion of additional spatial parameters and socioeconomic variables providing to the model even better fit to the data. The potential for future research is naturally large. Using more intermediate images would improve the temporal granularity of the panel data, thus allowing more precise correlation of development events with changes in the rate of urbanization. Furthermore, more elaborate functional model specifications could be used to expose spatial autocorrelations and other hidden relationships.

Forecasting of future urbanization is another interesting problem that could be examined if more dense sequences of images were analyzed. The development of multinomial models (eg. Zhou and Kockelman, 2008) for the prediction of more land-uses –instead of the pair urbanized/rural– is essential for their integration in LUTI models. High resolution images could be used for more accurate land-use extraction. Finally, the development of hybrid choice spatial models in the future, could increase the forecasting accuracy of land-use change.

3.7 Optimal site selection of demand-oriented transport facilities

3.7.1 Introduction

The problem of facility location has been a subject of research for many years. The increasing use of Geographical Information Systems (GIS) over the last two decades has contributed to the development of new techniques for placement of public and private facilities (eg. Church, 2002), such as hospitals, schools, and power plants. Hakimi (1964), for example, suggested an algorithm to locate the switching centers in such a way, that the maximum distance from the farthest community is minimum.

The most widely known technique for positioning is multi-criteria analysis (MCA), introduced by Carver (1991) to find the suitable sites for disposal of radioactive waste. This technique takes into account a variety of factors by assigning weights and then returns a value for each location. The integrated geo-spatial multi-criteria analysis and GIS, provide the user with the appropriate tools to evaluate alternative decisions based on different multiple criteria. Dobson (1979) applied an automatic technique in a GIS environment, which used multiple criteria to evaluate different scenarios of energy facility location. The graphical environment helped to better examine and analyse the results of each siting-scenario. Brandeau and Chiu (1989) made a first attempt to classify and categorize the different kinds of location problems. Ion et al. (2009) applied MCA with fuzzy logic for site selection of electric-car-sharing stations. Current et al. (1990) analyzed 45 papers of different facility location cases based on multi-objective location and concluded that the complexity in modeling the objectives of the location problem is based on four categories: 1) cost minimization, 2) profit maximization, 3) demand oriented, 4) environmental concerns, and their subcategories. MCA has been applied to many other studies so far, such as the location of hospitals (Marks et al., 1991), wind farms (Hansen, 2005), wave energy conversion systems (Nobre et al., 2009) and others.

A different method for facility location based on network analysis was introduced by the integration of Hakimi' s theory of the p-medians, network analysis algorithms and linear programming. This method was first introduced by ReVelle and Swain (1970) and was extended by Rosing et al. (1979) who applied linear programming formulation for siting location, using the p-median problem to minimise the distance of p facilities. Teitz and Bart (1968) and Densham and Ruston (1992) suggested two widely known algorithms for network analysis of location-allocation of facilities.

In this research, two different methodologies for the allocation of demand-oriented transport facilities (kiosks/stations) are presented. The first approach is based on MCA (Section 3.7.3) and the second on spatial econometric models (Section 3.7.4). A case study for the optimal location exploration of a Mobility Management Center in Thessaloniki, Greece, is used for the analysis.

3.7.2 Data collection and preparation

Collection

The data used in this research were collected in the context of the project "Development and operation of a pilot mobility center in the Municipality of Kalamaria in the framework of the project MOBINET". The project was implemented by the Hellenic Institute of Transport and it aimed at establishing a mobility center in the Municipality of Kalamaria (Thessaloniki, Greece, a map of which is shown in Figure 3.27). The objective of the mobility center is to provide services that will assist the mobility of the citizens in the greater urban area.

The main services of the mobility center established in the Municipality of Kalamaria are: point-to-point mobility guidance and support; mobility guidance to predefined points of interest and the region's gateways (port, airport, etc.); information provision about urban transport and points of interest; support for mobility impaired people; and ticketing services for urban and interurban transport. The mobility center is in operation since July 2008.

A questionnaire survey was conducted prior to the mobility centre development in March 2008 aiming to acquire the mobility characteristics in the Municipality of Kalamaria, the needs and requirements of the citizens, the factors that affect the choice of the mode to be used in their trips, their preferences on the services of a mobility center, and other mobility oriented attributes. 600 adult citizens responded to the survey (driving is only allowed to people above 18 years of age in Greece, and excluding younger participants was essential in making the mode choice realistic). The questionnaire was disseminated to a random sample at the data-collection sites.

The survey was organized in two parts. First, a series of questions pertaining to the travel choices of the respondents and the mobility center were asked. Then, demographic and socioeconomic questions were asked, including age, sex, marital status, number of kids, education level and occupation. An advantage of this ordering of the two parts of the survey, i.e., asking socioeconomic questions after the main part of the survey, is that the

Figure 3.27: Map of Kalamaria

respondents are not made explicitly aware of their socioeconomic status when answering the questions, and as such response bias may be reduced. The questions included in the first part of the questionnaire address the following topics:

- Transport mode(s) the respondents use in their daily trips
- Frequency of the use of each transport mode
- Indication of the three most important destinations
- The degree that a list of factors affect the choice of mode the respondents use for their daily trips
- The degree that a list of factors discourage the respondents from using transit services
- Level of effectiveness of several means of awareness (tables of journeys located in the stops, Internet, VMS, etc.) in support of the respondents' mobility
- Usefulness of a series of services of a mobility center

- Frequency in the potential use of a mobility center

In addition to the data collected in the above survey, additional information was collected for the needs of the present research. This information refers to bus routes, average vehicle speed in the study area, number of lanes in the road network, distances from key facilities (stadiums, parking stations, etc.).

Preparation

The first phase in the conducted research included the variables selection and data preprocessing. The most important parameters need to be identified and collected to the same reference level. In this research, these variables are the number of population, number of bus stations, number of bus routes, points of interest, total length of road network around the reference points, distance (logarithmic) from shopping center, distance (logarithmic) from stadium, distance (logarithmic) from parking charging schemes, average speed, average number of road lanes. The weights of the selected locations need to be allocated, with smoothing decay function, to street intersections. The logarithm of the main distance variables was considered as this captures the perceived differences in distances in a more natural way. This is a common transformation for entering distances in model specifications and reflects the non-linearity in the perception of distances.

The process was performed using ArcGIS 10 (ESRI, 2011) for data processing and transformation, and R (R Development Core Team, 2014), using the functionality included in the packages "maptools", "spgwr", "spdep" (Bivard et al., 2008), for data analysis and "ggplot2" (Wickham, 2014) for visualization.

The variables used in these research are summarized in Table 3.31 and explained below. Columns 2-5 of Table 3.31 refer to four candidate locations of the mobility management center, as per the original project. The data have been defined based on the street intersections in the study area. Whenever needed, the assumption of a coverage radius of 200 meters was made. This assumption is consistent with a reasonable walking distance for mobility services.

- Residential population: The total number of residents who live 200 meters around the street intersections was computed. Population provides a quantification of the service's demand.
- Number of points of interest: The total number of points of interest located 200 meters around street intersections. The concentration of people around them in peak-hour is increased and it suggests a demand indicator.
- Number of bus routes: The total number of bus routes 200 meters around street intersections. Correlation with the number of bus stations was measured to be

approximately the same all around the area, indicating that every station is being used by about four bus routes. As a result, the number of stations was not included as separate variable in the dataset.

- Total length of road network: The total length of road network 200 meters around the street intersections.
- Average income: The study area was divided in four income zones. The average income 200 meters around the intersections was calculated.
- Average speed: The average speed limit for vehicles 200 meters around the intersections.
- Average number of road lanes: The average number of street lanes 200 meters around the intersections.
- Logarithmic distance from the commercial way: The Euclidian distance of each intersection from the "Kapodistrion" shopping pedestrian was first computed. Since the demand of this 'line of interest' increase disproportionate with the distance, its logarithmic value was used.
- Logarithmic distance from stadium: Same methodology as the "logarithmic distance from the shopping center".
- Logarithmic distance from parking: Same methodology as the "logarithmic distance from the shopping center".

Kernel density (Siverman, 1986) had been applied to the weights of the selected by the respondents points, so as to adjust, by smoothing, the initial z values (score) to points within a radius of 700 meters around the candidate positions. Kernel density applies a symmetric probability density function (normal density or weights) to naive estimator. The result was then normalized to be the same as the input "score" of the survey.

Table 3.31: Values of variables on suggested locations

Variable	**KP**	**CSC**	**CBS**	**CS**
Number of points of interest	30	2	5	4
Total population (*100)	31.25	35.32	31.65	26.39
Average income (*1000)	15.33	13.83	15.19	13.83
Total number of bus routes	10	2	5	11
Length of road network (*1000m)	43.21	43.33	34.00	28.24
Logarithmic distance from stadium	5.79	6.62	7.32	7.31
Logarithmic distance from commercial way	0	6.20	6.99	6.99
Logarithmic distance from parking	0	6.00	6.18	6.97
Average speed	45	44	48	47
Number of lanes	1.23	1	1.40	1.08
Score (%)	43.03	20.08	13.93	9.84

KP: Kapodistrion Pedestrian
CSC: Citizen Service Center
CBS: Central Bus station
CS: Central Square

3.7.3 Multi-criteria analysis (MCA)

Introduction

In this section, an “open” model for optimal site selection of transport infrastructure locations is developed. An overview of the methodology followed is presented in Figure 3.28. The available data were evaluated in order to determine those more relevant to the problem at hand. The collected data were processed to detect and rectify possible inconsistencies. Data was then rescaled so that they could be used together in the same multi-criteria framework. A scale from 1 to 10 was selected for all variables.

The optimal allocation algorithm was then formulated as a multi-criteria optimization problem. Like multi-objective optimization, multi-criteria is generally solved by combining different variables into one-scale overlay. Sensitivity analysis is then performed, in which the weight of each objective is allowed to vary. Although weights are allowed in general to range between 0% and 100%, some restrictions can be set based on a priori expectations, according to the expected impact of the variables on the results. The variables used in this stage are: “total population”, “total number of points of interest”, “average income” and “distance from parking”.

The model’s formula is:

$$AP_i = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \tag{3.11}$$

Where, y represents the score for each scenario, x_i is the variable and α_i is the coefficient (percentage/weight) of each variable that maximizes the perception: α_{pop} = total population, α_{poi} = total number of points of interest, α_{dpar} = distance from parking, α_{inc} = average income.

The resulting multi-criteria optimization can be written as:

$$AP_i = f(\alpha_{pop}, \alpha_{poi}, \alpha_{dpar}.\alpha_{inc}, pop_{it}, poi_{it}, dpar_{it}, inc_{it}) \tag{3.12}$$

Where, pop_{it} = total number of population 200m around the position i, poi_{it} = total number of points of interest 200m around the position i, $dpar_{it}$ = average logarithmic distance of position i from shopping way, inc_{it} = average income 200m around the position i.

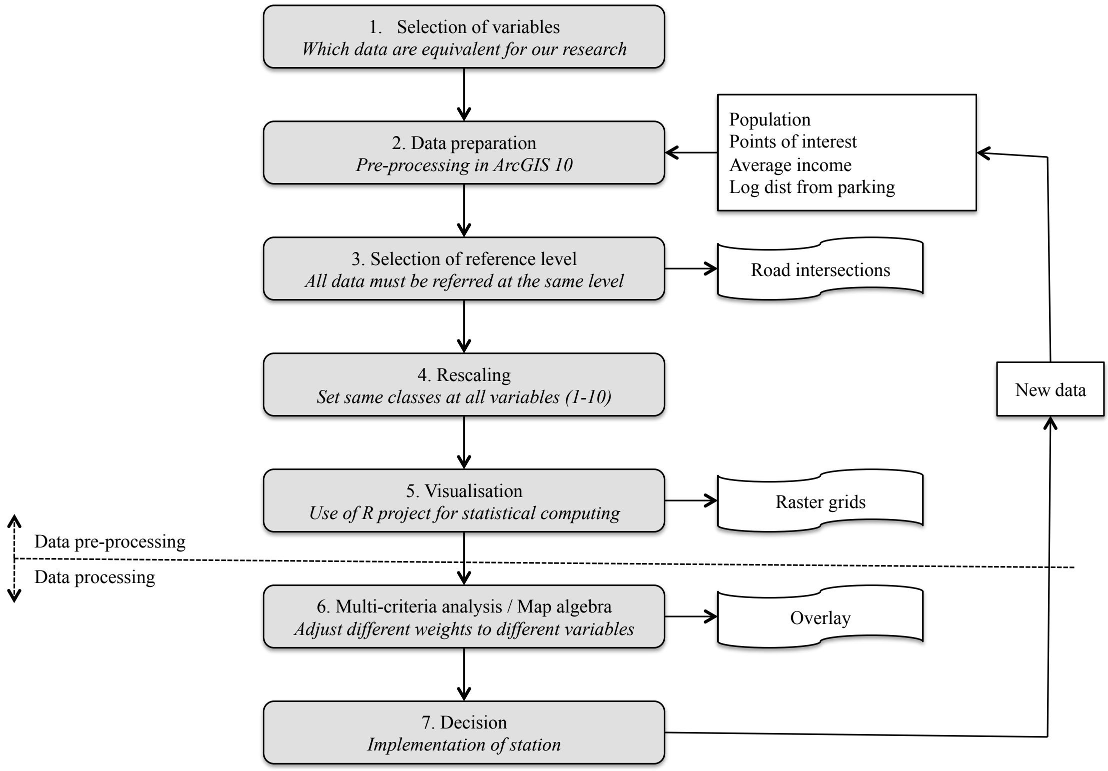


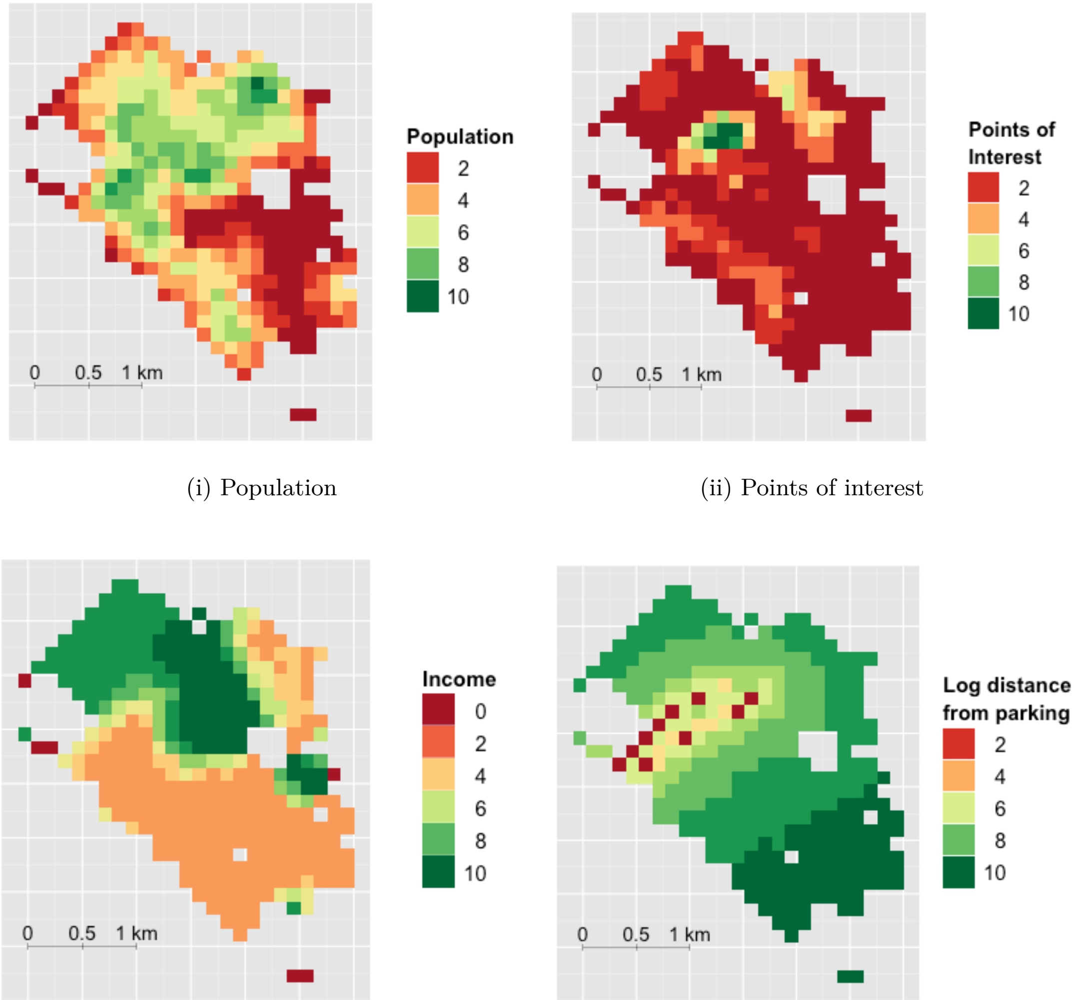
Figure 3.28: Flow diagram of MCA methodology

Results

One of the key questions in every multi-objective problem formulation relates to the weights assigned to each of the components. In this case, instead of determining a priori exact values, reasonable ranges of values have been considered for each attribute. Population is directly related with the demand factor, so a higher range (20-40%) was considered. Electric car owners are expected to be wealthier, so the range for income but also for logarithmic distance from parking was considered between 20-30%. On the other hand, the number of points of interest and the distance from parking may be less important, so a range of 20-25% was considered for both. Considering the constraint that the four factors should add up to 100%, the experimental design consists of the 13 feasible combinations/overlays shown in Table 3.32. The vector attributes were rescaled in a common scale, from 1 (minimum) to 10 (maximum), so as to have comparable and interpretable results, regardless of the scale of each parameter. The lower value of each original variable was assigned to 1, the highest to 10 and intermediate values were considered by selecting equal intervals. Then, the rescaled vector data were converted to raster with grid step ($25x30$ pixels), so as to obtain a complete colored-surface of the area. Figure 3.29 shows the dispersion of the variables around the surface of interest.

Table 3.32: Sensitivity analysis

Overlay	**Population (%)**	**Points of interest (%)**	**Income (%)**	**Distance from parking (%)**
1	20	20	30	30
2	20	30	20	30
3	20	30	30	20
4	30	30	30	10
5	30	30	20	20
6	30	30	10	30
7	30	20	30	20
8	30	20	20	30
9	40	30	20	10
10	40	30	10	20
11	40	20	20	20
12	40	20	10	30
13	40	20	30	10



(i) Population

(ii) Points of interest

(iii) Income

(iv) Logarithmic distance from parking

Figure 3.29: Distribution of key explanatory variables (normalized)

Sensitivity analysis

One of the issues with a sensitivity analysis involving a large number of reasonable alternatives, is to develop a decision rule for making a final selection. Violin plot, which is a combination of a box plot with kernel probability density (indicated by the width of the "violin" shape), is suggested for the representation of the results, because the statistical information needed for the decision is all included in this graph. The outcome of the sensitivity analysis depicted in Figure 3.30 shows that the values of the overlays have different distributions. It is reminded that an overlay is the combination of the four variables used with different weights. According to the results, overlays 6, 2, 12, 8 and 1 have higher density of data points with score below 0, because of the high weight of logarithmic distance from parking. A low population coefficient (20%) results in violins "narrow" around the mean value and "wide" at the upper and lower quartiles (overlays 1, 2, 3), which are not preferred. The combination of a high population coefficient (40%), a significant high weight for income (30%), a relatively low weight of points of interest (20%) and a negative, with low absolute value, distance from parking (-10%), gives a more uniform density of data points which cover a high range of values. Overlay No. 13 satisfies the above criteria and is selected for our purpose. When additional data are available (e.g. electric car ownership, demographic characteristics, locations of jobs or households), they should be counted in step 2 (Figure 3.29) and the procedure should be re-evaluated so as to achieve the optimal expansion of the network.

Having completed the selection of the preferred overlay, the actual question is determining the proposed locations for the electric-vehicle charging stations. Figure 3.31 presents the selected overlay in a combined graph that presents the estimation results as contours, but also indicates the density across the two axes, as densities along the x and y axes.

Findings

In this research, a method for optimal site selection for transportation infrastructure location is proposed and a case study in the municipality of Kalamaria, Greece, is presented. The methodology is based on the Multi-Criteria Analysis that was first introduced by Carver (1991). Four variables were combined and appropriately weighted for the purposes of the sensitivity analysis: "total population", "total number of points of interest", "average income" and "distance from parking". The results of this analysis showed that the optimum location of electric chargers is largely affected by the population and income, to a lesser extent by the availability of points of interest nearby, while it is negatively affected by the distance from parking places. The reliability of the above results, and as a consequence of the optimum location of electric chargers, depends a lot on the availability of a wide range of diverse data, such as electric car ownership, demographic characteristics, and location of jobs or households. The proposed method was tested for a certain type (variables) and amount of data. The importance, however, in this research is that it was demonstrated how the optimum location of electric

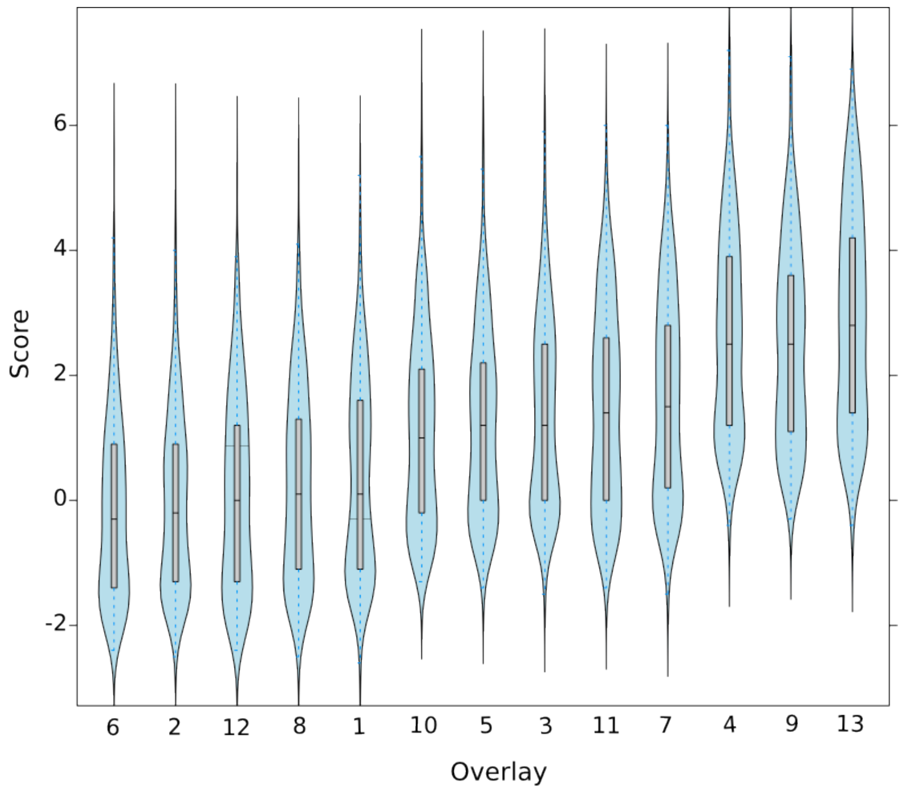


Figure 3.30: Comparison of different MCA scenarios

chargers can be investigated and determined using and combining several methods, i.e. multi-criteria analysis, GIS and sensitivity analysis. The proposed method has a lot of application fields beyond electric chargers. It can be used for determining the optimum location of other demand oriented services or facilities, such as mobility centers, info points, park and ride stations, bike and ride stations, etc. It can be proved a useful tool at the hands of transportation and infrastructure planners in order to maximize the use of such services and facilities.

3.7.4 Spatio-econometric approach

Introduction

The aim of this research is to develop a methodology that can suggest the most suitable location of a mobility center so that it better serves the residents and visitors of a city in their mobility choices and needs. The methodology is based on spatial econometric models. Three different types of models (linear regression, GWR and SAR) are developed and estimated to find the optimal location of a mobility center. Several criteria were used for this analysis, including points of interest, population, income, and bus routes. The data used was collected in the framework of a research project in Greece

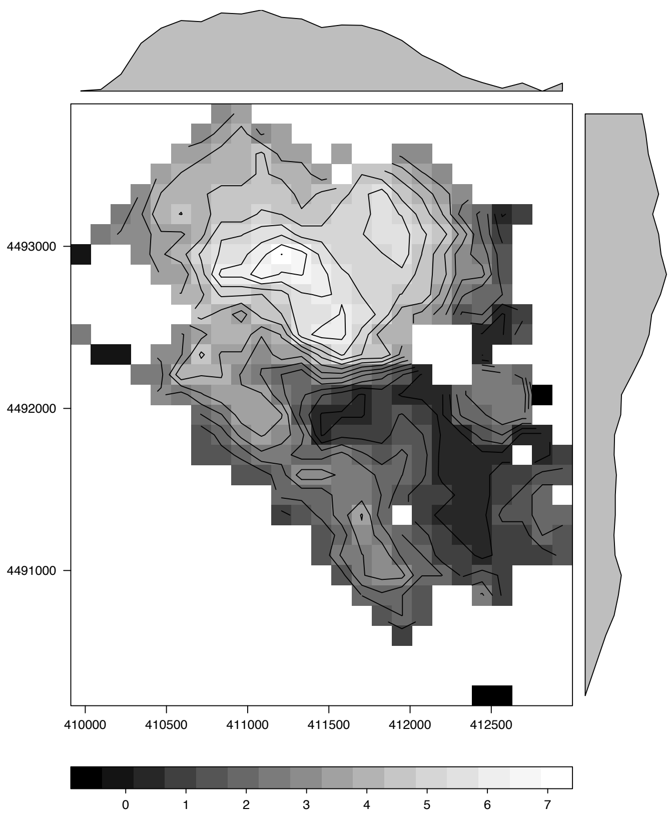


Figure 3.31: Selected overlay (No 13)

during the development of a mobility center in Kalamaria (Thessaloniki, Greece). The paper provides a discussion and recommendations on how the proposed method can be practically used by local authorities in the choice of a mobility center location. It is noted that the proposed methodology is generic and can also be used for other transportation applications, such as car/bike-sharing stations or electric chargers, as these systems have similar characteristics

Model estimation and results

A linear regression model has been developed, with dependent variable the percentage of respondents that selected each candidate location. The results are summarized in Table 3.33. The model results are statistically significant and explain about 79% of the preference of people about the location of the mobility center. The relationships between the attributes are supported by the data and all variable indicators are statistically significant. In addition, as it was expected, income, distance from the commercial center, from controlled parking facilities and from the stadium, but also the average number of lanes, are negatively associated with the perception of optimal position.

Before examining the exact model results, it is important to consider some implementation details. In particular, the various vector attributes were rescaled in a common scale, from 1 (minimum) to 10 (maximum), so as to have comparable and interpretable results, regardless of the scale of each parameter. The lower value of each original variable was assigned to 1, the highest to 10 and intermediate values were considered by selecting equal intervals. Therefore, each location can be associated with a "score" of 1 (lowest) to 10 (highest). This score can be used to interpret the estimated coefficients.

The selection of the mobility center location, increases with the population, by 1.7 "score" points for every 1000 additional population, with the number of points of interest (1.29 points every one point of interest), with the number of bus routes, average speed and length of road network. The OLS (global) model gives significant results, however, the non-stationarity of its variables is not considered. For that reason, local models are also estimated. The results of the SARmix model are presented in Table 3.34 and of GWR in Table 3.35.

Table 3.33: Linear regression results - optimal location

Variable	**Estimate**	**t-value**	**Pr**
Number of points of interest	1.291	12.54	0.00
Total population (*100)	0.174	8.36	0.00
Average income (*1000)	-0.747	-6.29	0.00
Total number of bus routes	0.221	5.60	0.00
Length of road network (*1000m)	2.926	11.23	0.00
Logarithmic distance from stadium	-4.50	-14.65	0.00
Logarithmic distance from commercial way	-0.72	-7.33	0.00
Logarithmic distance from parking	-0.842	-5.48	0.00
Average speed	1.401	19.13	0.00
Number of lanes	-18.488	-12.18	0.00

Residual standard error: 6.442 on 1428 degrees of freedom
Multiple R-squared: 0.791
Adjusted R-squared: 0.789
F-statistic: 541.2 on 10 and 1428 DF, p-value: 0.00
AIC: 9450.147

According to the SARmix model estimation results, when the speed is reducing, the tendency to select the site is increasing, and the distance from the commercial center is not of statistically significant.

Considering the GWR application, the inadequacy of fixed-kernel method was expected, as the density of data points over the area of interest was not uniform. For that reason, adapted kernels were used for the estimation. The optimal number of neighbour points was calculated by the minimization of the AIC. This method suggested that 11 neighbour data points (or the 0.0081% of the total) are needed. To verify the model's accuracy, GWR models with different values of fixed or adaptive kernels were estimated. The

Table 3.34: SDM model results - optimal location

Variable	**Estimate**	**t-value**	**Pr**
Number of points of interest	0.151	4.34	0.00
Total population (*100)	0.051	5.41	0.00
Average income (*1000)	-0.085	-4.14	0.00
Total number of bus routes	0.060	5.44	0.00
Length of road network (*1000m)	0.422	4.56	0.00
Logarithmic distance from stadium	0.480	3.12	0.00
Logarithmic distance from commercial way	0.01	0.471	0.64
Logarithmic distance from parking	-0.086	-2.57	0.01
Average speed	-0.105	-2.19	0.03
Number of lanes	-1.151	-1.84	0.07
Multiple R-squared: 0.791			
AIC: 3732.2			

results (Table 3.35) suggest that the AIC minimisation indeed led to the optimal fraction of points that should be considered. The results in Tables 3.36 and 3.37 correspond to that model.

The significance of spatial variability is examined by the Leung test F3 (Leung et al., 2000). The results show that all the variables apart from the number of lanes and logarithmic distance from controlled parking facility, are spatially inhomogeneous. GWR has higher R^2 than the linear regression model and lower residual sum of squares, two factors that give it an advantage. It is noted, however, that these goodness-of-fit measures should be treated with caution, as explained in (eg. Fotheringham et al., 1998), as they are local, approximate versions of the regression statistics and the do not always behave in the same way. The results of Moran's I are shown in Table 3.38. The data exhibit positive spatial autocorrelation when those with high values are located nearby on the surface and those with low values are also located nearby, while negative spatial autocorrelation is exhibited when both data with high and low values are located near each other Fotheringham et al. (1998). Data of the same dataset can have different values of spatial autocorrelation, both negative and positive. The global model ignores these differences of spatial autocorrelation, which sometimes can be strongly positive or negative on different region of same surface. Moran's I test is a localized version of spatial autocorrelation Fotheringham et al. (1998). The autocorrelation of the residuals was tested for all models, considering weight matrices of the k-nearest-neighbours, where k=11 the value that resulted by the AIC minimization for the GWR model.

While SARmix achieves the lower AIC which indicates the goodness of fit, but also the lowest (pseudo) R squared, it doesn't solve the autocorrelation problem. Despite the high R squared of the OLS model, all three parameters are worse than the local-spatial models. GWR is the equivalent for this methodology, as it solves the autocorrelation problem and results in low AIC and high R squared values and its use is recommended for further research. Figures 3.32 and 3.33 present the estimated coefficients of variables that are spatially non-stationary in the GWR model with 11 neighbours. Figure 3.34 presents a comparison between the predicted values of the global (OLS) (Figure 3.34i) and the suggested local (GWR) (Figure 3.34ii) model. There seems to be a higher concentration of high values in neighbour points within a region, something that makes the selection of the location, easier. Figure 3.34iii shows that local R squared values are high and almost uniform around the surface. It is interesting to note that following an empirical analysis and decision process, the mobility center has actually been located since 2008 in the Citizen Service Center (CSC, shown in Figure 3.34). Furthermore, from the model estimation results shown in Figure 4, it becomes evident that the empirically chosen location is very close to the model results. This provides additional support to the observation that the model is capable of capturing the problem aspects.

Table 3.35: GWR model results - optimal location

Bandwidth	**Number of points**	**AIC**	**R-squared**	**Sum of squares**
	11	4504.48	0.993	1034.660
Adaptive	43	6443.63	0.962	6429.69
	71	7131.73	0.936	10920.79
	150	5375.41	0.985	2462.37
Fixed	200	6136.94	0.971	4856.20
	500	8135.68	0.867	22835.91

Table 3.36: GWR model coefficient ranges - optimal location

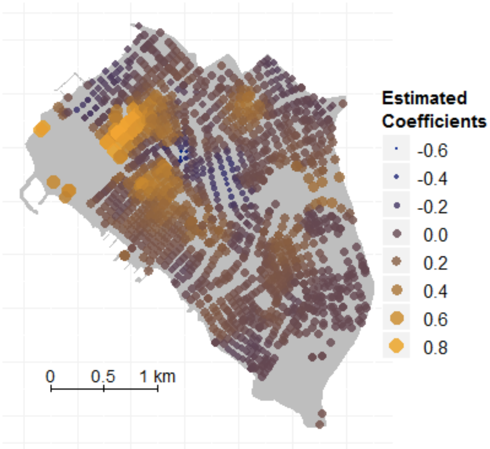
	Min.	**1^{st} Qu.**	**median**	**3^{rd} Qu.**	**Max**
Number of points of interest	-1.922	-0.045	0.145	0.562	2.868
Total population (*100)	-0.616	-8.439e-04	0.066	0.149	0.815
Average income (*1000)	-1.429	-0.454	0.458	4.414	6.486
Total number of bus routes	-1.183	-0.053	0.019	0.090	0.859
Length of road network (*1000m)	-5.835	-0.913	0.001	1.070	17.920
Logarithmic distance from stadium	-70.640	-9.393	-0.001	9.676	132.100
Logarithmic distance from commercial way	-126.700	-11.320	-0.484	2.937e-06	89.890
Logarithmic distance from parking	-101.500	-3.644	-0.282	0.513	35.570
Average speed	-23.920	-0.320	-4.155e-04	0.547	9.277
Number of lanes	-99.260	-8.109	-0.215	5.350	258.500
Quasi-global R squared: 0.993					
AICc: 4504.5					

Table 3.37: Spatial non-stationarity test - optimal location

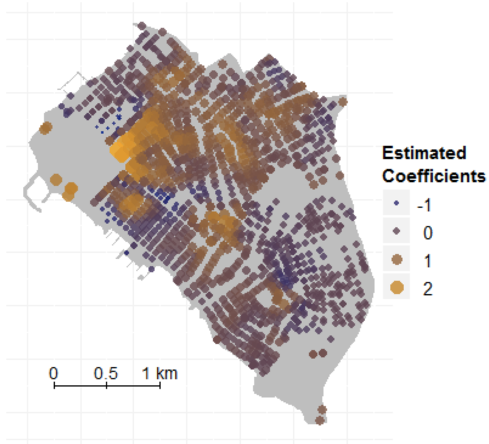
	F statistic	Numerator DF	Denominator DF	Pr(>F)
Number of points of interest	2.43	203	1166	0.00
Total population (*100)	4.83	118	1166	0.00
Average income (*1000)	5.46	71	1166	0.00
Total number of bus routes	5.09	345	1166	0.00
Length of road network (*1000m)	19.96	517	1166	0.00
Logarithmic distance from stadium	2.08	54	1166	0.00
Logarithmic distance from commercial way	5.57	93	1166	0.00
Logarithmic distance from parking	0.90	53	1166	0.00
Average speed	13.54	100	1166	0.00
Number of lanes	0.85	16	1166	0.619

Table 3.38: Model estimations comparison - optimal location

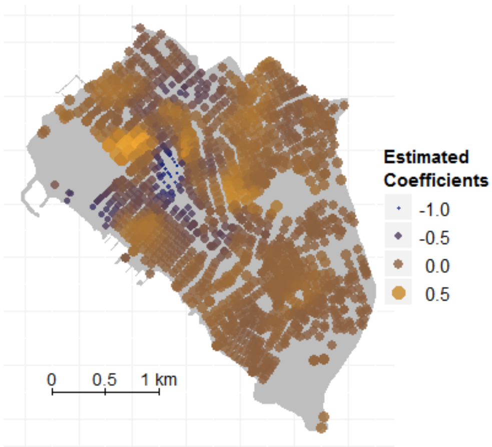
	AIC	R-squared	Moran's I
OLS	9450.147	0.7898	0.678
GWR	4037.89	0.9934	0.211
SARmix	3732.200	0.9936	0.313



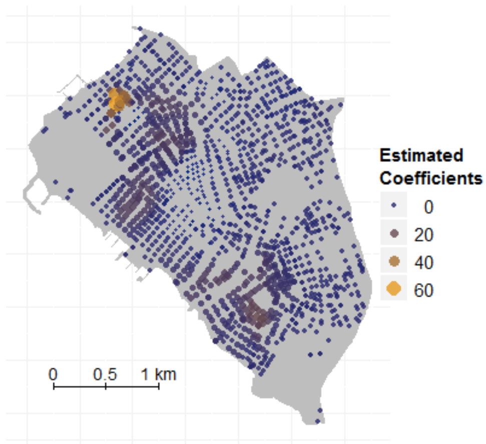
(i) logarithmic distance from stadium



(ii) logarithmic distance from pedestrian shopping zone

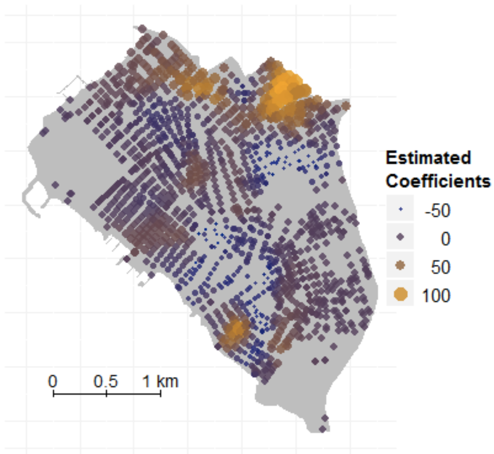


(iii) average speed

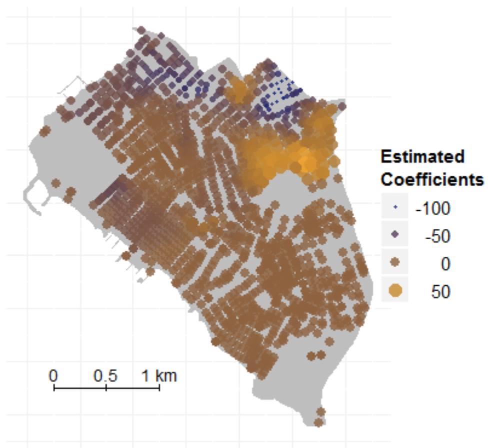


(iv) length of road network

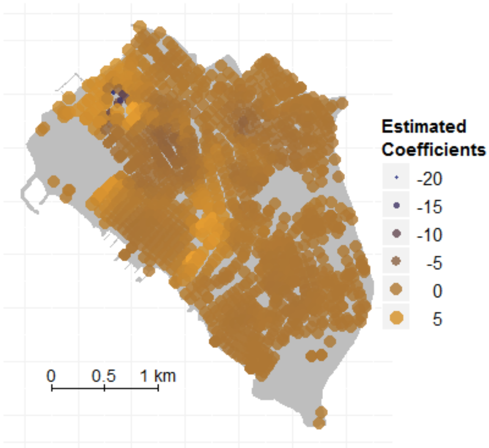
Figure 3.32: Coefficient estimates of GWR for Kalamaria 1/2



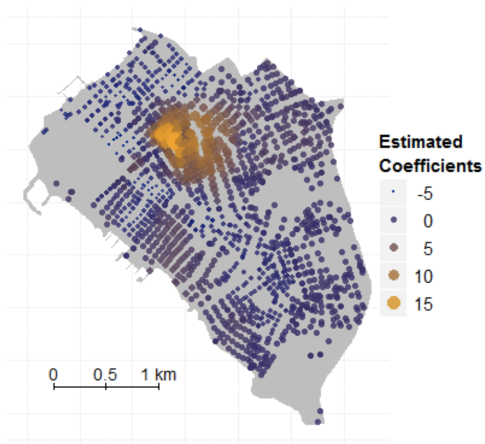
(i) logarithmic distance from stadium



(ii) logarithmic distance from pedestrian shopping zone

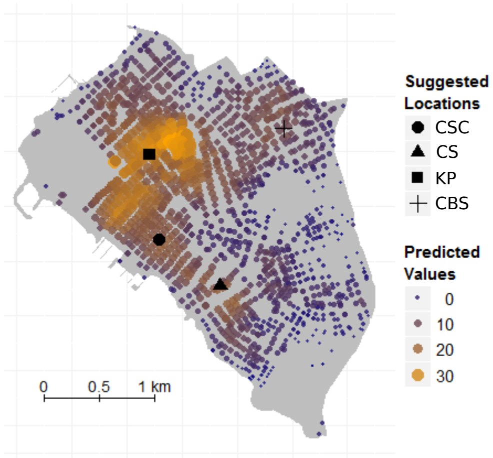


(iii) average speed

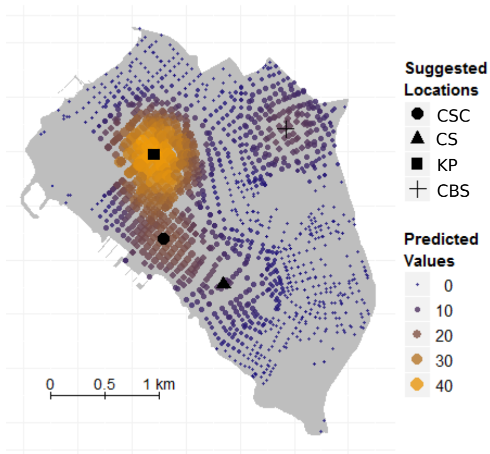


(iv) length of road network

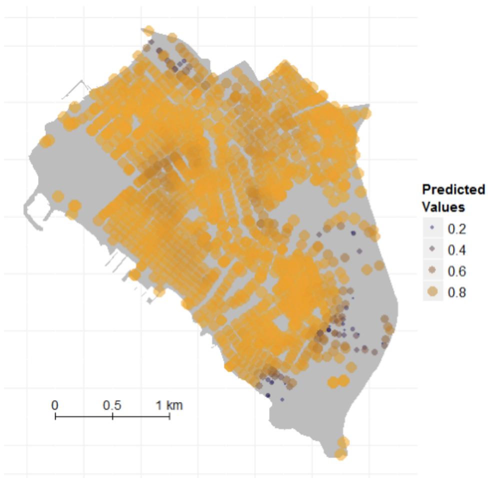
Figure 3.33: Coefficient estimates of GWR for Kalamaria 2/2



(i) OLS



(ii) GWR



(iii) local R-squared for GWR

Figure 3.34: Facility allocation model predictions (where KP = Komninon pedestrian zone, CSC = Citizen Service Center, CBS = central bus station, and CS = central square)

Findings

The problem of facility location has been addressed occasionally by different methodologies, from multi-criteria analysis to location-allocation algorithms, but not explicitly by spatial modelling point of view until now. The use of spatial models meets considerable growth because of their advantages over global models in terms of solving the autocorrelation problem, but also for their integration with GIS, which offers tools for the visualisation of their results. The geographically weighted regression and simultaneous autoregressive models have been applied for the prediction of house prices, development of regions, traffic analysis and other purposes.

In this chapter, OLS, SARmix and GWR have been applied to model the perception of people about the optimal location of mobility centers. Among the considered variables, the selection of the appropriate location of a mobility center seems to be positively influenced by the residential population around the location, the number of nearby points of interest, the number of bus routes, the vehicles' average speed and the length of the road network. Other factors, such as distance from commercial center, income and parking stations, have negative influence on the selection of the mobility center location.

Concerning the three models used, SARmix seems to better fit the data, while the Moran's I test shows that only the GWR solves the autocorrelation of the residuals, something the other models (OLS and SARmix) fail to accomplish. The spatial model's predicted values are localized (unlike OLS), making the decision easier. Unlike MCA and linear regression, the local-model methodology results in different coefficient estimates for each variable and each point in the modelled area, using for each point a subset of the neighboring data to obtain a local estimate. This, of course, makes interpretation of the local values more difficult Additional information (such as the results of a relevant survey) could be used as additional input for the more effective selection of the optimal site location.

The proposed method is not system dependent and can thus be used for a number of other applications, including the positioning of car and bike sharing stations or electric chargers for electric vehicles (EV's), using the appropriate system-related variables.

The selection of the appropriate location for a transportation facility is a crucial matter when a local authority or governmental agency plans transportation investments that attract massive population. Residents and visitors must be able to find easily this facility in order to use the services offered. Moreover the facility should be located in such a place so that it meets the various characteristics and particularities of the greater area, such as population, bus routes, road network and others. Therefore, this is a complex problem that requires scientific approaches and tools in order to uncover and take into considerations factors that play a decisive role in the selection of the appropriate location.

3.8 Discussion

Spatial econometric models have been developed to measure the impact of transportation infrastructure locations on real estate prices and land-use change. Purchase and rent price models of dwellings for the two major Greek cities Athens and Thessaloniki have been presented, using real data collected form on-line sources. The impact of transportation infrastructure locations differs depending on the type of the facility and the time of construction. Some of the results come in accordance with similar studies that have been performed in European, American and Asian Metro areas, while others differ. Moreover, the price–to–rent model is introduced. The increased availability of on-line data facilitates the development of such models, and its use in LUTI models should be further explored.

The effect of the current financial crisis on real estate purchase prices and rents has been investigated, focusing on the ability of transportation infrastructure to maintain the prices higher, or not. While this research forms a first substantial attempt to quantify the crisis effects at –disaggregate– agent-based level, LUTI models should be used to further investigate the impact in more fields (society, environment etc).

Finally, two methodologies for optimal location identification of demand-oriented transport facilities such as carsharing, bikesharing stations, and electric-car chargers, based on MCA and spatial econometrics are presented. The location of transportation infrastructure has been proved be crucial for determining the urban development and should be taken into consideration for planing purposes in LUTI models.

In conclusion, the results of this chapter indicate that spatial econometric models perform better than traditional linear regression in solving spatial autocorrelation, and therefore their integration in the next generation LUTI models is highly recommended. One of the main factors considered to make their implementation challenging, is the uncertainty in the number of nearest neighbor points that should be used for the construction of the weight matrixes, and whether its structure should change at every simulation step. Moreover, due to their computational complexity, the time required for their estimation was regarded as an obstacle. In order to address these concerns, it is first suggested the implementation of a generally simple form of spatial model, the Spatial Error (SEM), which applies the autoregressive term to the error, and the use of a single, static weight matrix, same at every simulation step. The forecasting capabilities of these models should be further investigated.

4 Policy evaluation in LUTI

4.1 Research objective and chapter structure

The objective of this Chapter is to present a number of methodologies that have been developed for the purpose of this doctoral research, expanding in different steps of LUTI model implementation, from data collection to policy analysis. Section 4.2 presents a methodology for associations generation in synthetic population, to support the synthesis of disaggregate demographic data from aggregate measurements. In Section 4.3 is described the development of the first zone-level UrbanSim & MATSim LUTI model presenting the Brussels case study. Section 4.4 demonstrates how LUTI models can be used for qualitative policy assessment using urban quality indicators, and Section 4.5 suggests a multidimentional indicator analysis methodology for quantitative policy evaluation, based on distributions rather than single aggregate measurements.

4.2 A graph-theoretic solution for associations generation in synthetic population

4.2.1 Background

Individual demographics, as well as home and work locations, are commonly collected in a census, but this data is typically not available for microsimulation due to privacy concerns. This has led to work on population synthesis, which aims to generate virtual people with the same demographics as the real population. Generally two data sets are needed for synthesis: a microsample (which typically does not contain any location information) and aggregate measures for the smallest available unit (the name of which varies by country) (Williamson et al., 1998). The idea is that this synthetic population would behave similarly to the real population, allowing for its use in microsimulation and other applications without any of the data confidentiality concerns that arise when real census data is used. Synthetic population generation has applications across many fields of research (Fotheringham and Rogerson, 2009; Zaidi et al., 2009; Birkin and Clarke, 2011).

Individual attributes are important factors in travel behavior (which includes route choice, mode choice and other decisions), but do not tell the whole story. Households, and particularly an agent's position within the household, play an important role in transportation decisions as well (Jones et al., 1983; Ballas and Clarke, 1999; Fisher et al., 2007; Ben-Akiva et al., 2012). Two-worker households are a particularly significant group and make collective decisions about household location which affect travel behavior (Surprenant-Legault et al., 2013). Other research papers have considered the influence of social interactions on travel behavior (eg. Dugundji and Walker, 2005; Arentze et al., 2012; Ben-Akiva et al., 2012).

The approach described in this research to Gargiulo et al. (2010) and Barthelemy and Toint (2013) in the sense that the association between synthetic persons and household positions is generated after the generation of individuals and households. That is to say that all three approaches are long-list based rather than fitted table based. However, it should be noted that only in this work are the agent-level marginals/conditionals before the matching process preserved in the associated population. Both Gargiulo et al. (2010) and Barthelemy and Toint (2013) use Monte Carlo simulation to generate the association, which intrinsically means that the marginals at the end of the process will not be an exact match to the individual and household marginals they started with. To avoid unmatched agents, they require a large quantity of oversampling. Also, they are based only on one realization of the simulation. In the case of Gargiulo et al. (2010), only age is used to sample the individuals and fit them to households. Their algorithm iteratively draws household size, age of head, household type, spouse age, and children's ages in order from probability distributions conditional to the last attribute chosen. Size and type are the only household attributes considered. Barthelemy and Toint (2013) also follows a more or less sample procedure where the individual type is used for the match. The procedure presented in this research evaluates all attributes simultaneously and is a deterministic solution so the results do not vary from one run to another. More attributes are used, three for individuals and six for households, compared to three total in Gargiulo et al. (2010), which increases the complexity.

A previous work by Farooq et al. (2013b), looks at inter-agent interactions, modelling the marriage and rental housing markets as bipartite graph matching problems computed by the Hungarian algorithm (Kuhn, 1955). The association generation problem in this research is formulated in the same way and also uses the Hungarian algorithm. Persons are matched to household positions in order to have a 1-to-1 matching, required for a valid bipartite graph. Households and individuals are considered to be two independently generated synthetic populations, so there are no prior associations between them. This is another difference compared to Gargiulo et al. (2010), who construct households around an identified household head instead of defining all the open positions in the initialization phase.

4.2.2 Methodology

An example of a bipartite graph is shown in Figure 4.1. Set U represents the list of persons while set V represents the list of households positions, since the matching for each position is run separately. These two lists are connected by edges E, which all have a length L, associated with them; the length is a transformation of the utility function where smaller values represent better matches. In this application, every element of list U has an edge connecting it to every element of list V. Since every connection is possible, the edge lengths can be stored in a cost matrix. The solution for a bipartite graph is a set of edges such that no two edges share an endpoint. In some algorithms, it is necessary for the two lists to have the same length. Here is used one, where unequal length is permitted; in this case, all elements of the shorter list are connected by an edge while some elements of the longer list are left unconnected.

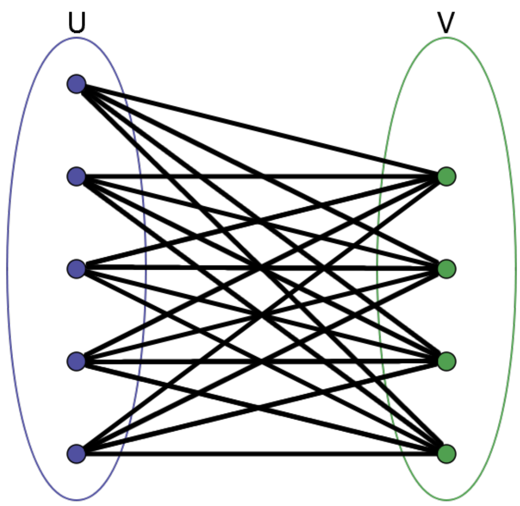


Figure 4.1: Sample bipartite graph

The formulation of household-population matching as a bipartite graph problem is similar to the procedure used by Farooq et al. (2013b). The goal is to determine the one–to–one matching between people and household positions. Mathematically, the problem is to find a graph $G^* = (U, V, E^*, L^*)$ such that the cardinality of E^* is equal to the cardinality of V (but not U, as there is a surplus of people). The sum of lengths L^* is guaranteed to be the minimum possible. This is equivalent to a maximum weighted bipartite matching, which can be solved by linear programming algorithms (Burkard et al., 2009).

The Hungarian algorithm (Kuhn, 1955) is perhaps the most common for bipartite graph matching. However, the Hopcroft-Karp algorithm (Hopcroft and Karp, 1973) has better performance in terms of time complexity ($O(\sqrt{n})$ vs. $O(n^4)$). The Hungarian algorithm is currently being used for matching due to the relative ease of implementation. It is an optimisation algorithm that solves the assignment problem in polynomial time. The implementation used is a modified version of the original algorithm (Munkres, 1957), which is important because the original requires graphs of equal length, whereas we have a population graph which is larger than the household graph. The modification also reduces the time complexity to $O(n^3)$. A switch to the Hopcroft-Karp algorithm is planned as an extension to this work but has not been completed at this time. The two algorithms use the same problem formulation, so in any case the matching results should be identical. The only advantage is in terms of computation time, which would be significant for a large enough population. For the head matching experiments done in this study (3,100 households), the total computation time is around 5-10 minutes. Computation time was shorter (1-2 minutes) for spouse matching as both graphs were smaller (people matched to head positions were removed from the synthetic population and only 1,200 households had a spouse position).

The population synthesis procedure is the same as Farooq et al. (2013a), a Markov chain Monte Carlo simulation approach which uses a Gibbs sampler to draw from agent attribute distributions. This study considers the problem of matching synthetic people into household head and spouse positions. These are the most important positions, particularly for household-level decision making. Matching people into other household positions is outside the scope of this work, but is planned to be considered in future research. A synthetic population of people and a set of real households from the census data is first used. A sample of 3,000 real households (with the real heads and spouses) is taken from the Lausanne area to mimic the actual data. Not all households have a spouse position so that sample was smaller, around 1,200. This was one of the advantages of having access to full census data, as these records were drawn from the complete population and did not require reweighting. The use of this set for model calibration is equivalent to the use of small-area samples in other studies. Multinomial logit models are then developed, with dependent variables the household size or type, and explanatory variables person and household attributes. The models are fitted to the sample using BIOGEME (Bierlaire, 2003; Bierlaire and Fetiarison, 2009). The result is a utility function for possible matches, which is transformed into a shape length attribute for the bipartite match. The bipartite matching algorithm used was written to minimize the distance between potential matches. This, of course, does not work with a utility function since larger values represent a better match. Various transformations are possible to invert the utility function. In this instance, a linear transformation was used to ensure that all functional values are positive and that the best matches are represented by the lowest values. Three different utility functions, described in more detail in the Section 4.2.4, are used. The quality of the matches is evaluated by comparing the total distribution of in the matched and real populations as well as by comparing the count of each unique combination of attributes in the two populations. A flowchart of the method used is shown in Figure 4.2.

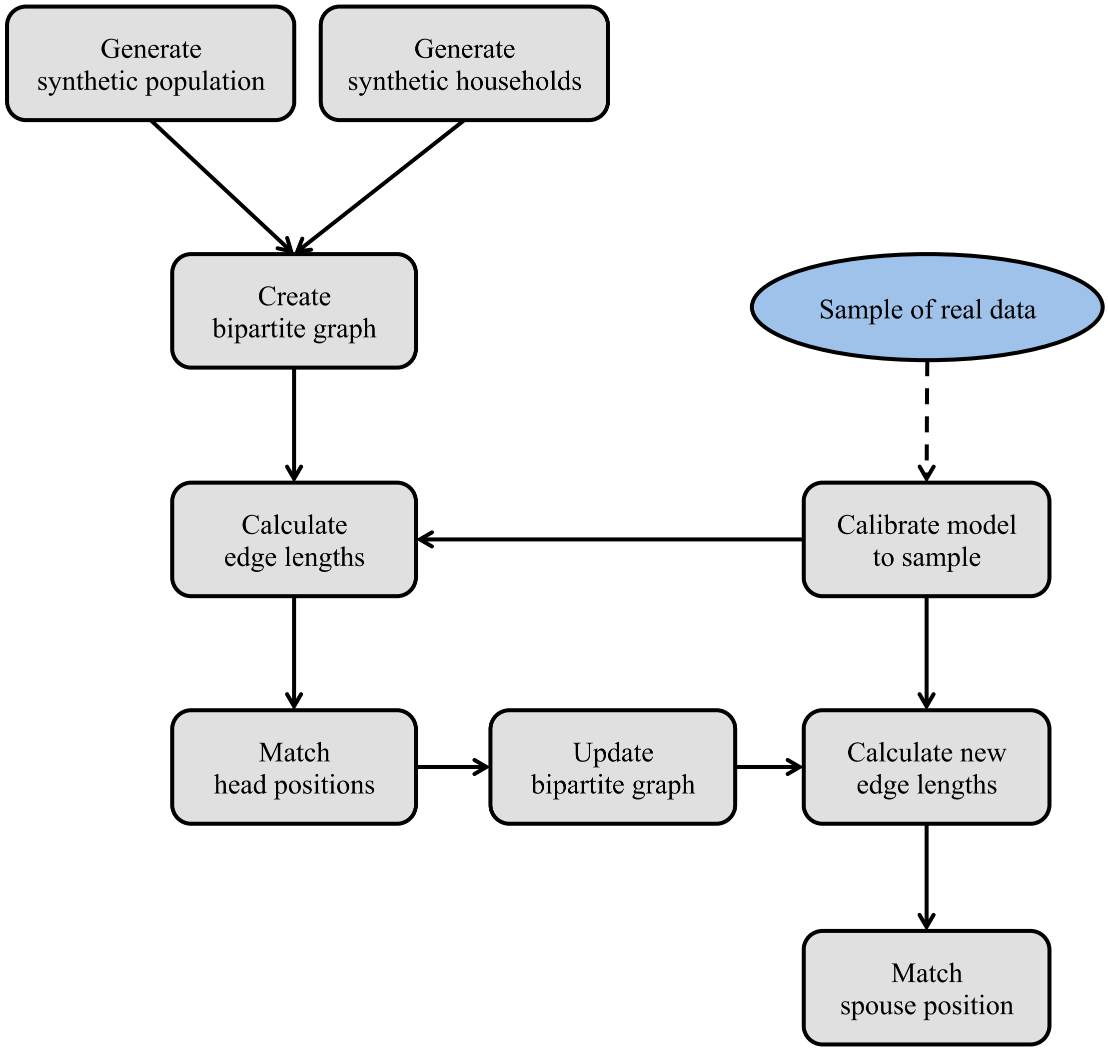


Figure 4.2: Flowchart of associations generation method

4.2.3 Data description

The main data source for this project is the 2000 Swiss Census, which has been made available for research purposes. The person population used in the matching experiments was generated from marginals calculated from the census data using the procedures described in Farooq et al. (2013a). The household list used is all households in postal code 1004, which is located in Lausanne. This zone has 3,100 households, and 1,205 of these households have a spouse position. The census data is at the individual level, which means that household level variables had to be derived based on the unique household ids. Since the synthetic population was very large (over 1,000,000 records), a sample of 6,458 records (equivalent to the population of zone 1004) was taken to shorten the runtime. After the head matching is complete, individuals which have been matched into head positions are removed from the synthetic population for the spouse matching.

Households have seven variables: size, number of workers, dwelling type, number of cars, type, nationality, and language spoken. All of these were calculated from the census data by performing a dissolve on the household ID field and taking a sum/minimum of the other fields. Size has categories for 1, 2, 3, 4, 5, and 6+ persons. Workers is an integer field with a range of 0 to 6. Dwelling contains a unique building ID. This field is unused at the moment but will be incorporated in future work when a list of building IDs and types can be obtained. Cars is an integer field with a range of 0 to 4. This was actually calculated by looking at the number of household members who drive a car to work, since there was no field for car ownership. We believe this to be a reasonable substitute because it is safe to assume that every person who drives a car to work is a car owner. However, there may be cars unaccounted for with this method because they are not used for work trips. Household type has three categories: one person, family, and non-family. All married couples are contained in the family category, regardless of children. Nationality is binary variable with choices Swiss or other. Language variables exist for all of the official languages of Switzerland. Since the study area is in the French-speaking area, we use the French language variable which is also binary (French spoken at home, or any other language spoken at home).

Persons have four variables: age, gender, household size, and education. Age has 8 categories: <15, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+. Sex is either male or female. Size has the same categories as on the household side: 1, 2, 3, 4, 5, 6+. Education has four categories: none, primary, secondary, and tertiary.

One limitation of the Swiss Census is that there is no field which directly measures income, however number of workers, number of cars, and building type (presently unused) but especially education have some correlation with income.

4.2.4 Model development

Head Matching

For the head matching, two different models are used: one which constructs the operation as a choice between different household sizes, and a second which chooses between different household types.

Each value in the household size category (ranging from 1 to 6+ persons) is considered as an independent choice. One person households are the base case and have a utility (after transformation for the bipartite match) of:

$$C_{ij} = -ASC_j + L \tag{4.1}$$

where i represents people and j households. L is a constant parameter to ensure that edge length remains positive. For all other household sizes:

$$\begin{aligned} C_{ij} = -(ASC_j + AgeConst[Age_i] - 0.328 * Sex_i + EducConst[Educ_i] - 0.419 * Nationality_j \\ - 0.393 * Language_j + WorkersConst[Workers_j] + CarsConst[Cars_j]) + L \end{aligned} \tag{4.2}$$

where $AgeConst$, $EducConst$, $WorkersConst$, and $CarsConst$ are arrays which contain constants for different values of the respective attribute. The attribute value is used as the index. After trying various values, the parameter L was set to 20.

Type Choice

Each value in the household type category (one person, family, and non-family) is used as an independent choice. One person households are the base case and have a utility (after transformation for the bipartite match) of:

$$C_{ij} = -ASC_j + L \tag{4.3}$$

For the two other types:

$$\begin{aligned} C_{ij} = -(ASC_j + AgeConst[Age_i] - 0.390 * Sex_i + EducConst[Educ_i] - 0.327 * Nationality_j \\ - 0.376 * Language_j + WorkersConst[Workers_j] + CarsConst[Cars_j]) + L \end{aligned} \tag{4.4}$$

This is very similar in construction to the Size Choice model but the use of a different attribute changes some of the constants and leads to a different fit as can be seen in the results section. The parameter L is also set to 20 for this model.

Spouse Matching

Since all households with a spouse position are assigned the family type by the census, the type choice model would be meaningless for the spouse choice problem. Therefore, a variant of the size choice model is used for all spouse matching. It is run twice because the available population depends on the head matching results.

The model is similar to the head size choice model. For two person households:

$$C_{ij} = -ASC_j + L \tag{4.5}$$

where i represents people and j households. For all other household sizes:

$$\begin{aligned} C_{ij} = -(&ASC_j + AgeConst[Age_i] + 0.534 * Sex_i + EducConst[Educ_i] - 0.410 * \\ &Nationality_j - 0.146 * Language_j + WorkersConst[Workers_j] + \\ &CarsConst[Cars_j]) + L \end{aligned} \tag{4.6}$$

The parameter L is also set to 20 for this model.

Implementation

The implementation used is a modified version (Munkres, 1957) of the original Hungarian Algorithm written in Java. As mentioned earlier, there is assumed to be an edge connecting every element of U, the person set, to every element of V, the household set. This allows edge lengths, E, to be stored in a cost matrix. The code iterates through persons and households, filling in the cost matrix according to the formulas given above. This matrix is passed to the Hungarian algorithm, and a list of edges is returned as the solution.

4.2.5 Results

The specification of variables is contained in Table 4.1 and the estimated values along with the t-test values in the three models are shown in Table 4.2. Nearly all variables are significant at the 95% level. The four exceptions were kept in the models, as suggested by Ziliak et al. (2008), because of the desire to have a constant for each attribute value and to have the same attributes involved in all models. The ρ values suggest that the Type Choice model is the best fit, followed by Size Choice and Spouse (although the lower value for Spouse may be a result of the smaller number of observations).

We note that, while the age variables are highly significant, the variable $Education_1$ in Type Choice, and the variables $Workers_2$ and Car_2 in both Size and Type Choice have similar or higher t-test values. This suggests that the approach of Gargiulo et al. (2010), which uses only age, household size, and household type, omits several highly significant variables that could be incorporated to improve the match quality. Their choice of attributes may be explained by the model's application, demographic evolution, since most other attributes (education, marital status, employment, nationality, language spoken at home) are variable on a long enough time scale. But even for this application, the existence of more person and household attributes in the present state can only help to predict transitions into future states (and if not, the extra variable can be ignored).

Many different approaches are possible for measuring match quality. In this research, it has been chosen to look at two different measures: the population-level distributions of attributes, and the fit in terms of unique combinations of attributes. Attribute distributions over the whole population are used, because the area from which the households are taken is actually somewhat small. In an aggregate model, this might be one zone. And in a broader sense, the goal is to produce a synthetic population which will behave similarly. If the attribute distributions match, then the real and synthetic populations have similar characteristics and diversity and should have similar behavior as well. The other measure, which looks at attribute combinations, can help to show if there is systematic bias in the model. A good model would have a trendline slope close to 1 and a high R^2 value. If this is skewed in either direction, there is an overrepresentation or underrepresentation of some types of people. In that case, the behavior of the synthetic population may not match the real even if the attribute distributions are a close match.

Figure 4.3 shows a comparison of the attribute distributions for real household heads to the matched populations from Size Choice and Type Choice models. In Figure 4.3i, it is observed that both models produce distributions which are close to that of the real population for all age groups. Size Choice produces a slightly older population, overestimating ages 45-74 while underestimating ages 15-44. Type Choice significantly overestimates ages 65-74 but has no overall trend and is relatively close on all other groups. Looking at Figure 4.3iii, it is observed that both models have the correct trend (more men than women in head positions). Size Choice slightly overestimates men while Type Choice is almost exactly equal to the real population. For education (Figure 4.3ii), Size Choice overestimates secondary education and underestimates primary while Type Choice does the inverse.

Figure 4.4 looks at the same attribute distributions for the spouse results. Although the model used is the same for both Size and Type Choice, results differ because the remaining population is dependent on the results of the head matching. In the age results (Figure 4.4i), it is observed a shift towards the younger age groups with both matched populations underestimating ages 55-74 and overestimating 25-34. The sex distribution in Figure 4.4iii shows that both matched populations are close to the true distribution with Type Choice performing slightly better. The spouse education distribution (Figure 4.4ii) is not a particularly good fit, with both populations under– or over–estimating all categories to some degree.

Figure 4.5i shows the fitting results for the Size Choice model. The head results, show that most points are relatively well spread out on either side of the $y = x$ line. The absence of points on the x-axis shows that all unique combinations of attributes which exist in the real population are represented to some degree in the fitted population. The shape of the trendline shows a slight trend toward underrepresentation among unique attribute combinations. The spouse results, showed in 4.6i are also good. The trendline slope and the R^2 are better than the head position. However, there are some points on the x-axis which indicates combinations unrepresented in the matched population.

Figure 4.5ii shows the fitting results for the Type Choice model. Comparing the head results presented in Figure 4.5ii to the Size Choice presented in Figure 4.5i, it is observed that the trendline is more skewed away from the $y = x$ line. This indicates that there are more undersampled groups in the Type Choice model. Results for the spouse model (Figure 4.6ii) are similar to those for Size Choice, which is expected because the model is the same.

Overall, the four fitting plots show that the Size Choice model performs better than the Type Choice model in this application. Although the difference is not so great as to preclude the use of the Type Choice model; in fact, it might perform better than Size Choice in other cases.

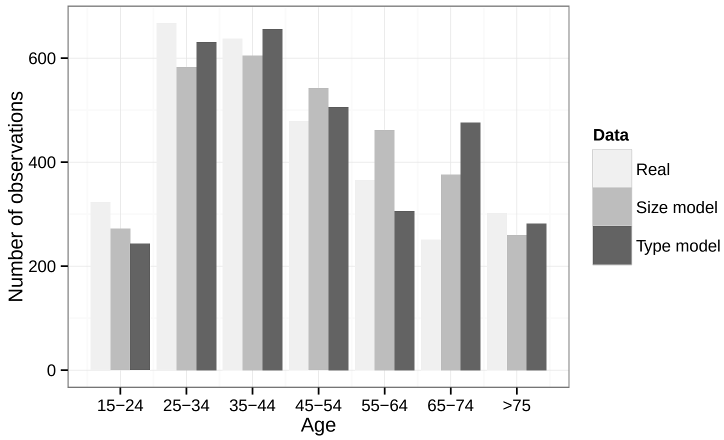
Table 4.1: Specification of variables

Variable	Specification	Name
Alternative Specific Constant		ASC_1
Alternative Specific Constant		ASC_2
Alternative Specific Constant		ASC_3
Alternative Specific Constant		ASC_4
Alternative Specific Constant		ASC_5
Alternative Specific Constant		ASC_6
Alternative Specific Constant		ASC_{one}
Alternative Specific Constant		ASC_{notfam}
Age 15-24	dummy, 1 if TRUE	Age_1
Age 25-44	dummy, 1 if TRUE	Age_{23}
Age 45-54	dummy, 1 if TRUE	Age_4
Age 55-64	dummy, 1 if TRUE	Age_5
Age 65-74	dummy, 1 if TRUE	Age_6
Age 55-74	dummy, 1 if TRUE	Age_{56}
Age >75	dummy, 1 if TRUE	Age_7
Age >65	dummy, 1 if TRUE	Age_{67}
Education 0	dummy, 1 if TRUE	Education_0
Education 1	dummy, 1 if TRUE	Education_1
Education 1 and 2	dummy, 1 if TRUE	Education_{12}
Education 3	dummy, 1 if TRUE	Education_3
Female	dummy, 1 if TRUE	Female
Workers 1	dummy, 1 if TRUE	Workers_1
Workers 2	dummy, 1 if TRUE	Workers_2
French speaking	dummy, 1 if TRUE	Fr
Swiss Nationality	dummy, 1 if TRUE	Nat
Car 2	dummy, 1 if TRUE	Car_2

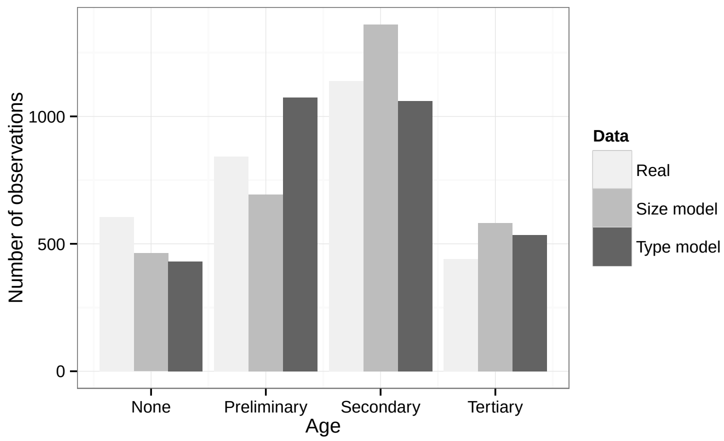
Table 4.2: Model estimations

	Head				Spouse	
	Size n=3026		Type n=3026		Size and Type n=1202	
Variable	**Value**	**t-test**	**Value**	**t-test**	**Value**	**t-test**
ASC_1	1.91	9.90	-	-		
ASC_2	-	-	-	-	1.12	3.16
ASC_3	-0.737	-11.77	-	-	-	-
ASC_4	-0.993	-14.49	-	-	0.027	0.31*
ASC_5	-1.86	-19.16	-	-	-0.821	-7.30
ASC_6	-2.63	-19.15	-	-	-1.63	-10.61
ASC_{one}	-	-	1.33	7.18	-	-
ASC_{notfam}	-	-	-2.62	-26.66	-	-
Age_1	-	-	-	-	-1.35	-4.56
Age_{23}	1.59	8.89	1.49	8.24	0.681	3.32
Age_4	1.57	7.78	1.44	7.08	-	-
Age_5	-	-	1.17	5.70	-1.73	-6.03
Age_6	-	-	1.49	6.89	-	-
Age_{56}	1.47	7.94	-	-	-	-
Age_7	1.16	5.39	0.987	4.44	-	-
Age_{67}	-	-	-	-	-2.57	-6.46
$Education_0$	-	-	0.142	1.05*	-0.502	2.16
$Education_1$	-	-	0.590	5.00	0.380	2.02
$Education_{12}$	0.150	1.28*	-	-	-	-
$Education_3$	-0.657	-3.91	-0.577	-3.89	-0.648	-2.45
Female	-0.328	-3.52	-0.390	-4.13	0.534	2.77
$Workers_1$	0.393	3.63	0.473	4.17	0.731	2.83
$Workers_2$	5.40	10.53	5.49	10.67	0.563	2.15
Fr	-0.393	-3.69	-0.376	-3.48	-0.146	-0.82*
Nat	-0.419	-4.19	-0.327	-3.18	-0.410	-2.32
Car_2	6.96	13.26	7.77	14.78	-	-
	ρ =0.328		ρ =0.431		ρ =0.268	

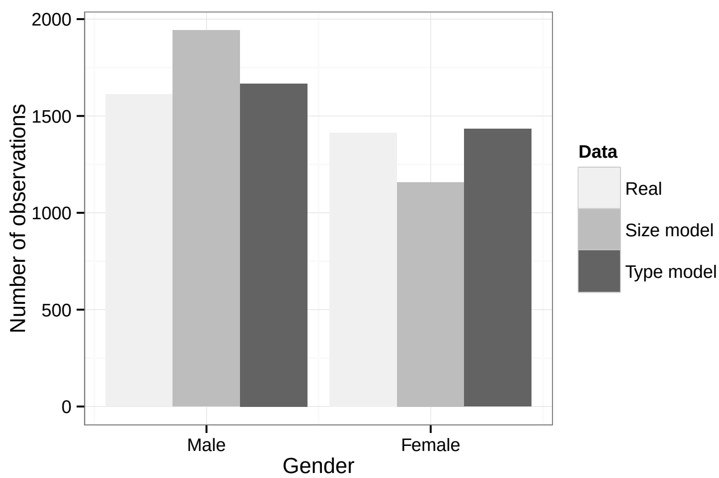
*Not significant at 0.05 level



(i) Age

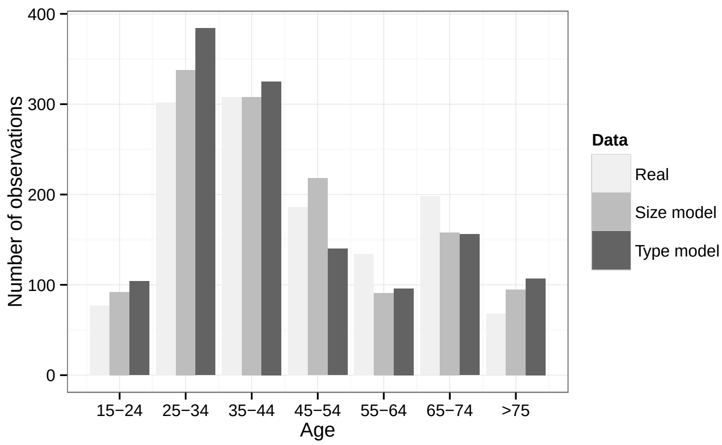


(ii) Education

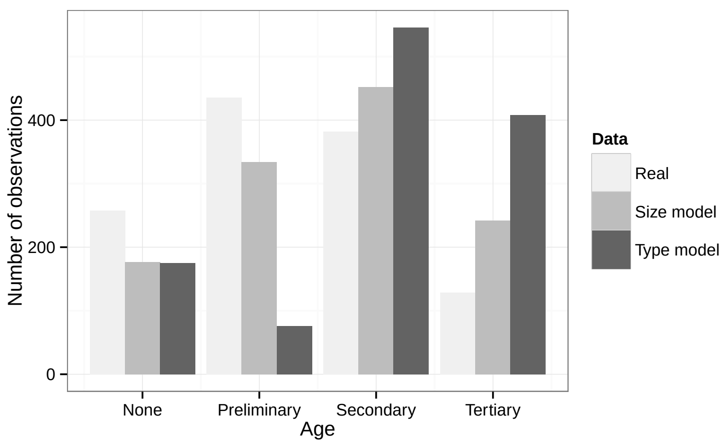


(iii) Gender

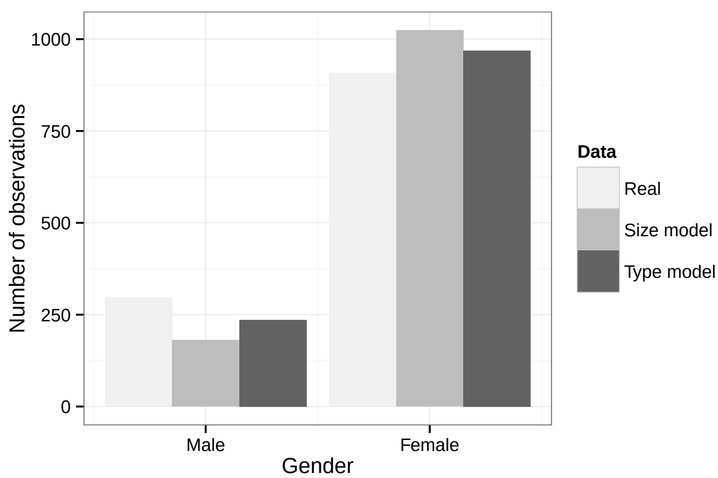
Figure 4.3: Household head distribution comparison



(i) Age

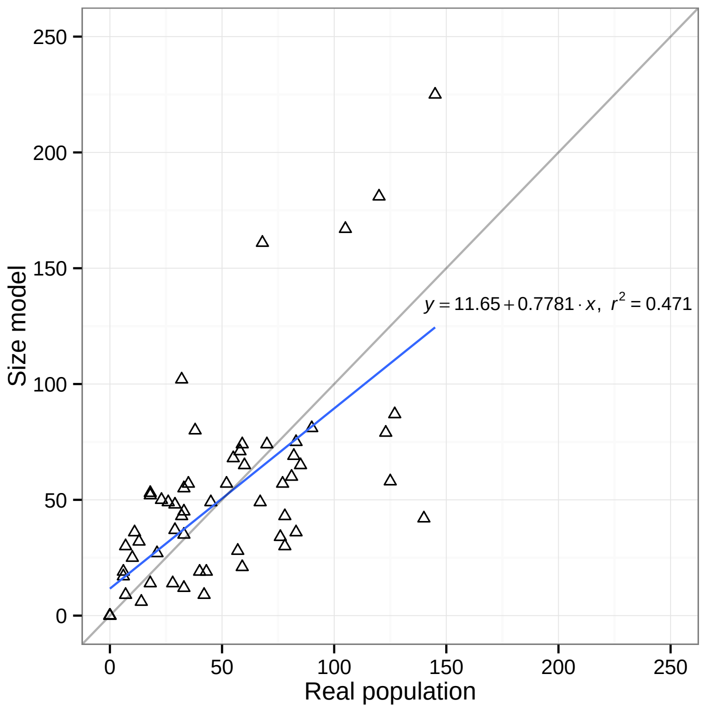


(ii) Education

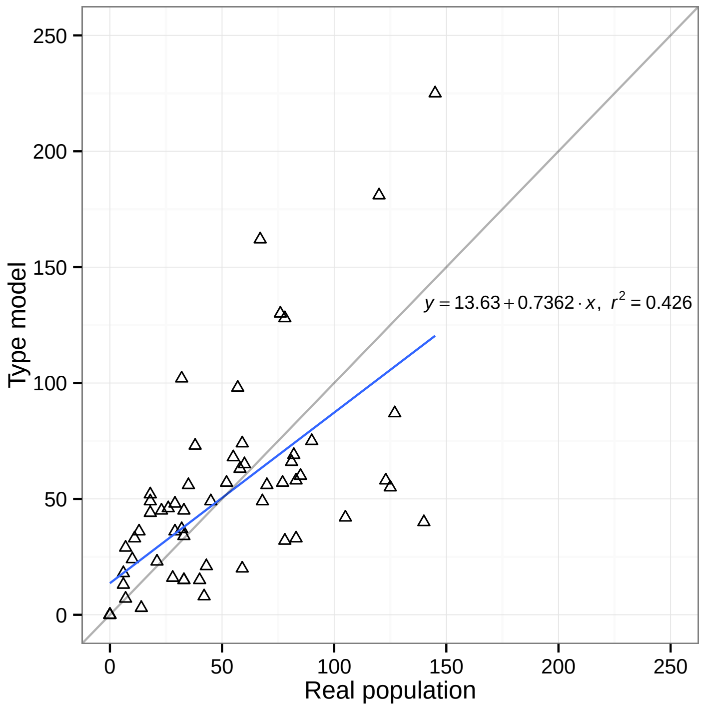


(iii) Gender

Figure 4.4: Household spouse distribution comparison

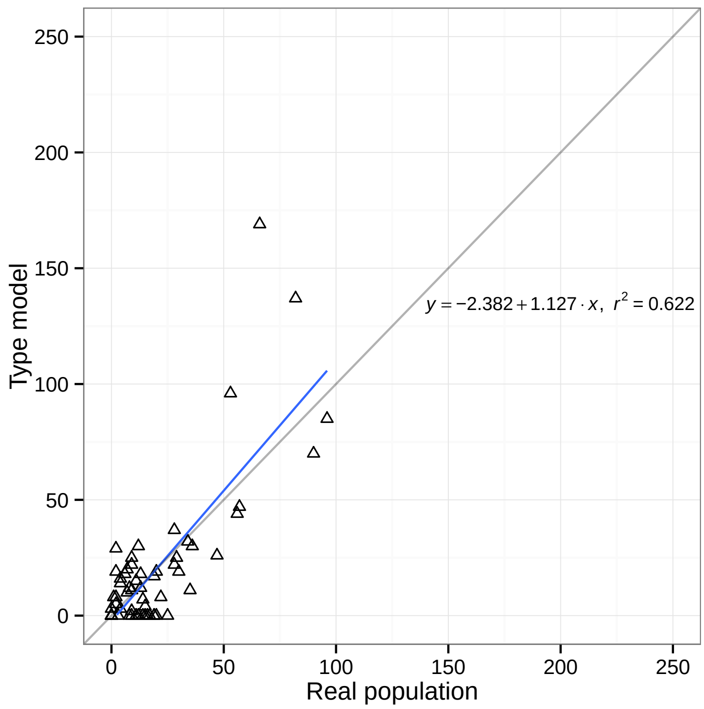


(i) Head - Size Choice

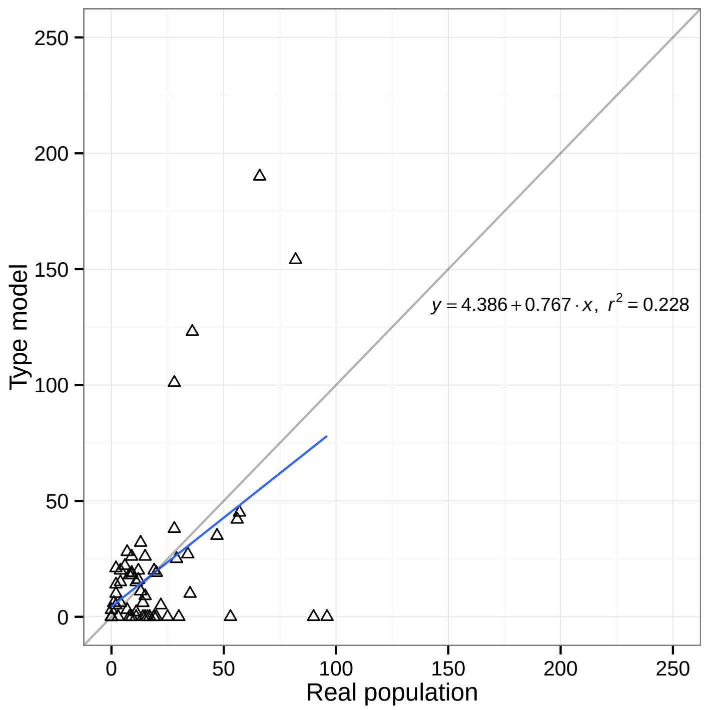


(ii) Head - Type Choice

Figure 4.5: Household distribution comparison - head position



(i) Spouse - Size Choice



(ii) Spouse - Type Choice

Figure 4.6: Household distribution comparison - spouse position

4.2.6 Conclusions

In this research, bipartite graph-based method for associations generation in synthetic population is proposed. It is then applied to matching of head and spouse positions for the case study of Switzerland. The presented methodology could be proven valuable for the development of Integrated Land-Use and Transport Models. Two approaches of associations generation were tested: one which formulated the problem as matching people to a particular household size, and a second which looked at matching people to household type. Both of these approaches were used to match people into household head positions, whereas only the size model could be used for spouse positions because the type was not meaningful for this problem. All three models (as size choice was calibrated separately for head and spouse positions) are multinomial logit models calibrated to a sample of the real population using BIOGEME. The resulting utility functions were then transformed to a length attribute and the head matching was set up as a bipartite graph. A modified Hungarian Algorithm was used to solve the graph, and the remaining unmatched population was used to run a second bipartite match for the spouse position.

Overall, the attribute distributions are a relatively good fit for the head position. Some attributes are overrepresented in the fitted population and others are underrepresented, but all attributes are present and the general shape of the age, sex, and education distributions is correct. For the spouse position, the distributions are similarly good with the exception of education. The four education categories are over— or under–sampled to some degree. This could be caused by propagation of error. The population available for the spouse matching depends on which individuals are used in the head matching. Consequently, if a particular attribute is oversampled in the head matching, less individuals will be available for the spouse matching. It is observed that primary education is oversampled in the head results for Type Choice. In the spouse results, this is reversed, which suggests that there may be a correlation.

The fitting results show that all combinations of attributes in the real population are also present in the matched populations for the head position. There is room for improvement, but with results like this we would expect the real and matched populations to behave similarly in simulation. The spouse models result in good fit, but there are some attribute combinations which are absent in the matched populations. This is an area for improvement in future models.

These results are encouraging, and show that the idea of first generating a population of synthetic individuals and then matching them into household positions is feasible. Although disaggregate data was used in this study, the methodology is applicable without it. Model calibration could be done with a small area sample, and results could be evaluated by comparison to aggregate distributions, as has been done in Figures 4.3 and 4.4. It has been demonstrated the matching procedure for the head and spouse positions. There are many possible extensions of this topic. The next logical step would be to extend this procedure to all household positions. Once that is done, a complete synthetic population of people and households will be available, and could begin to be used it in microsimulation models and evaluate its performance by comparing to results obtained using other data sources. This method is capable of generating the large amount of individual and household-level data required for detailed transportation models while balancing the privacy concerns inherent with personal data.

4.3 Developing an Integrated Land-Use and Transport model

4.3.1 Data collection

One of the main characteristics of UrbanSim is the requirement of disaggregate data. The main datasets needed are the agents that generate the demand (jobs and households) and the available supply (buildings), while historical data of the developed projects (development events history) and the shapefiles that define spatial levels are integral parts needed for a basic UrbanSim run.

The main data sources used for this application were the 2001 Belgium Population Census and the Belgium Land Registry (a cadastre of real estate goods). Both datasets were obtained at an aggregate level from the Belgian Statistical Authority (SPF - Economie, http://www.economie.fgov.be). Aggregate data regarding employment by activity type and commune was collected from the ONSS (http://www.onssrszlss.fgov.be) and INASTI (http://www.rsvz.be) databases, from the same source. Additionally, individual level data for households and persons was obtained from the travel survey MOBEL (Hubert and Toint, 2002), performed in the area of study during 2002.

The data was available with different levels of spatial disaggregation. These levels are the ”communes”, that divide the Brussels Metropolitan Area in 151 units, and the ”zones” that divide it in 4945 units. Each zone belongs to only one commune.

For zone-level projects, as the one described here, UrbanSim makes the following assumption: Each zone contains one representative building per type, while the real number of units (buildings) of each type in the zone, is included in a field named ”number of units” in the buildings table. In our building data for Brussels, there are 14 categories of building types: four residential (detached, semi-detached, attached and apartments) and 10 non-residential (industrial, governmental, educational, quarrying, warehouses, office, shops, hotels/bar/restaurants, industrial). The main attributes of the buildings are: price (average value per unit), type, structural characteristics (average surface by unit type, surface by non-residential use) and capacity.

The households table contains information about its size, the number of workers, cars, income, location of residence, etc. Finding a complete dataset of households is usually im-possible. For the purpose of this research, a synthetic population was generated (Farooq et al., 2013a).

Since the employment data were available at the commune level, they were distributed to zones through Monte Carlo simulation, following the observed distribution of available non-residential surface. The number of job-sectors was assumed to be equal with the number of types of non-residential buildings, meaning that each type of job could be located in a specific type of building (e.g. jobs in industrial sector can be located in industrial buildings).

The integrated version of UrbanSim with MATSim requires a road network, which in our case was acquired from the Open Street Maps (openstreetmaps.org), and a list of commuters (workers) with information about their household (origin) and job (destination). This dataset was structured using the synthetic population (households) and the observed OD for Brussels.

Finally, UrbanSim requires a dataset with the residential and non-residential contraction projects (from 1990 to 2001), to be used for the estimation of the building development model, and other smaller, less disaggregate data, such as the annual employment control totals, annual household control totals and target vacancies.

4.3.2 Model Estimation

All the models were estimated using the OPUS/UrbanSim software platform. Figure 4.7 depicts the interaction between the land-use and transport models and sub-models used in this case study. One of the main advantages of using UrbanSim for model estimation, is the possibility of using OPUS for creation of interaction variables with data coming from different tables. Several different specifications were examined, given the available data and following what the literature and urban economic theory suggests as explanatory variables for each of the modeled phenomena (Picard and Antoniou, 2011). Final specifications were selected following estimate-significance and theoretical-consistency criteria. For all choice models a linear-in-parameters utility function specification like the following was chosen:

$$V_i = \sum_k \beta_k x_{in}^k \tag{4.7}$$

where β_k is the k-th parameter to estimate and $x_{in} = x_i \cdot x_n$, is a variable describing an interaction between a k-th attribute of the alternative i and a characteristic of the decision maker n.

MATSim for UrbanSim

MATSim for UrbanSim (Nicolai, 2012; Nicolai and Nagel, 2011) is able to use a single accessibility measurement per zone, instead of OD matrices. These accessibility indicators are calculated by the "logsum" (the logarithm of the denominator of the logic probability) as follows:

$$A_i = \frac{1}{\beta_{Scale}} \cdot ln(\sum_{j=1}^{J} (W_j \cdot exp(-\beta_{Scale} \cdot c_{ij}))) \tag{4.8}$$

where, A_i is the workplace accessibility at location i, $i \in I$ the origins, $j \in J$ the destinations, β_{Scale} is a scale factor related to the scale of a logic model, W_j is a weight giving the number of jobs at location j, $exp(-\beta_{Scale} \cdot c_{ij})$ is a deterrence function, c_{ij} is the generalized travel cost from location i to location j.

The generalized travel cost c_{ij} is:

$$\begin{split} c_{ij} = (\alpha \cdot ttime) + (\beta \cdot ttime^2) + (\gamma \cdot ln(ttime)) + (\delta \cdot tdistance) + (\epsilon \cdot tdistance^2) + \\ (\zeta \cdot ln(tdistance)) + (\eta \cdot tcost) + (\theta \cdot tcost^2) + (\iota \cdot ln(tcost)) \end{split} \tag{4.9}$$

where $ttime$ is the travel time in minutes, $tdistance$ is the distance in meters, $tcost$ is the monetary travel cost, α to ι are the marginal utilities

In this study, the default values of MATSim were used: $\beta_{Scale} = 1$ and $\alpha = -12$. The other values were set to zero.

Home and work locations are distributed randomly on the nodes within each zone, in order to avoid that all the household and workplace locations are attached at the same link of the road network. Another option would be to be distributed to the nodes in a given distance from the zone centroid, however, because of the uneven sizes of zones in our case, the first method was selected.

Real Estate Price Model

The Real Estate Price Model (REPM) used in UrbanSim, is a semi-log linear regression based on Ordinary Least Squares (Franklin and Waddell, 2003; Waddell and Ulfarsson, 2003). It predicts the average value per unit, for every year of the simulation (eq. 1).

The results for the real estate price model are shown in Table 4.3. Two submodels were estimated for the REPM: one for the houses (detached, semi-detached and attached), and one for the apartments. There were no available observations of non-residential real

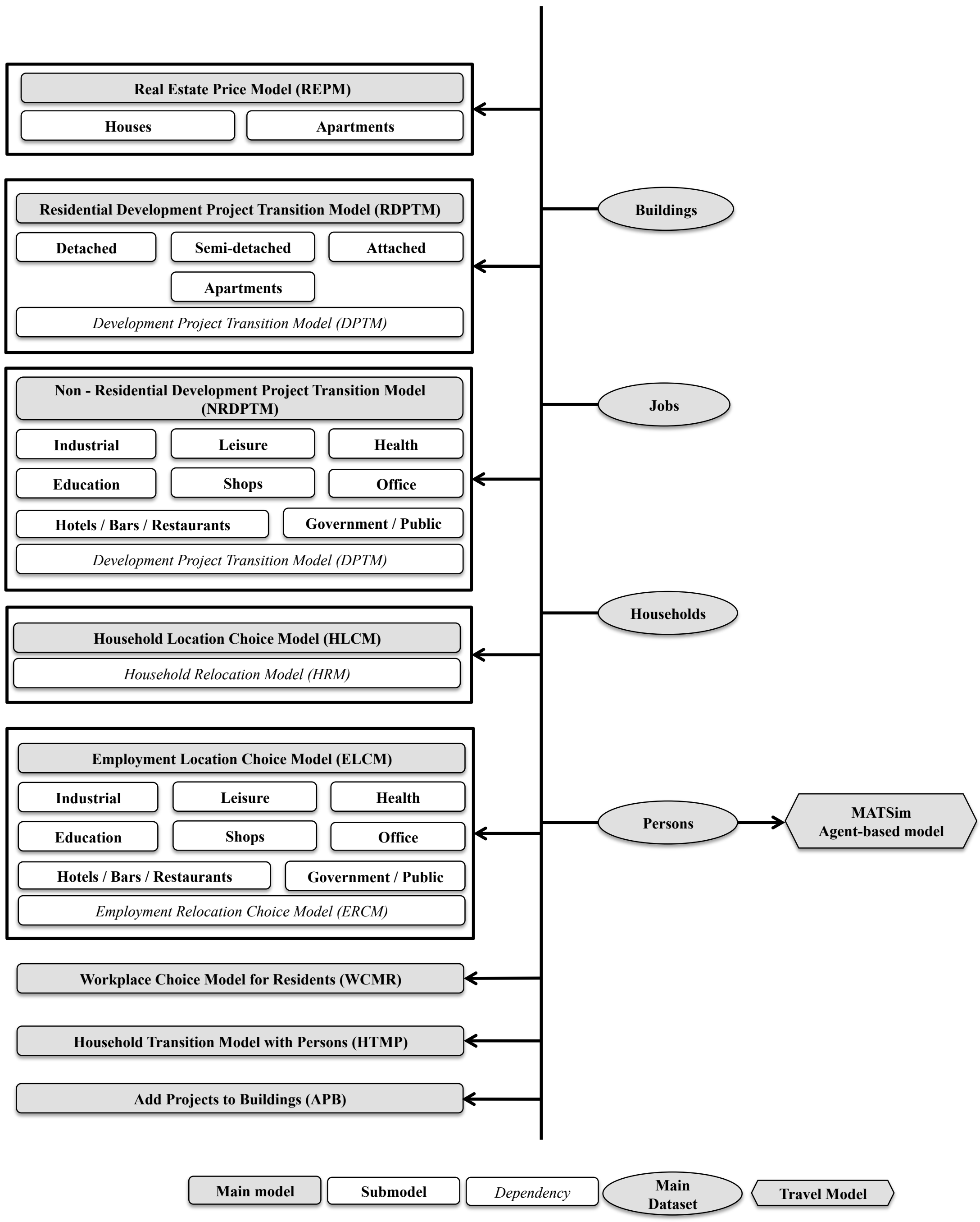


Figure 4.7: Flow diagram of UrbanSim & MATSim LUTI Model of Brussels

estate prices, and therefore no model was estimated for this case.

The price of houses and apartments is positively affected by the car accessibility of the zone and the percentage of green areas in the commune. Sociodemographic characteristics that have a positive impact are the percentage of households with high income in the commune, and the logarithm of the population density.

In order to avoid endogeneity issues, an instrument variable (De Palma and Picard, 2005) was included in this model's specification. This is the communal housing tax, which is a percentage of the dwelling's price per year, and has a negative effect.

The real estate price model presented in Table 4.3 is mostly based on location (neighborhood or commune) attributes. The only building-specific attribute used is the residential surface, which is positive. Despite the fact that the literature shows that prices are largely explained by attributes of the buildings (Löchl and Axhausen, 2010; Efthymiou and Antoniou, 2013a), the presented models are still able to capture land use effects that should be relevant for the modeling purposes.

Table 4.3: Real estate price model

Houses (n=14835)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
constant	-			11.5407	857.94
$\beta_{\text{car-acc}}$	Car accessibility	zone	%	0.0020	4.09
β_{green}	Green area score	commune	0 to 1	0.1349	10.81
$\beta_{\text{income-high}}$	Percentage of high income (>3) households	commune	%	0.0260	60.02
β_{tax}	Housing tax	commune	%	-0.0681	-47.75
$\beta_{\text{pop-den}}$	Logarithm of population density	commune	ln(pop/hectare)	0.0591	56.33
β_{sqm}	Surface	building	m^2	0.0005	8.751
$\boldsymbol{R}^2$=0.59					
Apartments (n=4945)					
Name	**Definition**	**Level**	**Unit**	**Coefficient**	**t-value**
constant	-			11.2914	368.69
$\beta_{\text{car-acc}}$	Car accessibility	zone	%	0.0046	4.09
β_{green}	Green area score	commune	0 to 1	0.4128	14.24
$\beta_{\text{income-high}}$	Percentage of high income (>3) households	commune	%	0.0225	22.67
β_{tax}	Housing tax	commune	%	-0.0334	-10.13
$\beta_{\text{pop-den}}$	Logarithm of population density	zone	ln(pop/hectare)	0.0020	1.82
β_{sqm}	Surface	building	m^2	0.0002	1.89
$\boldsymbol{R}^2$=0.31					

Household Location Choice Model (HLCM)

The Household Location Choice Model (HLCM) estimation results are presented in Table 4.4. All parameters are statistically significant and have the expected signs. Price has a negative effect in the utility for all households, no matter their income level. The presence of high income households attracts other households of high income but makes locations less affordable for low income households. Households with university degree holders prefer to be located in zones with high ratio of university degree holders. This is consistent with the expected social agglomeration and segregation effects usually observed in residential location.

Car accessibility increases the utility of car-owning households. Households with workers prefer to be located closer to the central business districts while households not owning any cars select locations close to the rail stations. Communes with with percentage of green areas are more attractive.

The spatial alternative-specific constant that accounts for unobserved attributes, indicates the attractiveness of the central locations. This constant is active when the location is inside the Brussels Capital Region.

Table 4.4: Household location choice model

(n=48526)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{car-access}}$	Households with car * Car accessibility	household * zone	0 or 1 * (logsum)	0.0106	2.95
β_{educ}	Household with high education level * Ratio of university degree holders in zone	household * zone	0 or 1 * ratio	3.6401	27.97
β_{green}	Green area score	commune	0 to 1	0.1924	2.62
$\beta_{\text{income-low}}$	Households at income class 1 or 2 * Ratio of hh with high income over all hh	household * zone	0 or 1 * ratio	-2.8948	-13.12
$\beta_{\text{income-high}}$	Households at income class 4 or 5 * Ratio of hh with high income over all hh	household * zone	0 or 1 * ratio	4.8074	12.95
β_{workers}	Households with workers * Log distance from CBD	household * zone	0 or 1 * ln(meters)	-0.1453	-12.44
β_{rail}	Households without cars * Distance from rail station <1000m	household * zone	0 or 1 * meters	0.3681	12.82
β_{price}	Logarithm of transaction price	building	log(euros)	-1.0298	-29.09
ASC_{BCR}	Central Brussels area (central communes)	commune	0 or 1	0.8738	44.64
Log-likelihood=−95751					

Employment Location Choice Model (ELCM)

The Employment Location Choice Model (ELCM) is subdivided in eight submodels, one for each type of economic activity. Table 4.5 shows the estimation results for each submodel. Jobs in the agricultural and mining sectors are not considered for modeling purposes, because of their low share overall jobs.

Employment location choice of each sector is positively affected by the density of jobs of the same sector in commune. The logarithm of non-residential surface and the car accessibility in zone, have a positive impact, when significant. The results for the industry jobs location sub-model shows that there is a negative effect of the density of jobs in the zone, which is observed to be the case only in this particular sub-model. Office jobs prefer to locate in zones with agglomeration economies and therefore favor density of jobs of the same type. Job density also has a positive effect in the utility for office jobs, probably because office jobs are service providers and prefer to locate near potential clients. Retail jobs also have benefits from the agglomeration economies and therefore the presence of jobs of the same type and of jobs in general have a positive effect in their location preferences. Finally, it is noted that health and leisure activities prefer to be located in communes with high percentage of high income households.

Table 4.5: Employment location choice model

Name	Definition	Level	Specification	Coefficient	t-value
Industry (n=13943)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{job-den}}$	Logarithm of job density	zone	ln(jobs/hectare)	-0.0627	-7.43
β_{sam}	Logarithm of non residential surface	building	ln(m^2)	1.2514	118.65
$\beta_{\text{ind-den}}$	Density of jobs in industry sector	commune	jobs/hectare	0.0782	27.47
Log-likelihood=-13634					
Office (n=14937)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{gov-den}}$	Density of jobs in public sector	commune	jobs/hectare	-0.0212	-6.36
$\beta_{\text{off-den}}$	Density of jobs in private sector (office)	commune	jobs/hectare	0.0152	4.93
$\beta_{\text{job-den}}$	Logarithm of job density	zone	ln(jobs/hectare)	0.6641	70.19
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	-0.0057	-12.08
β_{sqm}	Logarithm of non residential surface	building	ln(m^2)	0.5227	72.36
Log-likelihood=-22791					
Retail (n=3886)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{car-access}}$	Car accessibility	zone	logsum	0.0384	3.50
$\beta_{\text{ret-den}}$	Density of jobs in retail sector	commune	jobs/hectare	0.1643	4.43
$\beta_{\text{job-den}}$	Logarithm of job density	zone	ln(jobs/hectare)	0.0780	5.09
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	-0.0036	-2.19
β_{sqm}	Logarithm of non residential surface	building	ln(m^2)	0.8906	51.24
Log-likelihood=-6443					
Hotels/Bar/Restaurants (n=2013)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{car-access}}$	Car accessibility	zone	logsum	0.0427	3.21
$\beta_{\text{job-den}}$	Logarithm of job density	zone	ln(jobs/hectare)	0.3854	22.82
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	-0.0076	-7.10
β_{sqm}	Logarithm of non residential surface	building	ln(m^2)	0.3377	23.73
$\beta_{\text{hbr-den}}$	Density of jobs in hotels/bar/restaurants	commune	jobs/hectare	0.2018	10.44
Log-likelihood=-4923					
Government and public service (n=8471)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{off-den}}$	Density of jobs in private sector	commune	jobs/hectare	0.0125	6.69
$\beta_{\text{job-den}}$	Logarithm of job density	zone	ln(jobs/hectare)	0.7523	58.37
$\beta_{\text{pop-den}}$	Logarithm of population density	commune	ln(pop/hectare)	-0.0045	-7.69
β_{sqm}	Logarithm of non residential surface	building	ln(m^2)	0.5081	44.25
Log-likelihood=-10973					
Education (n=3775)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{edu-den}}$	Density of jobs in education sector	commune	jobs/hectare	0.2208	14.08
$\beta_{\text{job-den}}$	Logarithm of job density	zone	ln(jobs/hectare)	0.1824	11.37
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	-0.0075	-7.65
β_{sqm}	Logarithm of non residential surface	building	ln(m^2)	0.8405	46.04
Log-likelihood=-5995					
Health (n=5099)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{high-inc}}$	Percentage of households in high income scale (>3)	commune	%	0.0564	9.86
$\beta_{\text{hea-den}}$	Density of jobs in health sector	commune	jobs/hectare	0.1832	17.10
$\beta_{\text{job-den}}$	Logarithm of job density	zone	ln(jobs/hectare)	0.3708	32.00
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	-0.0129	-13.04
β_{sqm}	Logarithm of non residential surface	building	ln(m^2)	0.4908	41.29
Log-likelihood=-10493					
Leisure activities (n=1315)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{high-inc}}$	Percentage of households in high income scale (>3)	commune	%	0.0837	6.58
$\beta_{\text{leiz-den}}$	Density of jobs in leisure sector	commune	jobs/hectare	0.2978	16.43
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	0.0133	11.17
β_{sqm}	Logarithm of non residential surface	building	ln(m^2)	0.6327	27.56
Log-likelihood=-2349					

Residential development project location choice model (RDPLCM)

Estimation results for the Residential Development Location Choice Model (RDPLCM) are presented in Table 4.6. The models are estimated using data for real estate developments that took place in the ten year period previous to the base year (2001) and, therefore, are not representative of all existing supply in the city. All types of residential development tend to agglomerate and therefore have a positive parameter for the logarithm of the number of buildings of the same type. The dwelling categories "semi-detached" and "attached" were grouped for the purpose of this research, because of their similar characteristics.

Residential buildings are developed in zones with high price number of residential units. The population density of the commune has a positive impact for semi-detached, attached and apartments, but is insignificant for detached houses.

Non-residential development project location choice model (NRDPLCM)

The Non-Residential Development Project Location Choice Model (NRDPLCM) models the location of the developed non-residential projects. Eight sub-models were estimated in UrbanSim, one for each of the building types: 1) industrial, 2) office, 3) shops, 4) hotels/bar/restaurants, 5) government and public service, 6) education, 7) health, 8) leisure activities. Since there were not a significant number of development projects in the past, sub-models regarding quarrying and agricultural buildings were not estimated.

The estimation results for the location choice model of non-residential real estate developments are shown in Table 4.7. New non-residential supply tends to locate in places that already show agglomeration and with high concentration of other activities in general. Locations with good car and public transport accessibility tend to be attractive for the location of new developments.

Development of projects of buildings that host private services is negatively affected by the population density of the commune and the logarithm of the total population in the zone, and positively by the logarithm of total number of jobs in zone. Another factor that affects the development of retail buildings is the number of jobs in the zone. The more jobs, the more preferable the zone is for the development of such infrastructure. The number of citizens in a zone is a positive determinant of the location, while the density at a commune level is negative.

Workplace choice model for residents (WCMR)

This model assigns jobs to workers of the households. For its estimation, a table with each individual person with information about its household and at the base year (household_id and job_id), was created. The WCMR contains two variables: the car accessibilities per zone, which is positive, and the distance between the house and work location of the person, which is negative.

Table 4.6: Residential development project location choice model

Name	Definition	Level	Specification	Coefficient	t-value
Detached (n=59558)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
β_{price}	Logarithm of price of detached houses	building	ln(euros)	1.5334	59.31
β_{units}	Logarithm of number of detached house units	building	ln(sum)	1.6578	338.21
Log-likelihood=-160082					
Semi-detached and Attached (n=20119)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
β_{price}	Logarithm of price of semi-detached and attached houses	building	ln(euros)	0.3013	7.06
β_{units}	Logarithm of number of semi-detached and attached house units	building	ln(sum)	1.1172	164.97
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	0.4097	37.48
Log-likelihood=-58729					
Apartments (n=5119)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
β_{price}	Logarithm price of apartments	building	ln(euros)	0.1823	2.38
β_{units}	Logarithm of number of apartment units	building	ln(sum)	0.1823	2.38
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	1.0609	85.62
Log-likelihood=-12286					

Table 4.7: Non-residential development project location choice model

Industry (n=2770)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{car-access}}$	Car accessibility	zone	logsum	-0.0659	-6.25
$\beta_{\text{ind-den}}$	Density of jobs in industrial sector	commune	jobs/hectare	-0.0705	-10.32
$\beta_{\text{ln-jobs-zone}}$	Logarithm of total number of jobs	zone	ln(sum)	0.4251	33.19
Log-likelihood=-10949					
Office (private sector) (n=767)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{car-access}}$	Car accessibility	zone	logsum	0.0906	3.61
$\beta_{\text{off-den}}$	Density of jobs in private sector	commune	jobs/hectare	0.0269	3.32
$\beta_{\text{pop-den}}$	Population density	commune	pop/acre	-0.0318	-5.86
$\beta_{\text{ln-jobs-zone}}$	Logarithm of total number of jobs	zone	ln(sum)	1.1741	34.43
$\beta_{\text{ln-pop-zone}}$	Logarithm of total number of population	zone	ln(sum)	-0.1460	-5.30
Log-likelihood=-1953					
Shops (n=1466)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{ln-jobs-zone}}$	Logarithm of total number of jobs	zone	ln(sum)	0.4451	21.84
$\beta_{\text{ln-pop-zone}}$	Logarithm of total number of population	zone	ln(sum)	0.3899	12.62
Log-likelihood=-5451					
Hotels, bar, restaurants (n=107)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{car-access}}$	Car accessibility	zone	logsum	0.2219	2.73
$\beta_{\text{hbr-den}}$	Density of jobs in hotels/bar/restaurants	commune	jobs/hectare	0.1600	1.56
$\beta_{\text{pop-den}}$	Population density	commune	pop/hectare	-0.0365	-2.89
$\beta_{\text{ln-jobs-zone}}$	Logarithm of total number of jobs	zone	ln(sum)	0.7093	8.81
Log-likelihood=-359					
Government and public service (n=264)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{ln-jobs-zone}}$	Logarithm of total number of jobs	zone	ln(sum)	0.7184	15.57
$\beta_{\text{ln-pop-zone}}$	Logarithm of total number of population	zone	ln(sum)	0.1059	2.24
Log-likelihood=-932					
Education (n=140)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{ln-jobs-zone}}$	Logarithm of total number of jobs	zone	ln(sum)	0.3591	6.21
$\beta_{\text{ln-pop-zone}}$	Logarithm of total number of population	zone	ln(sum)	0.3539	5.77
Log-likelihood=-533					
Health (n=225)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{ln-jobs-zone}}$	Logarithm of total number of jobs	zone	ln(sum)	0.3523	5.45
$\beta_{\text{ln-pop-zone}}$	Logarithm of total number of population	zone	ln(sum)	0.5394	7.55
Log-likelihood=-840					
Leisure activities (n=970)					
Name	**Definition**	**Level**	**Specification**	**Coefficient**	**t-value**
$\beta_{\text{car-access}}$	Car accessibility	zone	logsum	-0.0240	-1.70
$\beta_{\text{lei-den}}$	Density of jobs in leisure sector	commune	jobs/hectare	3.2041	8.41
$\beta_{\text{ln-jobs-zone}}$	Logarithm of total number of jobs	zone	ln(sum)	0.2371	11.14
Log-likelihood=-3867					

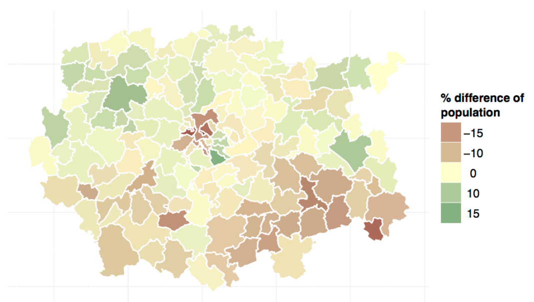
4.3.3 Simulation – validation

For validation purposes, a baseline scenario between the years 2001 and 2013 was simulated. The simulation results are compared with real data, acquired from the official Belgian databases. The results indicate the strengths and the weaknesses of a zone-level integrated land-use and transport model in predicting the real socioeconomic and transport situation. The time interval for each simulation was set to 1 year for UrbanSim, and 9 years for MATSim (2001, 2010 and 2019). A 10% of the agents that use car for trips to work was randomly selected for the traffic simulation using MATSim.

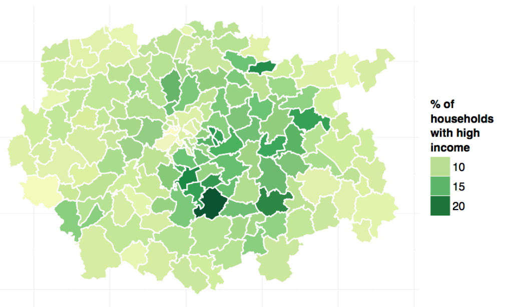
Figures 4.8i, 4.9i and 4.9ii demonstrate the efficiency of the model in predicting, with considerable accuracy, the real population in the great majority of the communes. However, the model under-predicts the population in the central communes *Brussels*, *Schaerbeek*, *Sint-Jans-Molenbeek* and *Saint-Josse-ten-Noode*, and the south-eastern *Wasseiges*. This can be explained by the under-prediction of the house prices in the city center, affecting the real estate development model, which also from its side then under-predicts the generation of new supply in those communes.

Despite the model's success to capture the increased trend of low to medium prices in 2008, as indicated in Figure 4.9iii, it fails to predict the higher values. These are mainly observed in the central commune *Brussels*, where the difference goes beyond 20% in 2008, and the south-eastern commune *Lasne*, which has the highest percentage of households with high income, as shown in figure 4.8ii. Figure 4.9iv shows that the real estate price submodel of the apartments is even less accurate in predicting the average transaction prices of 2008, which occurs because of the low R^2=0.31 of this particular submodel. In general, despite the fact that the real estate price model considers socioeconomic dynamics, the hedonic price model partially fails to fulfill its purpose with success, probably because of the following reasons: 1) the hedonic price model does not capture market effects, such as supply or demand surplus (Hurtubia et al., 2012), 2) the first come - first served market clearing mechanism is oversimplified and may introduce a bias in the location choice of households and jobs, and 3) the ordinary least squares (OLS) fail to capture the spatial autocorrelation, an issue that can be solved by spatial econometric models (Sections 3.2, 3.3, 3.4 and 3.5). Another reason could be the lack of variables related with building-specific attributes, as they were not available for the particular case study presented in this research.

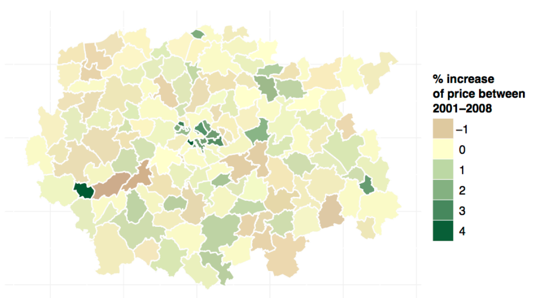
Figure 4.8vi shows the car accessibilities per zone, which is the output of MATSim, in the year 2008. For the majority of the zones, the higher the distance from the city center, the lower its accessibility –with some exceptions in the outskirts of the area of study.



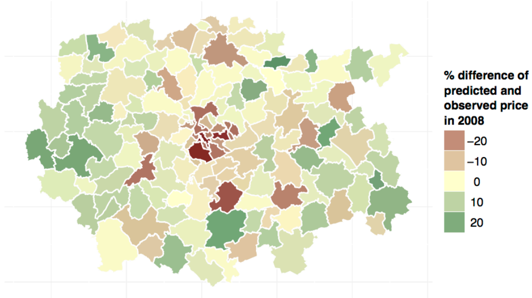
(i) Difference of Observed and Simulated Population in year 2008



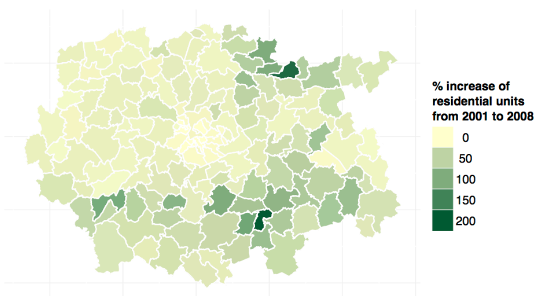
(ii) Percentage of Households with High Income in 2008



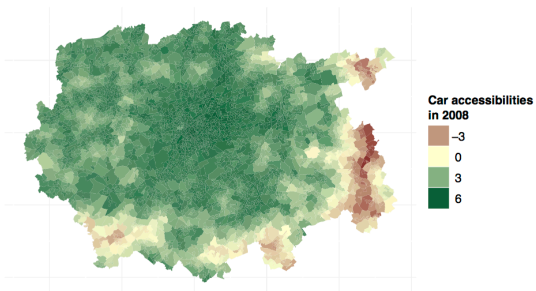
(iii) Increase of Price Between 2001 and 2008



(iv) Difference of Predicted and Observed Price in 2008

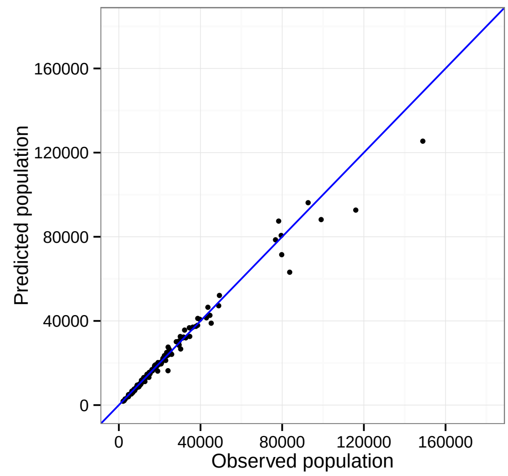


(v) Increase of residential units 2008

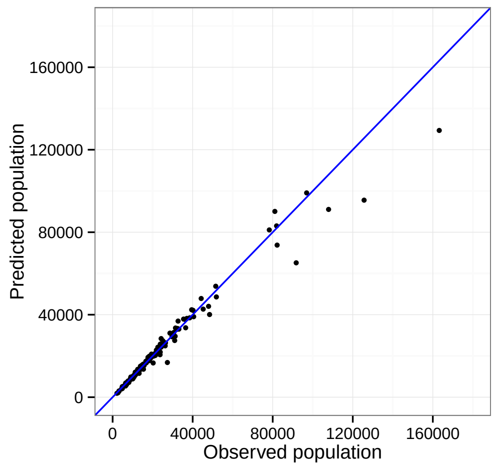


(vi) Car accessibility in 2008

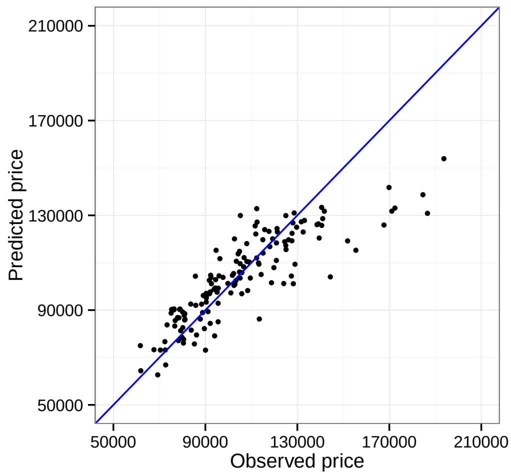
Figure 4.8: Basecase scenario validation 1/2



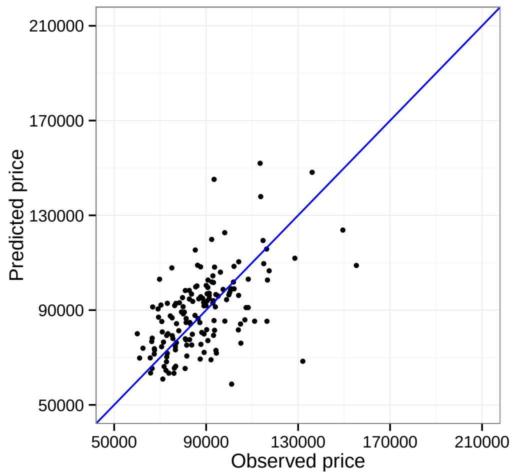
(i) Comparison of observed and predicted population in year 2008



(ii) Comparison of observed and predicted population in year 2011



(iii) Comparison of observed and predicted prices of houses in year 2008



(iv) Comparison of observed and predicted prices of apartments in year 2008

Figure 4.9: Basecase scenario validation 2/2

4.3.4 Conclusions

In this section the development and implementation of an Integrated Land-Use and Transport Model for the city of Brussels, Belgium, has been presented. The model is developed using the agent-based microsimulation platforms UrbanSim and MATSim. The objective is to estimate the individual required models, interpret the estimated coefficients and identify the strengths and weaknesses of the modeling approach. This is done by validating the results of a base-case scenario, comparing them with available observed data and examining the possible margins of improvement.

The individual transition, development, relocation, location choice and price models were estimated, and a base-case scenario from 2001 to 2013 was then simulated. Results show that the model succeeds in predicting the spatial distribution of location of new households; however it tends to underestimate the prices for communes were a significant increase was observed. This deviation may be explained by the use of an ordinary least square hedonic model for the real estate prices, making them dependent on attributes of the location but independent of market conditions. Another possible cause of the error in price forecast is the underestimation of dynamic variables that explain the price, like the income distribution or location of other agents in a zone, that could be due to the simplified market clearing process (first come first served) considered by UrbanSim. Identification of the causes of this error is matter of future research, but some potential ways to improve the modeling results are to include a more realistic market clearing mechanism (Hurtubia and Bierlaire, 2012), or the use of spatial autoregression models for the real estate price (Sections 3.2, 3.3, 3.4 and 3.5).

Although, the question of whether the added value of the integrated aspect of the analysis outweighs the increased complexity is not directly answered, a clear advantage that emerges is the flexibility for policy evaluation. This benefit will be exploited in the next chapters of this dissertation.

4.4 Indicator–based urban quality analysis

4.4.1 Background

Current urban development processes coming along with increasing population growth in many urban areas worldwide results in reduced urban quality and quality of life of urban dwellers (Alberti, 2005). This includes quality aspects such as place qualities in economic development, sustainable resource use, quality of life in places as well as the regional and local identity in the (inter-)national context (Albrechts et al., 2003). How can the transition of existing spatial urban patterns be managed so that they comprise adequate urban assets (e.g., public space, infrastructure, key services, the right mix of activities and people, adequate housing etc.) suitable to serve human life quality also for future generations? Currently, great efforts are taken to develop new strategies and guidelines for transforming agglomerations into sustainable urban patterns (UN-Habitat, 2013).

It is widely acknowledged that trans-disciplinary processes of science practice collaboration, across disciplines and involving local actors, are required for enhancing urban quality because of the numerous actors, perspectives and conflicting values and interests at various spatial scales influencing urban development (Albrechts et al., 2003; Pacionne, 2003; Van Kamp et al., 2003; Healey, 2007; Steinitz, 2012). Furthermore, there is a need for planning at strategic stages based on a scientific understanding of the quality of landscape patterns for identifying undesired developments prior to actual design of development plans (Brown, 2003; Gordon et al., 2009; Steinitz, 2012). Particularly tools and indicators are required that show ecological, economic and social quality aspects in urban landscapes interact and change with scale (Van Kamp et al., 2003; Wu, 2010; Jenks and Jones, 2010a; Dempsey et al., 2012).

Regions are responsible for translating policy targets concerning housing, transport, the environment, and regional aspects of economic policy to realities of urban development (Albrechts et al., 2003; Nuissl et al., 2009). Implementing integrated planning still is a challenging task for regional planning actors since high complexity in the analysis of the urban system is required for sound political decision making (Alberti, 2005; Waddell, 2011). Spatial scenarios offer a promising method to reflect on possible urban future states from different perspectives (Wiek and Scholz, 2006). They offer an integrated, interdisciplinary approach to the analysis of possible future urban patterns. In this context, integrated behavioural land use and transportation models are recently emerging that can show resulting land use patterns of certain policy inputs and provide spatially explicit socioeconomic indicators for policy evaluation (Waddell, 2011; Nicolai T. and Waddell, 2011). However, bringing these models into the operational planning context and integrating them into the participatory decision-making process still requires significant efforts (Waddell, 2011). Brown (2003) points out that the major task is to prove that the provided information is useful to practitioners, that is to meet the needs of the planning actors with regard to the information fitting into the planning process and delivering evidence for decision-making.

In this research, we focus on an approach that allows actors in trans-disciplinary planning processes for cross-scale analysis of urban development scenarios addressing economic, social and environmental aspects of urban quality. The main objective is to identify practicable interfaces for implementing an integrated behavioral modeling system in order to calculate a set of spatially explicit urban quality indicators that are meaningful for the planning actors. Research questions addressed in this section are:

1. How to translate qualitative scenario information on urban development into input data for an agent based land use and transportation modelling system?

2. Which indicators can be calculated based on the modelling output and are meaningful for real-world spatial decision-making processes?

3. What kind of evidence do the results provide on possible urban patterns that can occur on different spatial scale?

4.4.2 Methodology

Framework for integrated analysis of urban patterns quality

Pacionne (2003) states that “in order to obtain the goal of a livable city, a wide range of social, economic and environmental needs must be satisfied”. This research refers to to the concept of the socio-environmental system and a set of basic human needs that the urban environment should fulfill for urban dwellers’ life quality (Smith et al., 1997; Matsuoka and Kaplan, 2008; Dempsey et al., 2012). It is used as a framework to translate specific conditions of urban patterns into indicators for assessing the pattern’s quality (for a review of other frameworks, see e.g. Van Kamp et al., 2003). Urban patterns are understood here as built-up areas, open space and infrastructure patterns tightly intermeshed with patterns of population groups and their travel patterns constituting the landscapes of cities and their agglomerations (Nassauer, 2012).

Urban patterns can be linked to dimensions of human needs and goals. Basic human residential needs include, e.g. regeneration, privacy or economic viability (Matsuoka and Kaplan, 2008; Dempsey et al., 2012). Concrete spatial expressions are required to fulfill these needs, e.g. recreational areas such as parks, private living space, affordable apartments or accessibility provided by the transportation system. A lot of these elements and patterns can be manipulated to meet the needs through urban planning and design and guided by regulatory or investment programs (Pacionne, 2003). However, implementation of measures to fulfill the demands occurs on multiple governance hierarchies, which have to be addressed in the analysis of their effects (Nassauer, 2012).

Urban regions are highly interlinked socio-environmental systems with human actions and natural habitats both affecting socioeconomic and ecological processes at various scales (Alberti, 2005; Nassauer, 2012). This means that enhancing the provision of certain services on one scale might reduce the provision of other services on another scale. For example higher housing density providing sufficient housing space for people can have an impact on the biodiversity on the regional level (Nuissl et al., 2009), on the access to open space for recreational purposes in walking distance on the district level (Dempsey et al., 2012), and on the social mix of the population on the local neighbourhood level (Keddi and Tonkiss, 2010). Due to the effects occurring across scale the impact of certain demands on other demands is not directly obvious and hardly assessable without integrated evaluation tools (Brown, 2003).

Objectives for sustainable urban development can evolve through the collaborative interpretation of a certain case and its possible transformations (Brown, 2003). Therefore, in this research a scenario is analyzed to show the effects of an alternative demand set on the fulfillment of all other needs. The aim is to translate qualitative information of integrated, interdisciplinary urban development scenarios into quantitative data as input for an integrated behavioral and transport modeling system. The modeling results provide spatially explicit information as well as enhanced spatial scenario data for further analysis of socio-economic and ecological aspects on the regional level and on other scales. A set of indicators was selected that reflects the fulfillment of basic human needs by the urban fabric. The indicator results are prepared in form of maps suitable for a spatially explicit analysis. This allows for cross-scale analysis of the scenario's effects.

Case study – a suburban agglomeration

The approach is illustrated in the case of the region Limmattal, a suburban agglomeration close to Zurich, Switzerland. The valley between the two city centers of Zurich and Baden is very typical for the unexpected spatial results of a growth process hardly coordinated at all. The economic and demographic expansion over the last 60 years consumed the landscape and overrode existing settlement structures. However, planning was not able to replace them by other convincing spatial structures. The region has an area of 85 km^2 and about 165,000 inhabitants and 118,000 employees in the year 2010. Its landscape is both dominated and fragmented by extensive, (inter-) nationally significant infrastructure in the form of linear transport axes. Current urban development is affected by high rates of new inhabitants (+12,200 meaning +10% from 2000-2010), comprising 32.6 per cent of non-Swiss citizens compared to the average of 22.4 per cent in all of Switzerland.

The target group collaboratively working together in this study, was a group of geographically oriented natural and social scientists (ecologists, sociologists, transportation and land use modeler), planners, urban designers and local stakeholders with different socio-economic and environmental demands. Spatial scenarios of possible urban development patterns in the Region Limmattal were developed, for identifying sustainable patterns with regard to the options of spatial conversion, economic productivity requirements, and environmental and social sustainability, in particular the design and the perception of public space and the provision of living space. The aim was to become aware of the broad constraints and possibilities concerning urban development. Furthermore, different aspects of urban quality were considered simultaneously in the earliest stage of the planning process.

The scales that were of importance for the target group can be characterized by both physical aspects and institutional hierarchies associated with policy objectives (Schaefer, 2011). On the regional level, the structure determined by the topography, the allocation of land uses and land use densities (zoning), and the infrastructure is controlled by spatial and infrastructure planning. On the city or district level the municipality can guide urban transformation by reserves of building zones (development potentials), which influence identity-giving qualities of quarters and settlements. On the local or neighborhood level, the urban design and the design of open spaces determine the culturally shaped form, which influence spatial as well as social characteristics (Schaefer, 2011). This scale between the municipality and the building level can serve as a meeting point between public and private sectors because it is no legal level but it is large enough to address concrete themes and actions that exceed the scale of a single property (Pérez and Rey, 2013).

On the regional level, the extent of the Region Limmattal was reduced to 8 municipalities located in the Canton of Zurich (Dietikon, Geroldswil, Oberengstringen, Oetwil an der Limmat, Schlieren, Unterengstringen, Urdorf, and Weiningen) comprising in total an area of 39 km2. Reducing the perimeter was necessary since the required base data of the current state was available for these municipalities only. One municipality, Dietikon (9.3 km2) was chosen for calculating indicators on the district level. Actually this municipality is a city. However, its area size is comparable to districts of the city of Zurich, e.g. the district Altstetten (7.5 km2), which is adjacent to the Region Limmattal. In the municipality analyzed on the district level, one neighbourhood was selected for calculating indicators on the local level. The neighborhood 'Niderfeld' in Dietikon comprises an area of approx. 0.45 km2 and is defined by a given plot for settlement development.

4.4.3 Scenario modeling and analysis

Qualitative regional scenarios

Input to the modeling approach was one scenario that was selected out of four existing qualitative regional scenarios. These stem from a scenario study implementing formative scenario analysis (Scholz and Tietje, 2002) with systematic science-practice collaboration. Of the region Limmattal the stakeholders comprised practitioners from different planning scales and various sectors (e.g., local and cantonal authorities, real-estate developer, building technology companies, infrastructure planner, regional planning associations, nature protection, social or cultural organizations).

The scenario storylines depict four different view axes giving more weight to either design, technological, economic or ecological aspects respectively and thus determine the nature of the urban development in the region Limmattal by the year 2030. In the selected scenario "Smart City", Limmattal is characterized by highest possible energy efficiency through high densities of services and short distances as well as an attractive modal-split infrastructure design. These are all interventions that are currently called for in order to secure a more sustainable urban development (Dempsey et al., 2012; Pérez and Rey, 2013; UN-Habitat, 2013) and thus their possible impact is of high interest for the stakeholders.

Translating qualitative scenarios into an urban modeling system

The scenario “Smart City” was analyzed to show the effects of the alternative demand set on the fulfillment of a set of basic human needs across scale implementing a behavioral and transport modeling system. A spatially explicit scenario information format was required as input to this system that allowed for an in-depth analysis with complex simulation programs. Therefore, we translated the qualitative storyline into ESRI’s shapefile format.

Required input to the simulation is a zoning plan on parcel level. The input parcels of the current state contained already information on the land use designation and the allowed floor area ratio. Using ESRI’s ArcGIS (ESRI, 2011), these attributes were changed for all parcels according to a consistent logic that the group of scientists developed for all the four scenarios. The result was a shapefile of the parcels’ zoning information for the scenario, that is, in the terms of the practitioners a ‘revised zoning plan’.

The parcel shapefile of the scenario was then used as input to the open-source UrbanSim model (Waddell, 2011). For the implementation of UrbanSim for the Region Limmattal, the different agent location choice, real estate price and supply development models were first estimated.

The control totals of households (56,195) and jobs (61,039) in the year 2030 for the scenario were defined by the regional stakeholders. They expect this increase of households and jobs within the given time frame. Their interest is to see the impacts of this increase according to the induced changes of land use designation and allowed floor area ratio in their scenario “Smart City”. The base year of the scenario simulation was 2000 and it was simulated until the year 2030. In the year 2020, a light railway line was put into operation.

Analysis of urban patterns’ quality across scale

A set of 10 indicators relating to demographic, social, economic and residential conditions has been determined and calculated from the input data for the current state in the year 2000 (base year) and the output of the agent-based simulation for the scenario in the year 2030. The indicator data has been prepared as maps or pie charts for three spatial scales subjected first to examine their correlation cross-scale. This was done to reveal patterns that highlight particular problems and problem areas as well as advantages of the urban development in the scenarios. In the following, further detail is given on how the indicators have been selected and calculated. There is much debate within the research community on which categories and types of indicators should be used. For more detailed information on how urban quality indicators can be selected, see (e.g. Alberti, 1996; McCool and Stankey, 2004; Pacionne, 2003).

The overall goal is that many parties involved in the regional strategic planning process should be able to grasp the concrete spatial differences between the possible development scenario and the current state. The indicator maps should thereby make the differences more visible than the storylines. The relevance of the indicators for practice and their tangibility for heterogeneous stakeholders in the case study site are therefore important aspects. Particularly key concerns to be faced by policies correlated to urban form, such as housing, accessibility, preservation of open spaces and recreation areas, social diversity, and economic viability (Jenks and Jones, 2010a; Schetke et al., 2012), should be addressed by the indicators. They should provide information on density, land use, public space and effectiveness of infrastructure and services, and mix of social groups useful for analyzing their contribution to human life quality. Furthermore, only indicators that could be calculated based on the modeling output have been taken into account. Thus, there is a limitation to objective indicators describing the environments within which people live and work. Finally, it has been decided to provide not too many indicators as this might be an information overload for stakeholders. Furthermore, communities tend to prefer defining a very limited set of indicators for detecting critical trends in their perimeter (Alberti, 1996). Three to four different indicators per level have been regarded as reasonable. Table 4.8 gives an overview on the indicators, which are defined below.

Table 4.8: Overview on selected indicators per spatial level

Level	Indicators
Regional level	**1. Accessibility:** Travel times by private car, by public transport, and by walking
(8 municipalities)	**2. Population density:** Number of people per hectare
	3. Population segregation: Theil index
	4. Available open space: Supply rate of inhabitants with recreational area
District level	**5. Economic viability:** Economic sustainability of commercial areas (local retailers)
(1 municipality)	**6. Residential density:** Number of dwellings per hectare
	7. Housing affordability: Rent prices per square meter of residential floor area
Local level	**8. Mix of uses:** Ratio of residential to commercial / industrial floor area
(1 neighborhood)	**9. Social mix:** Age structure of households
	10. Building density: Ratio of building footprint area to parcel area

Regional level

1. *Accessibility* can have various effects, e.g. on house prices, construction activity, population growth and composition of households (Section 2.5). There is some evidence that better accessibility supports greater use of local services in areas of higher density (Jenks and Jones, 2010b). In this research, the modeling output provides the travel times by private car and by public transport, which allow for an interpretation of the accessibility within the region Limmattal.

2. *Population density* has implications for the functioning of the urban ecosystems and for different aspects of social sustainability (Jenks and Jones, 2010b; Power, 2012; Dempsey

et al., 2012). The amount of inhabitants and employees is given as a direct output of UrbanSim per parcel. The spatial population density is presented as the number of people (inhabitants) per hectare for each parcel in the region.

3. *Population segregation* comes along with many negative effects, such as social unrest and crime due to negative socio-economic outcomes of isolated groups of low-income households (UN-Habitat, 2013). Therefore, the extent and concentration of poverty within a city is a relevant social indicator (Jenks and Jones, 2010b). It is related to the human needs for safety and sense of integration, which depend also on accessibility (Power, 2012). The Theil index decomposition has been implemented for measuring regional inequality of the population's wealth for the comparison of the scenarios. Based on the household characteristic "income" the economic inequality in the region is given per parcel (Section 2.5).

4. *Available open space*, and especially green space, is a key factor of ecosystem performance in urban areas and is thus very important in securing human life quality and creating sustainable urban patterns (Jenks and Jones, 2010a). As indicator with practical relevance, the supply rate of inhabitants with recreational area is calculated. This indicator is deployed in standard planning practice in the case study area (Grün Stadt Zürich, 2006) implementing the following formula:

$$\text{Supply rate (in \%)} = \frac{\text{Offer of public open space}}{\text{Simultaneous demand of open space}} \cdot \mathbf{100} \tag{4.10}$$

A supply rate of 100 per cent equates 8 square meters public open space per inhabitant, which is the guiding value declared by the City of Zurich. Public open space that is considered as suitable recreational area for inhabitants must have a minimum area of 2500 square meters. Furthermore, the capacity of the open areas based on its function is taken into account. Forests, fields and pastures (5 inhabitants / ha), parks (100 inhabitants / ha), sports grounds and cemeteries (20 inhabitants / ha) differ considerably in their general capacity for accommodating the inhabitants simultaneous demand of open space. Another precondition defined by the City of Zurich is that the public open space must be accessible within a walking-distance (400 m) from a parcel. Therefore, we implemented a network analysis in ESRI's ArcGIS using the walkable paths and streets of the street network as basis for calculating the catchment area of each suitable recreational area. With the actual amount of inhabitants in the respective catchment areas the overall supply rate of the residents with open space is determined. Further details of this indicator are described in Thalmann (2012).

District level

5. *Economic viability* can be measured from very different perspectives. From a social perspective, the accessibility to key services, such as supermarkets, is very important for social and community life (Jenks and Jones, 2010b). However, local retailers require a certain amount of inhabitants in their catchment area to be economically viable. In Switzerland, the local convenience stores providing basic supply for daily life require at least around 8,000 inhabitants in a walking distance for a shop sized 700 - 1,500 square meters. Neighborhood centers and discounters with full-range food supply and non-food supply require a minimum of around 20,000 inhabitants for a shop with 1,500 - 5,000 square meters (Research and Consulting, 2002; Ciari et al., 2008). For indicating the economic viability of potential local retailers in the scenarios modelling output, the commercial areas are first selected. Then the catchment area within 400 m for these areas is calculated implementing a buffer analysis in ESRI's ArcGIS. The rate of inhabitants in the catchment area as measured by the minimum requirement of inhabitants indicates the potential economic viability.

6. *Residential density* is often expressed in urban planning practice and policy targets as number of dwellings per hectare (Dempsey et al., 2012). It can be calculated directly from the modeling output of UrbanSim for each parcel. Increasing the residential dwelling density in suburbs is one strategy of compact city policies and the indicator can offer useful insights into the spatial implications (Chhetri et al., 2013).

7. *Housing affordability* is an important aspect of securing social equality since housing is a basic human need (Clarke, 2012). Urban transformations driven by fast growing population can cause accelerating housing market growth (e.g. due to a shortage of available land for development) and pressures particularly on sufficient supply of affordable housing for low-medium-income communities (Atkinson, 2009; Keddi and Tonkiss, 2010). Areas of renewal are often changed in their use, composition and image in order to attract more affluent residents leading to gentrification. Effects of this process can be displacement of lower-income groups and segregation of social groups or patterns of "micro-segregation" of diverse live-work and leisure social groups in a neighborhood coming along with competition for occupation of local spaces and services (Keddi and Tonkiss, 2010). Rent prices of dwellings can provide an indicator for the amount of affordable housing and thus the potential for such social processes in a district. Based on the UrbanSim modeling output, the mean monthly rent price (in CHF) per square meter of residential floor area for each parcel is calculated. They are classified into five evenly distributed classes from low to high rent prices (from 6 to 17 CHF/sqm without the inflation being taken into consideration).

Local level

8. *Mix of uses* can improve the easy access to work and basic local services and facilities, which is, e.g. highly valued by residents of densely populated Asian cities (Tu and Lin, 2008). Further it can increase opportunities for active and passive surveillance and

thus can enhance safety within neighborhoods (UN-Habitat 2013). The residential, commercial and industrial floor area per parcel is a direct modeling output. The ratio of residential to commercial and industrial floor area in the neighborhood is calculated as an indicator for the mix of working and living.

9. *Social mix* of residents is a characteristic often used for predicting urban areas of unsustainable housing as e.g. high proportions of ethnic minority residents are frequently associated with it (Cameron and Field, 2000). Age, however, also provides indications on the needs, which have to be taken into account. In practice, needs of marginal age groups such as children and the elderly are often neglected by housing development due to a stronger focus on middle-aged adults (Turcu, 2012). For the social composition of the neighborhoods it is provided an indicator on the age structure on the local level, by specifying the per cent of households according to four groups of age bands: 16-24, 25-49, 50-64, and over 64 years old. Please note that only the age of the head of the households is given in the simulation results.

10. *Building density* can be described as floor area ratio or as ratio of the building footprint area to the area of the plot Dempsey et al. (2012). Since the residential environment is important not only for recreational purposes but also for the more general neighborhood satisfaction (de Jong et al. 2012), and for enhancing social goods such as biodiversity and ecosystem services (Jenks and Jones, 2010b), the latter indicator is computed from the modeling output as it reflects the amount of open space. In the Canton of Zurich this ratio is usually given in per cent.

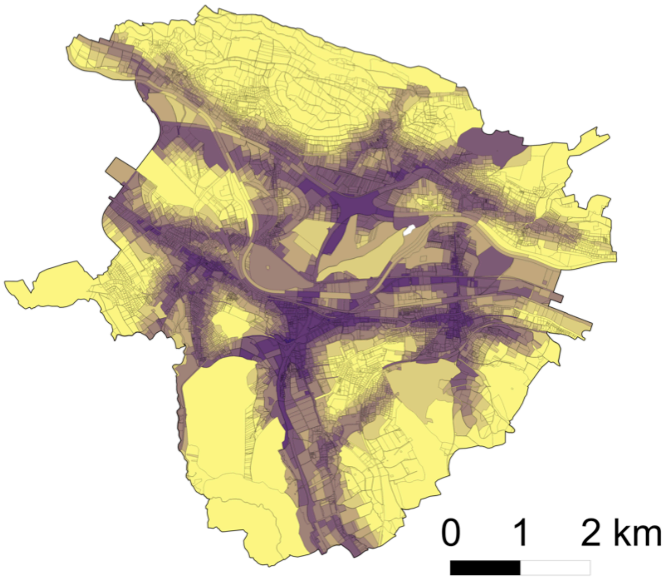
4.4.4 Results

The overview on all resulting indicator maps are provided in Figures 4.10 - 4.21. Figures 4.10 - 4.12 show the resulting indicator maps for addressing the hypothesis that higher accessibility has an impact on the viability of local services and the composition of households. In comparison to the current state, the quality of the accessibility increases strongly in the scenario “Smart City” in the middle of the region. This is where today the suburban railway is running through. In the scenario a new light railway line was put into operation in this area in 2020. Its positive effects on the accessibility are directly visible. Particularly in the western area of the region, the utility of the accessibility by public transport for the people increased significantly. Zooming on the district scale, there is evidence that there is a correlation of the enhanced accessibility with the existence of small retailers (Figure 4.16). Figure 4.14 shows the income inequality (Theil index) in years 2000 and 2030.

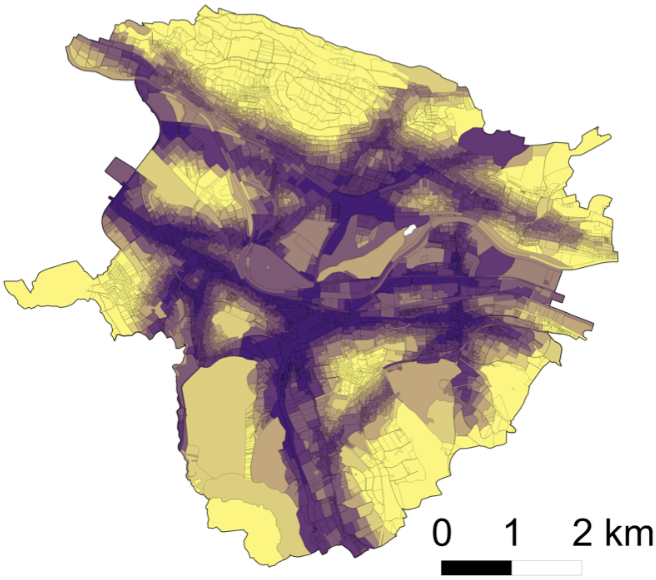
Based on the concept of livability, access to public open (green) spaces is very important for the quality of living (Figure 4.15). Population, residential and building density can have implications on its provision (Figures 4.16 and 4.17). While in the status quo only a few small commercial floor areas are available in the whole area of the municipality of Dietikon, there are much more in the scenario "Smart City" (Figure 4.16). Furthermore, the new areas for small retailers have rather high potential economic viability, in particular in the part of the municipality from the North West to the East. Since this is dependent on the amount of people in the catchment area of the retailers, this points to a high population density in that area (Figure 4.13).

Figure 4.18 shows the rent distribution across the area. Figures 4.19 and 4.21 provide indicator maps suitable to analyze effects of densities on local and district scale on the supply rate of the people with open space. In the current state the supply rate is rather good and only in a few areas the people suffer of open space. This is particularly the case in the core areas of the municipalities. However, in the scenario the areas with insufficient supply rate increased drastically. Looking at the densities on the selected district scale, there is a significant increase in dwelling density. The parcels showing densification stretch across the area from the North West to the East of the municipality. This is actually the area where there is not enough recreational open space available for the people. Further zooming-in on the local level (Figure 4.21), the density indicator displays the decrease of open space on parcel level due to the increase of building footprint area. Whereas the indicator on regional level is just pointing to the available public open space, the indicator on the local level highlights parcels where only little private open space is available. Areas that have low supply rates of both public and private open space cannot fulfill the recreational demands of the people. The cross-scale analysis of the selected indicators can effectively point to areas where the densification patterns cause loss in urban quality with regard to the recreational quality and, thus, potentially the degree of livability.

A further zoom-in on the local scale provides evidence on the shift in the population composition in the example of the neighbourhood of Niderfeld (Figure 4.20). In this perimeter the majority of the heads of the households are in the middle age (25 – 50 years old) and in the pension age (> 65 years old) in the year 2000. In scenario "Smart City" there is a strong decrease of households of the latter age group in favor of the middle age group and the age group in between those two (50 – 65 years old). From these patterns there is some evidence that the higher accessibility due to the new light railway line supports the viability of local retailers and leads to a composition with less old households in the selected perimeters.



(i) Status quo 2000



(ii) Smart city 2030

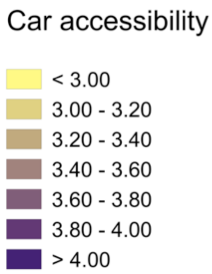
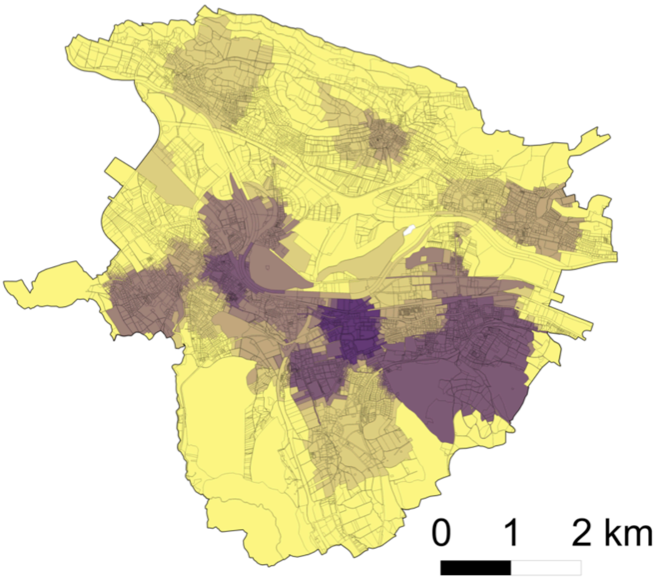
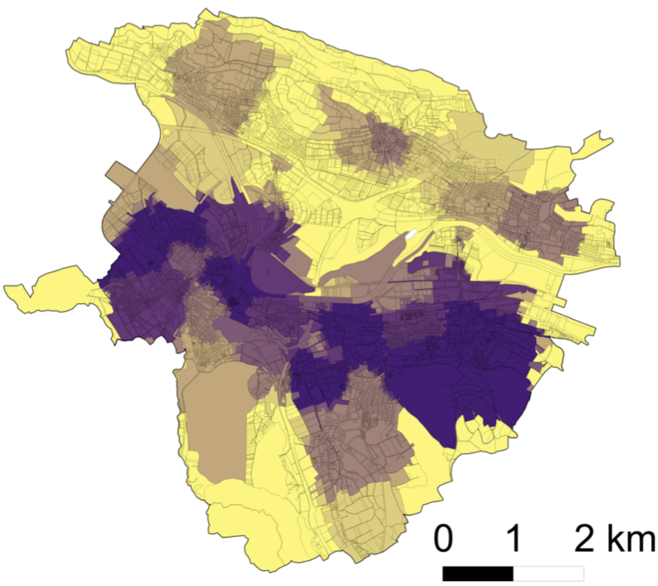


Figure 4.10: Car accessibility



(i) Status quo 2000



(ii) Smart city 2030

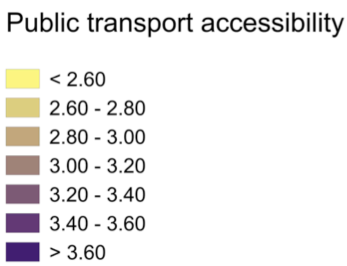
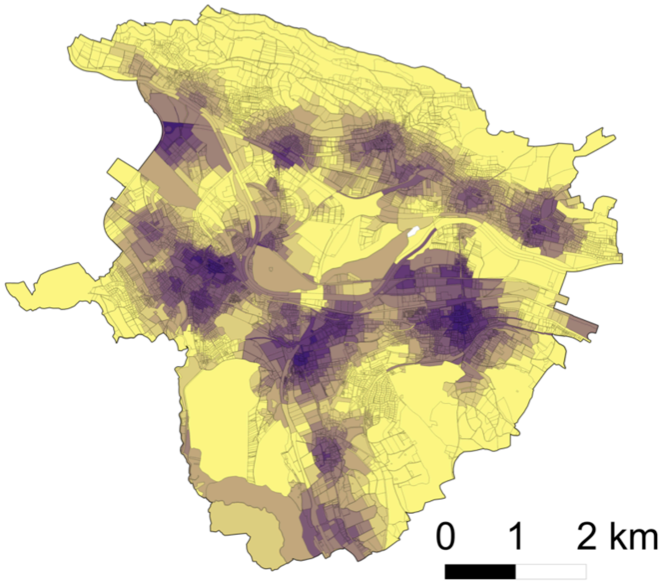
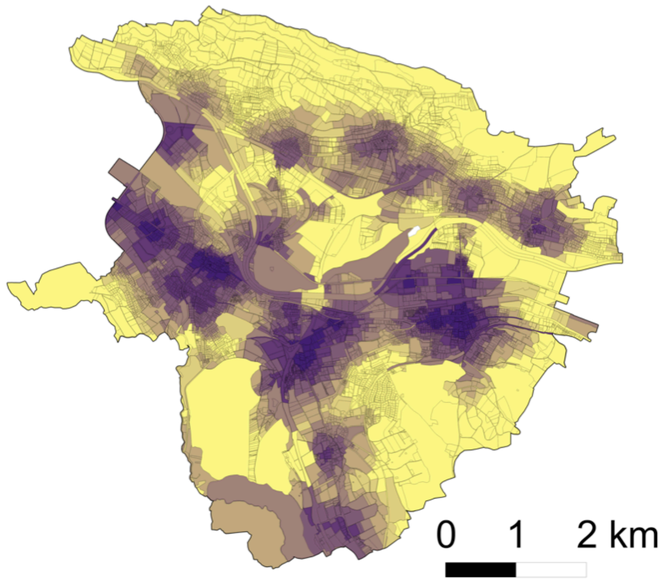


Figure 4.11: Public transport accessibility



(i) Status quo 2000



(ii) Smart city 2030

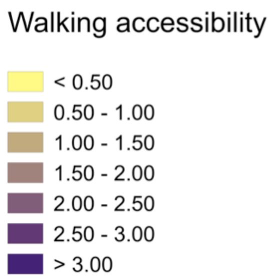
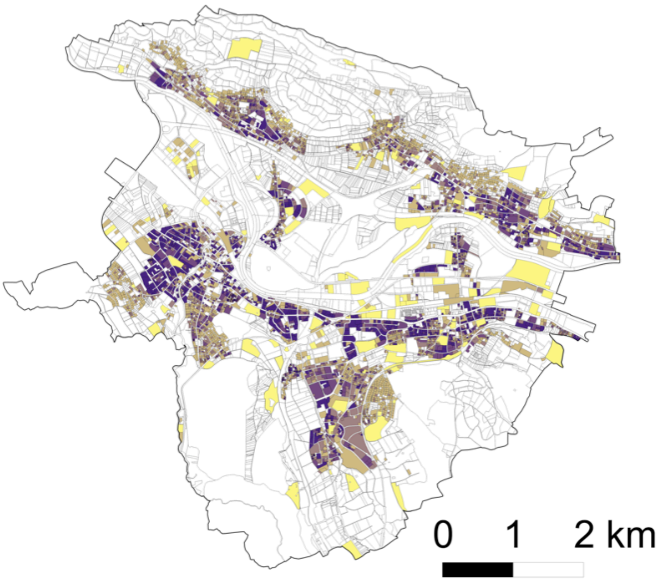
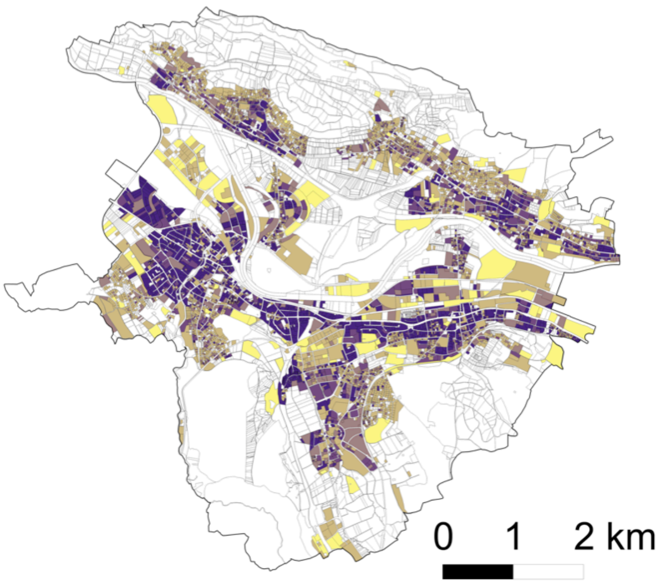


Figure 4.12: Walking accessibility



(i) Status quo 2000



(ii) Smart city 2030

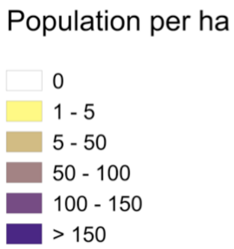
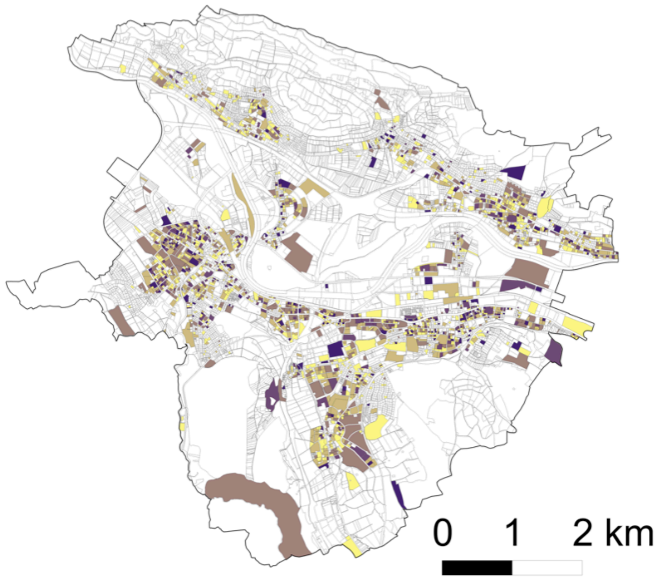
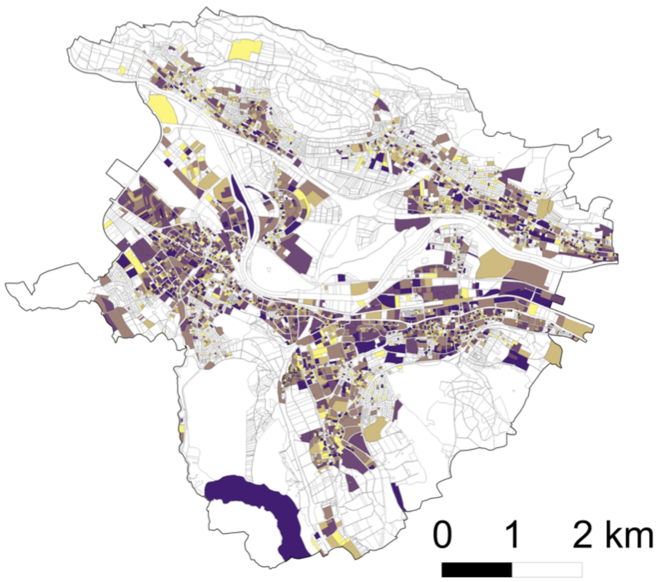


Figure 4.13: Population density (number of people per hectare)



(i) Status quo 2000



(ii) Smart city 2030

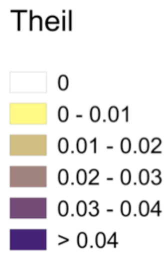
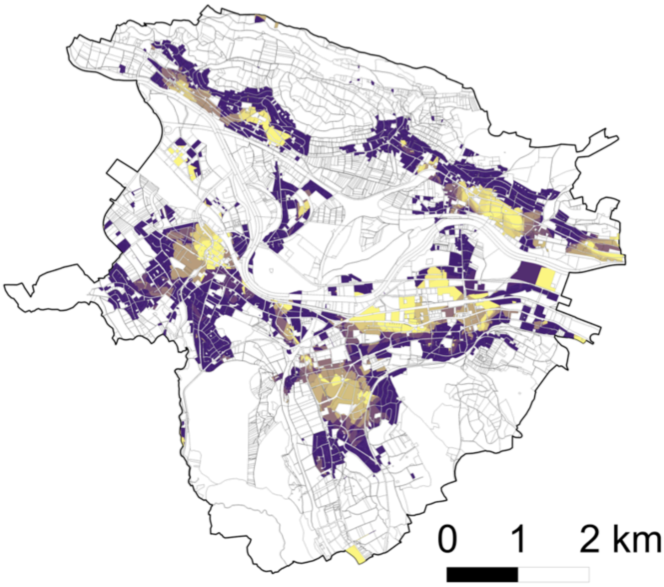
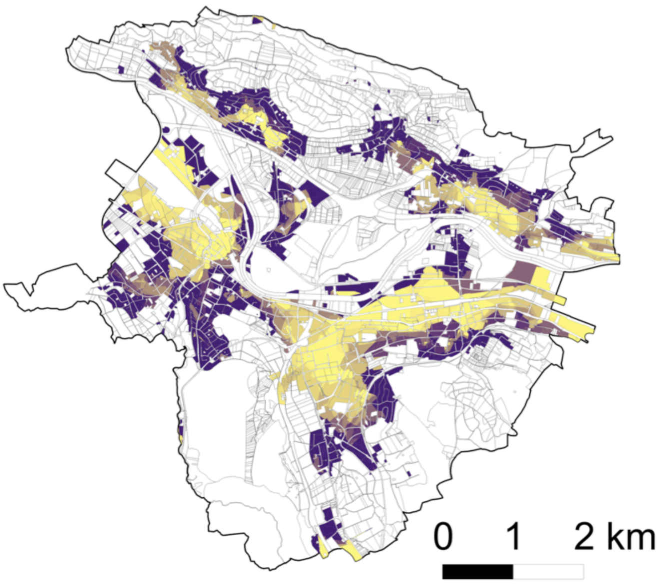


Figure 4.14: Population segregation (Theil index)



(i) Status quo 2000



(ii) Smart city 2030

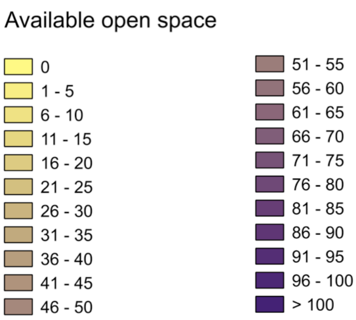
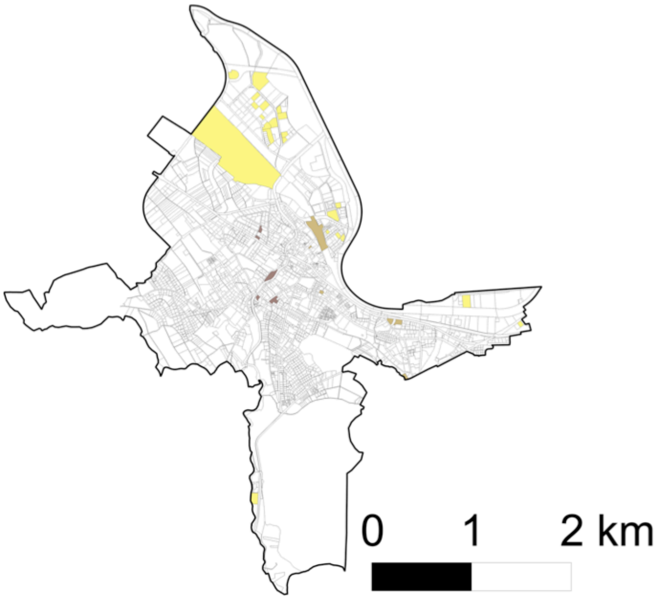
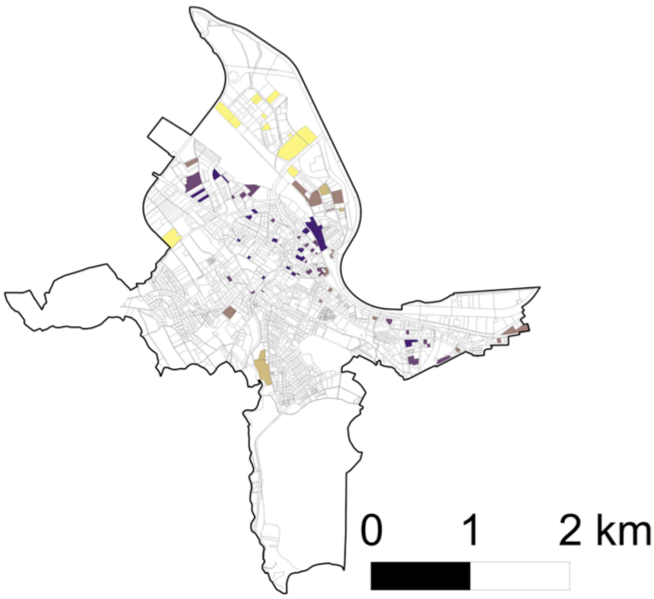


Figure 4.15: Available open space (supply rate of the people with open space [%])



(i) Status quo 2000



(ii) Smart city 2030

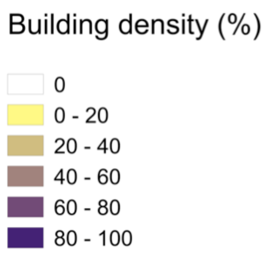
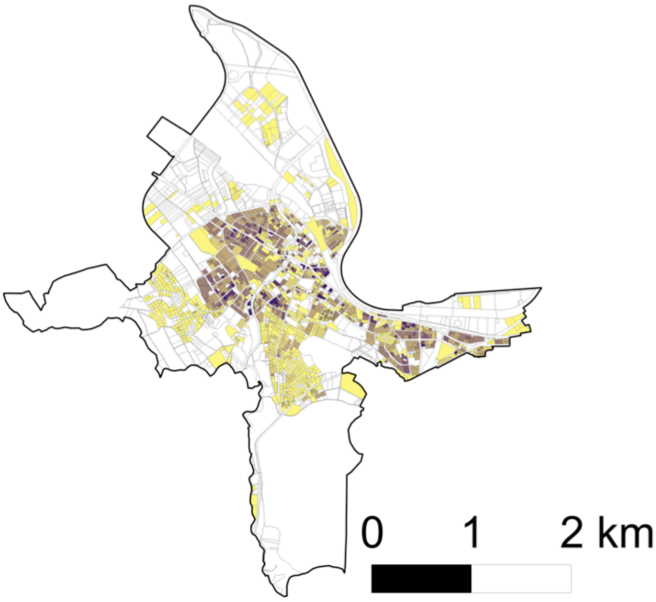
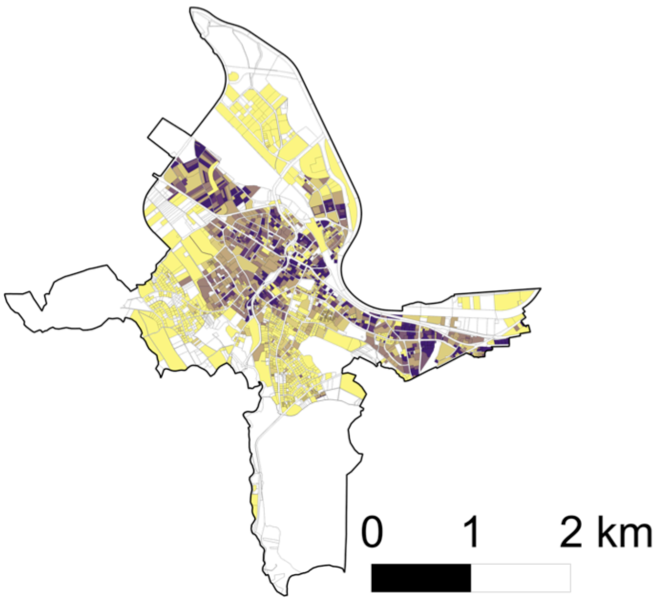


Figure 4.16: Economic viability of retailers (%)



(i) Status quo 2000



(ii) Smart city 2030

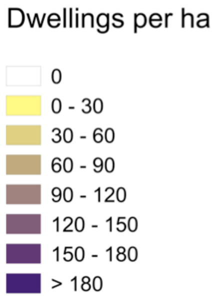
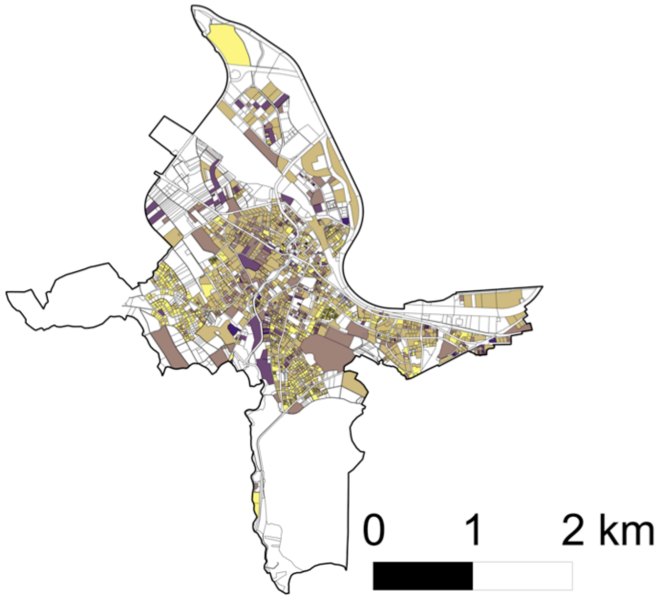
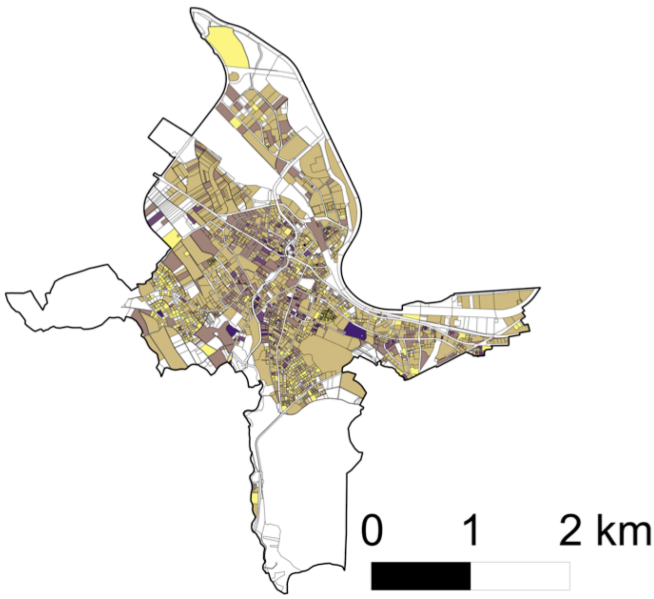


Figure 4.17: Residential density (number of dwellings per hectare)



(i) Status quo 2000



(ii) Smart city 2030

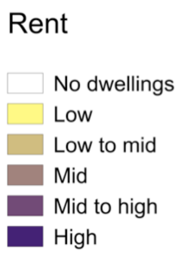


Figure 4.18: Housing affordability (rent per m^2 of residential floor area)

(i) Status quo 2000

(ii) Smart city 2030

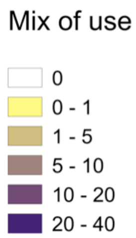
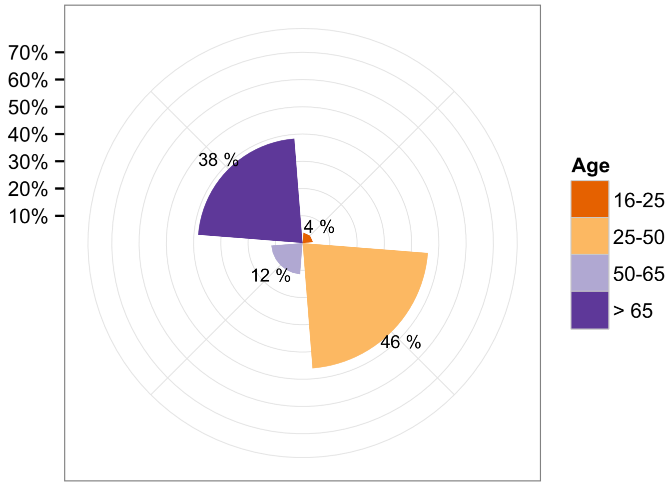
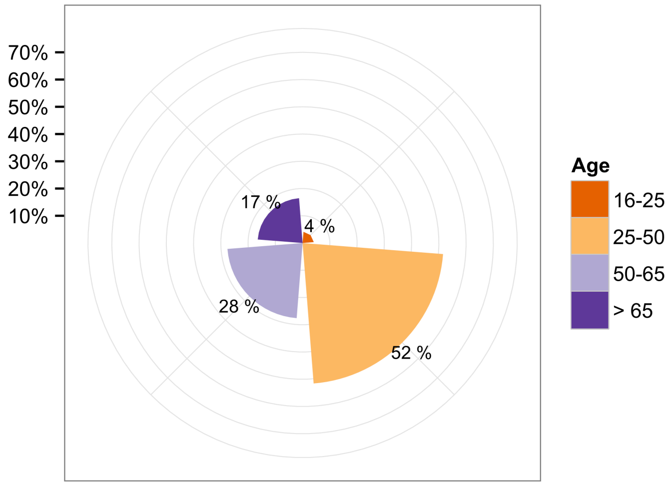


Figure 4.19: Mix of uses (ratio of residential to commercial and industrial floor area [%])



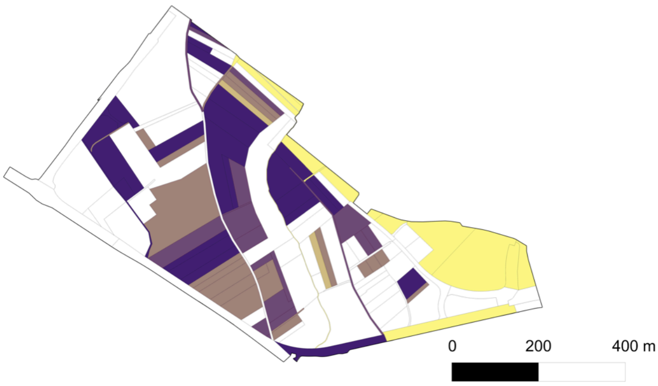
(i) Status quo 2000



(ii) Smart city 2030

Figure 4.20: Social mix (age structure of households)

(i) Status quo 2000



(ii) Smart city 2030

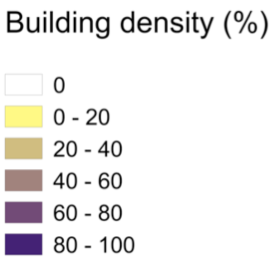


Figure 4.21: Building density (ratio of building footprint to parcel area [%]

4.4.5 Conclusions

The aim of this research to develop an approach that allows planning actors in transdisciplinary collaboration settings for an integrated assessment of interventions providing evidence on their possible long-term effect on different urban quality aspects across scale. It is demonstrated how this can be done taking into account the requirements of the praxis in terms of suitable interfaces and meaningful indicators.

In this research, UrbanSim is implemented integrated with MATSim for simulating a possible urban settlement pattern under the scenario constraints given in the input data. Qualitative information is transferred on an urban development scenario into spatially explicit and quantitative information by altering the land use designation and the allowed floor area ratio on parcel level. This is regarded as a feasible interface between practitioners and the complex urban modeling system. In practice, the preparation of the scenario input data equates to a revision of a zoning plan. Therefore, it is rather likely that the input data can be supplied by practitioners rather easily.

A set of indicators was elaborated that reflects the livability in an urban area. The indicators are calculated on the quantitative data that the simulation system delivers. Thus, they are linking objective, quantifiable aspects of the urban area and the fulfillment of social, economic and environmental needs. This is thought to ensure a high relevance of the indicators for heterogeneous stakeholder groups. It has been illustrated how the maps of the indicators on the different spatial scales can be implemented for gathering evidence on the impacts of the scenario on the urban quality from different perspectives. The two examples of analyzing the indicator maps across scale point up useful insights into focused questions, which can be addressed to the resulting urban patterns. The evidence on the positive effects of the light railway line on accessibility and viability of local retailers as well as the negative effects of certain degrees of densification on recreational quality has high probability to support planning discourses at strategic stages.

The approach that has been developed is generic. This means that it can be easily transferred to other case study areas. However, it has to be pointed out that setting up the agent-based simulation model is rather time intensive and requires a lot of further input data and calibration of the models. But once the baseline model of the current state is set up, scenario data can easily be simulated.

The selection of indicators for the different scales is based on literature review and subjective decisions. Therefore, the selection should be reviewed with regard to the informative value for other case study sites. Other indicators might be necessary to complement the presented set, such as an indicator for the connectivity of habitats for different plant and animal species (Termorshuizen et al., 2007). Ideally the indicators should be selected in a participatory process together with representatives of different stakeholder groups

(Alberti, 1996; Schetke et al., 2012). This could ensure that information is provided that is of real interest for concrete questions in the planning process. More empirical testing of the selection of indicator sets is recommended (Pacionne, 2003).

The approach can provide plausible results on the spatially explicit distribution of the urban development in the region. They can be seen as another means of scenario visualization. More precisely, they are representations of integrated processes at three selected urban landscape scales relevant for the stakeholders. Through the interpretation of relevant indicators, the stakeholders can infer theses of sustainable urban patterns or target values actually required for securing urban quality. Although the models in the behavioral and transport modeling system are rather complex, the values of the indicators are still based on a simplified model of the complex urban system and their validity is limited (Walz et al., 2007). But overall, implementing the approach in concrete trans-disciplinary research and case study settings has high potential to foster a deeper understanding of the coupled human-environment system required for designing sustainable urban patterns. Furthermore, as Albrechts et al. (2003) point out, the analysis of urban patterns across scale can effectively support the development of spatial logic and concrete images that can focus discussions on strategies for sustainable transformation of urban areas.

Wu (2010) states that in order to develop a sustainable urban future natural sciences and design must be fully integrated in a way that we understand how landscapes that we create work and how they can work better. The demonstrated approach in this paper can provide a useful tool for practice to support the understanding of urban development patterns and their impact on urban quality. The indicators chosen can support strategic spatial planning and inform political and administrative efforts in city regions to guide their development processes. However, in order to interpret and enhance urban quality not only physical, objective but also social, subjective indicators are required (Dempsey et al., 2012; Van Kamp et al., 2003). Furthermore, the value system of the assessors defines what urban quality actually is (Pacionne, 2003). Thus, the approach has to be integrated in the collaborative trans-disciplinary processes and complemented with further assessment and visualization tools that explore different components of urban quality.

4.5 Multidimensional indicator analysis

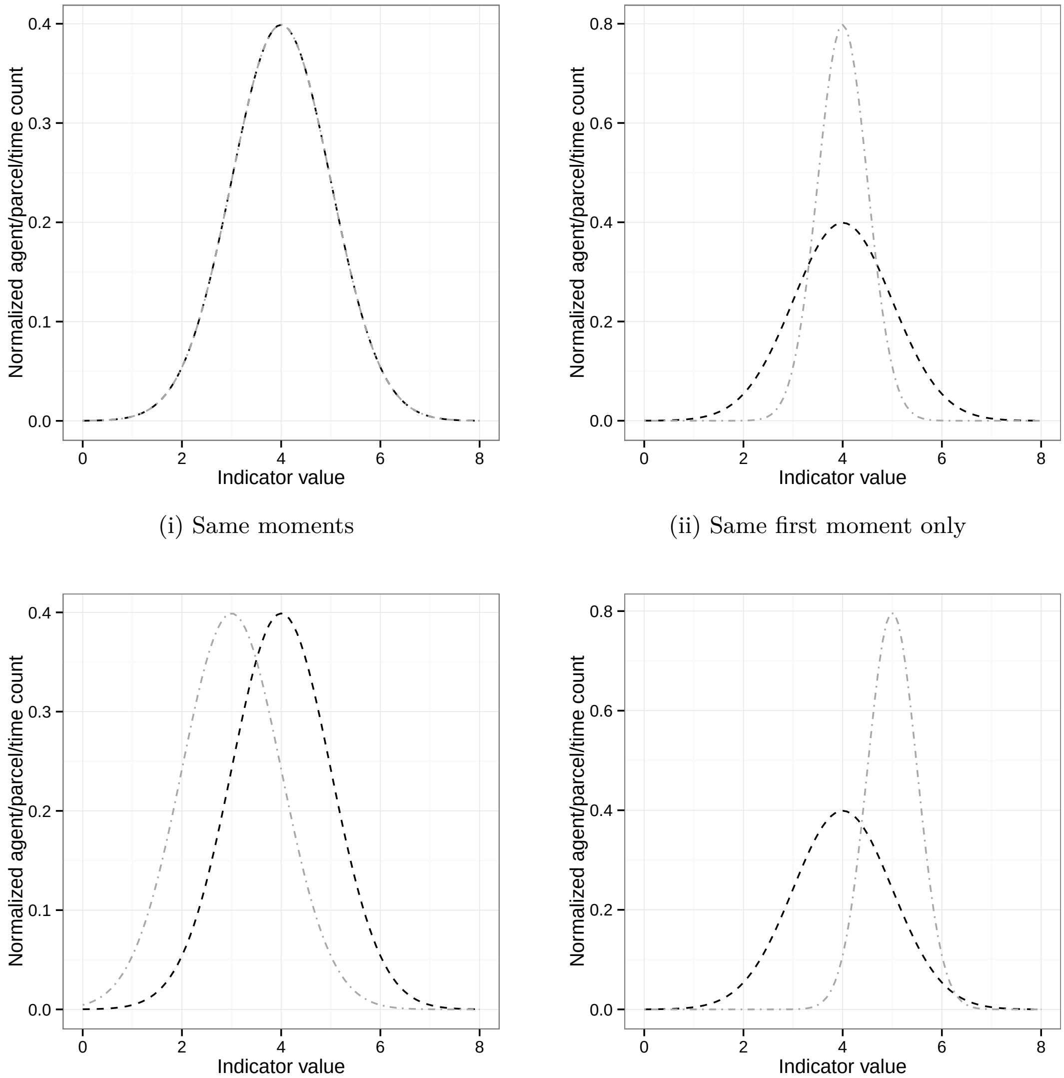
4.5.1 Methodology

In recent years, agent-based models have been established as the state–of–the–art of the LUTI models. In this research, a methodology that takes advantage of the disaggregate information provided at the agent, spatial and time level is proposed to develop the distributions of the policy indicators. The evaluation of different policy scenario is then based on comparison of the distributions and statistical analysis rather than the existing predominant practice of point comparison. Moreover, the methodology takes into account the simulation noise and if required treating it, in the development of indicator distribution. Suppose that we are interested in evaluating the effects of two policy scenario (A and B) on a region using an indicator over individual households (e.g. household income) that is generated from an agent-based LUTI model. The distribution of the indicator from the simulation run can be generated by counting agents/spatial-unit/time-unit and aggregating a histogram. The LUTI simulation is highly random as it involves drawing from various distributions (e.g. migration, birth, job growth rates) and evaluating different models (e.g. location, mode, duration choice models). This necessitates running the simulation several times for each scenario with different seeds; generating the required indicator for each run; and studying the variance within the simulations. If the simulation variance is high, it needs to be controlled using variance reduction techniques (e.g. Antithetic draws, control variates). A detailed discussion on these techniques can be found in Ross (1997). Ševčíková et al. (2007, 2011) proposed a method called Bayesian melding, for assessing the uncertainty about quantities of interest in urban simulation models; the proposed method encodes the information about model inputs and outputs in terms of prior probability distributions and likelihoods, and uses Bayes s theorem so as to obtain the posterior distribution of any quantity of interest.

Once the variance is investigated and corrected for, the distribution of the indicator is the average distribution from various simulation runs. At this point an in depth statistical tests on the indicator distributions from different scenario can be performed. In the literature various tests for normal or non-normal distributions can be found. These tests include Anderson-Darling, Levene's, Kruskul-Wallis, Manning tests and others.

If the first two moments (mean and variance) are to considered, then the comparison between the indicator distribution may result in four cases outlined in Figure 4.22. Case 4.22i represents where the two policies results in exactly the same first two moments. In this case the conventional point-estimates based policy comparison will have similar results as the null-hypothesis testing for the mean and variance of the two distributions. Case 4.22iii is something that will be completely missed using mean/median based policy analysis. Here the means are the same, but in one case the variance is lower. If the indicator under consideration is public transit accessibility then the policy with lower variance is preferred. As it means that at an average a larger population is able to enjoy the same level of public transit accessibility. Case 4.22ii is more clear for the decision maker, as it results in the same variance, but at an average one policy is better than other. Depending on the indicator type, even a better case will be where the variance is smaller for the policy with higher mean. An example of this can be the case where the indicator is accessibility. In Case 4.22iv, the conventional approaches will consider the policy with higher mean, but ignore the higher variance. This may not result in benefiting a large share of population and may only result in high benefits for few outliers. The mean may be higher, but then the variance is higher as well.

Note that this analysis can be conducted in all three dimension: agents, space, and time. Moreover, it can be performed at various aggregations (e.g. parcels, traffic analysis zones, municipalities) and/or cross-sections (e.g. young adults, household of conventional family type).



(i) Same moments

(ii) Same first moment only

(iii) Different first, but same second moment

(iv) Different first and second moments

Figure 4.22: Cases of indicator distributions

4.5.2 Scenarios definition

The Municipality of Limmattal comprises from 15 communities, including a part of the city of Zurich. The transit planing of Limmattal is nationally significant, as it serves the gateway to Zurich. Moreover, it offers a relief function for the agglomeration of Zurich, as it is a popular leisure destination. From an urban development perspective, Limmattal is characterized by high migration rates, not only in the ethnic-cultural mix, but also in the in -and out-commuters. The base year data used in this research, are presented by Zöllig and Axhausen (2011).

For the purpose of this research, two scenarios of development for the region with horizon the year 2030 are examined: A “Pure Dynamics” scenario, where the current trends are going on, and a “Charming Valley” scenario, where the region is characterized by a robust and cross-linked system of green spaces. The Charming Valley scenario assumes that the development of massive high blocks of buildings leads to settlement densification. The flexible open spaces and urban squares allow for temporary uses. The river Limmat is been highlighted and becomes an attraction, while the waterfront area becomes more recreational, and is integrated into a walk and bike network, as well as parks and squares. The main differenc comparing with the Pure Dynamics scenario, is the construction of the light train “Limmattalbahn” in 2020, which runs underground through the region. The private transport is expected to decline, due to higher densification at infrastructure intersections and optimized provision of cross-linked walkways. The reduction of private transport and the set of energy standards on buildings implies high energy efficiency. This scenario assumes that the increase of housing settlement facilitates people of all the social classes to live and work locally. Limmattal becomes very attractive to businesses and employees, because of the increase of the accessibility (see also Section 4.4). The different communities cooperate and are networking in a very conductive way (Wissen et al., 2012).

4.5.3 Results

The first step of the three dimensional analysis (space, time and agents), was to simulate the current situation in the next 30 years, using the Integrated Land-Use and Transport model UrbanSim. In order to generate the distribution of the spatial indicators in time, 10 simulations for each of the scenarios (base-case and Charming Valley) first run, starting from different seed.

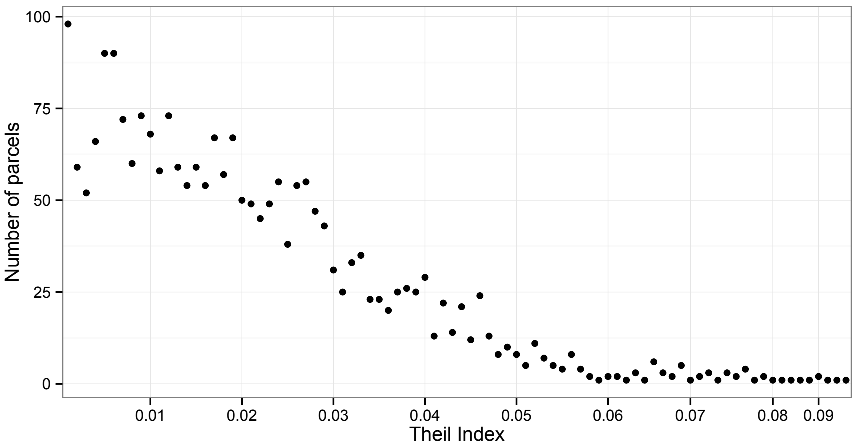
Parcels of zero inequality (i.e. one household) were not-considered for the analysis. Figures 4.23i, 4.23ii and 4.23iii, show the distribution of Theil at the first year of simulation (2001) and in 2030. The number of parcels that a given Theil decomposition measurement appears, differs significantly from one simulation to another in both scenarios. This verifies the expectations that the results should not be based in a single simulation, but many. Therefore, the mean number of parcels per Theil measurement should computed for the rest of the analysis.

Following the static inter-simulation analysis, the distributions between the years of the simulation are compared. Having observed that the distribution is non-linear, a non-parametric test Mann and Whitney (1947) has been applied to compare the distributions of the mean Theil measurements per year with 2001. The results of the test show that the null-hypothesis that the data are coming from the same population is rejected for 99% confidence level. The difference between the mean values of the distributions and the confidence intervals computed by the Mann-Whitney test are presented in Figure 4.24i and 4.24ii. It is observed that the mean increases by about 0.001 (±0.001) per 5 years. In 2020 and 2030 the mean remains the same, probably because Theil has been previously rounded at the third digit. Having used the Equation 2.21 (Section 2.5), the Theil decomposition of the whole area has been found to be 0.031 in both scenarios, with insignificant variance between the simulations.

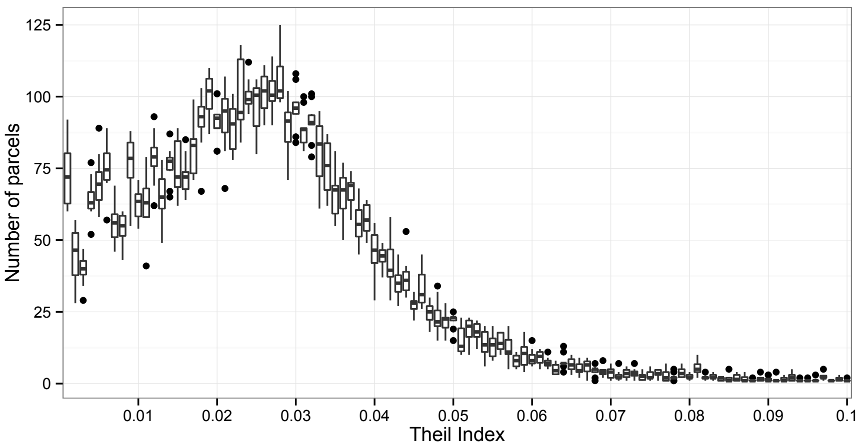
Apart form the Theil measurement, the distributions of the car and public transport accessibilities have been generated. Figures 4.25i and 4.25ii show the spatial distribution of these indicators at the base-year (2000), and 4.25iii, 4.25iv, 4.25v and 4.25vi, show the distributions at the year 2030 for the Pure Dynamics and the Charming Valley scenarios. The inter-simulation variance of the public transport accessibility becomes very high after 30 years of simulation, while the new light-rail line considered by the Charming Valley scenario seems to have an effect on that, since the variance in higher there. This indicates that, similar to Theil's case, the mean values should be computed. On the other hand, the car accessibility does not vary much between the simulations.

Figures 4.26i and 4.26ii show the total car and public transport accessibilities of the area per year. Accessibility indicates the Value of Travel Time Savings (VTTS) for trips to work Nagel (2013). Both accessibilities increase in time. An interesting finding of the analysis, is the sudden increase of the public transport accessibility after the year of implementing the new metro line, which is a side-effect of the decrease of the public transport travel time. In 2020, the public transport accessibility gets equal increase as the last 20 years all together.

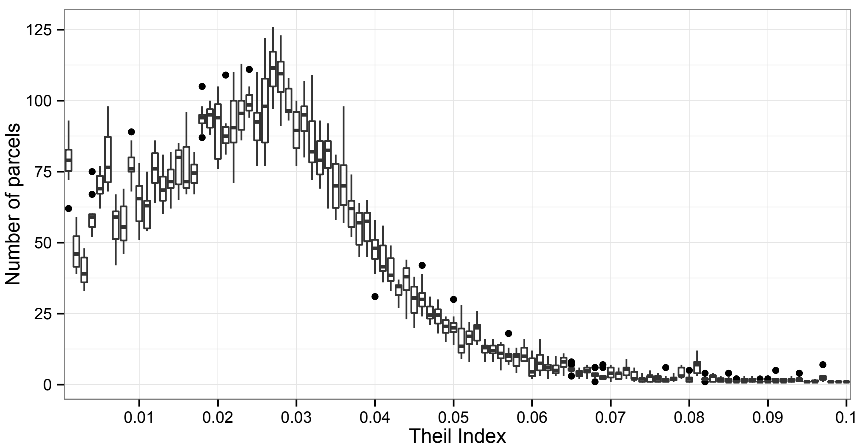
It has been observed that the accessibility distributions can be approximated by Weibull. The scale parameter of the Weibull distribution (which shows the statistical dispersion) of the car and public transport accessibilities are depicted in Figures 4.27i, 4.27ii, 4.27iii and 4.27iv. A loess curve has been fitted to make the result more visible. The effect of the new metro line is made clear at the Scale parameter of the beta distribution in the Charming Valley scenario (Figure 4.27iv). The parameters for both scenarios increase in time. This raise some questions about the trust that one should have to the results of long-term simulated scenarios.



(i) Year 2001

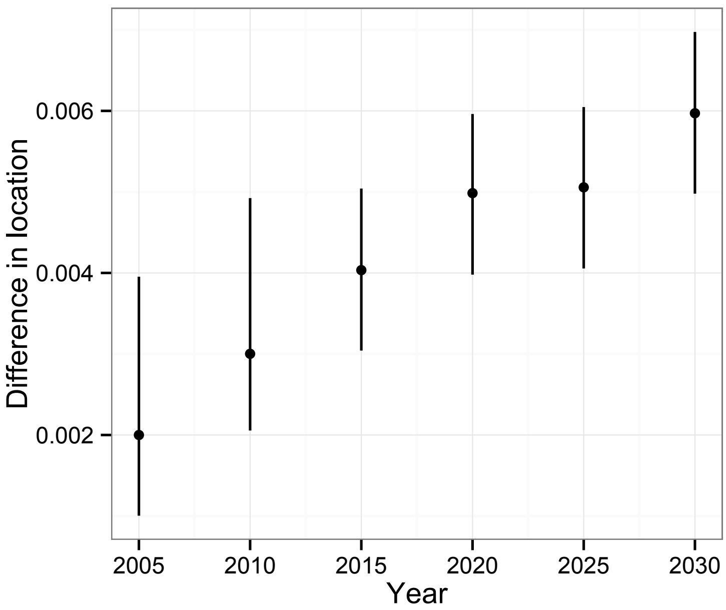


(ii) Year 2030 - Pure Dynamics scenario

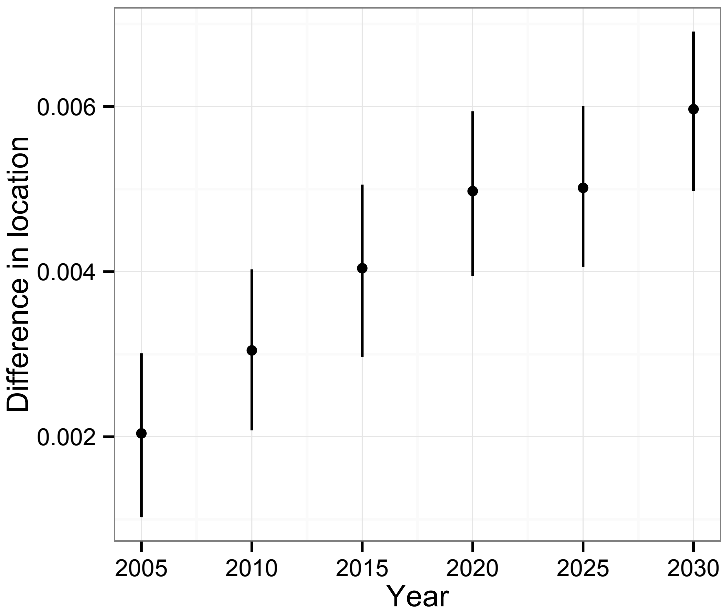


(iii) Year 2030 - Charming Valley scenario

Figure 4.23: Spatial distribution of Theil index

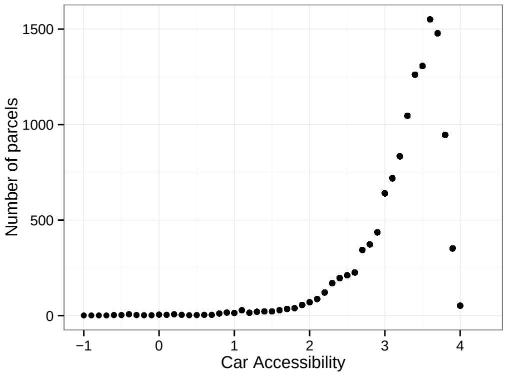


(i) Pure Dynamics scenario

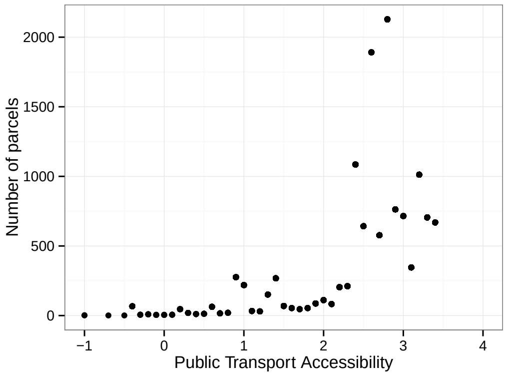


(ii) Charming Valley scenario

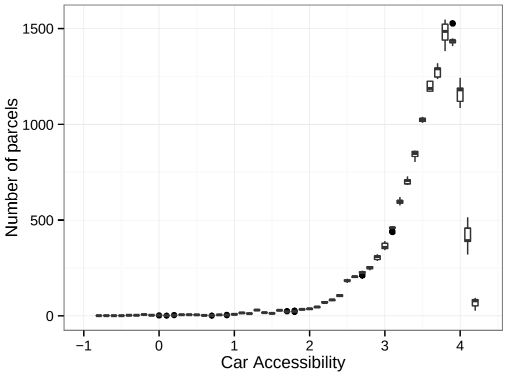
Figure 4.24: Mann-Whitney test



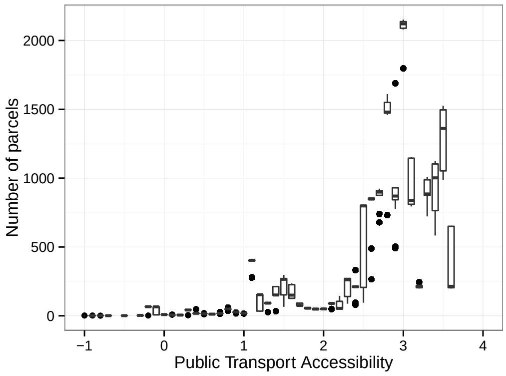
(i) Car accessibility in 2001



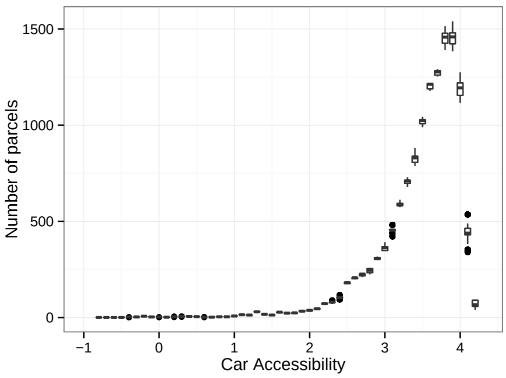
(ii) Public transport accessibility in 2001



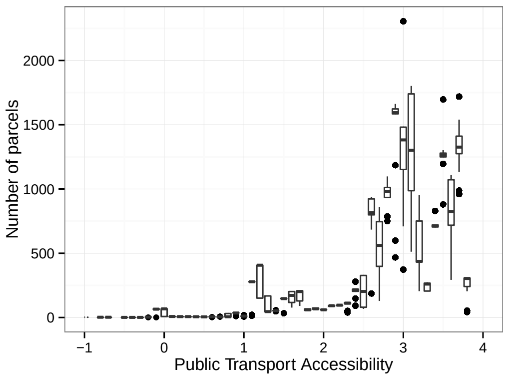
(iii) Car accessibility in 2030 - PD



(iv) PT accessibility in 2030 - PD

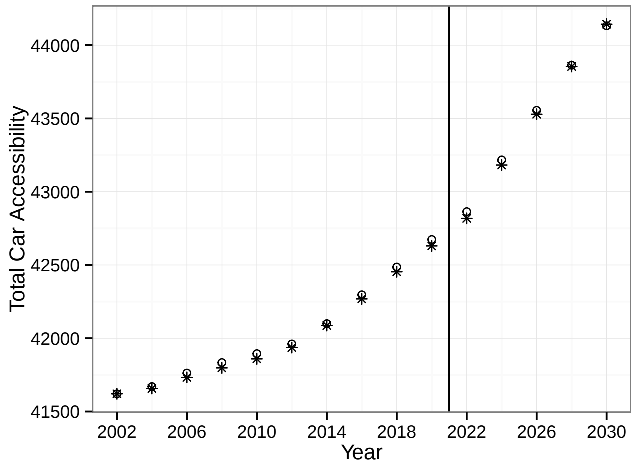


(v) Car accessibility in 2030 - CV

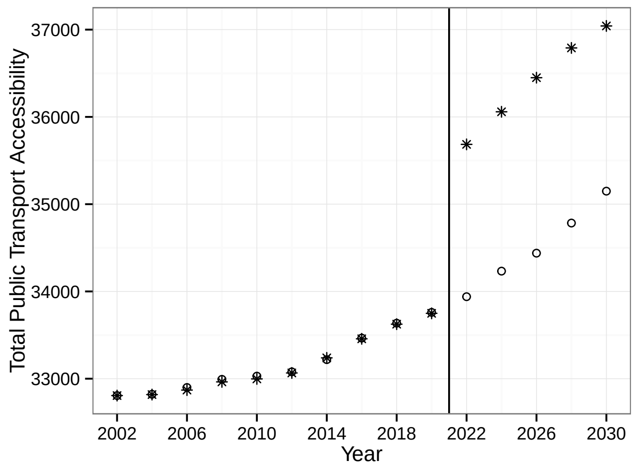


(vi) PT accessibility in 2030 - CV

Figure 4.25: Spatial distribution of accessibilities

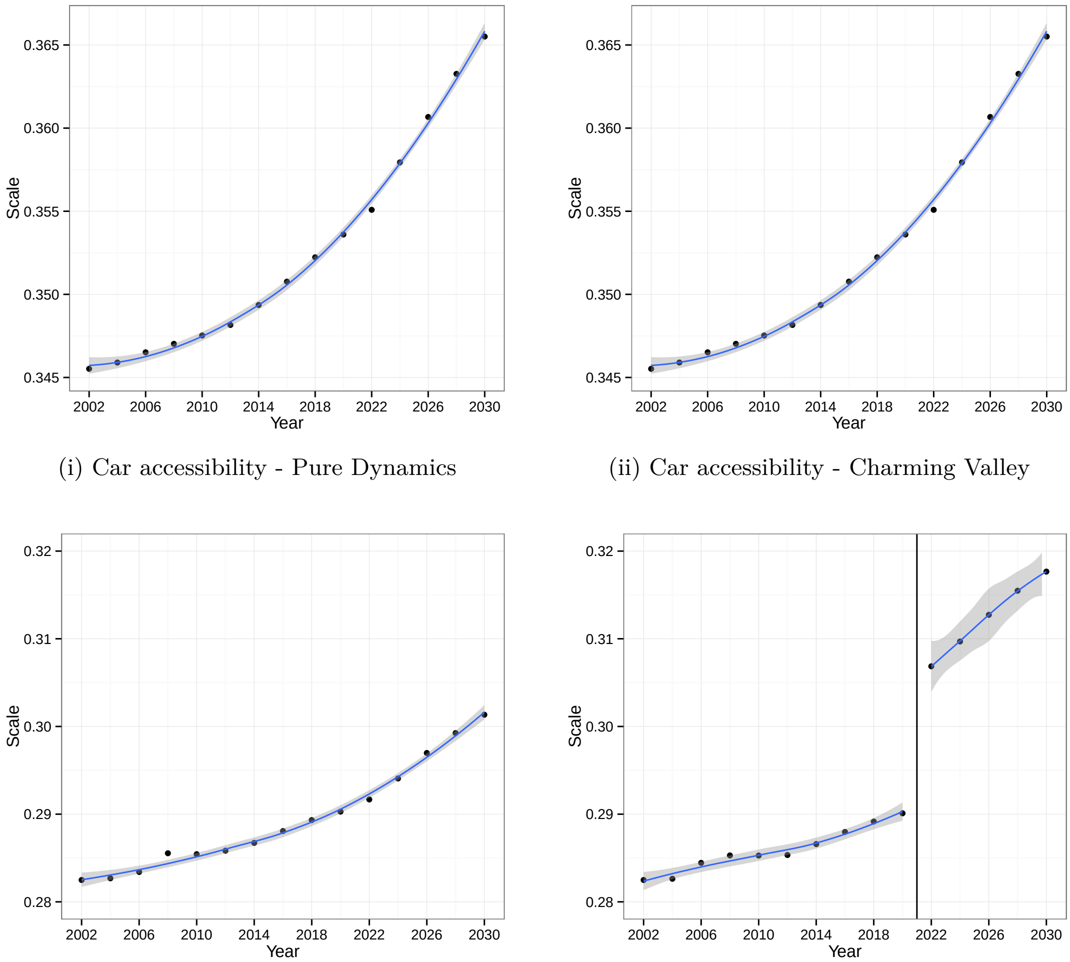


(i) Car accessibility in 2030; circles: Pure Dynamics scenario, stars: Charming Valley scenario



(ii) Public transport accessibility in 2030; circles: Pure Dynamics scenario, stars: Charming Valley scenario

Figure 4.26: Total accessibility



(i) Car accessibility - Pure Dynamics

(ii) Car accessibility - Charming Valley

(iii) Pt accessibility - Pure Dynamics

(iv) Pt accessibility - Charming Valley

Figure 4.27: Scale parameter of Weibull distribution

Inter-simulation analysis

Having completing the inter-simulation distribution-analysis of the indicators, we are attempting to analyze the variance between the Pure Dynamics and the Charming Valley scenarios, comparing their spatial distributions. Since the indicators follow non-normal distributions, we employed non-parametric tests for their comparison. These tests, which examine whether the samples originate from the same distribution, are the:

Kruskal Wallis test (null hypothesis: population variances are equal) (Kruskal and Wallis, 1952)

$$K = (N-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_i - \bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij} - r)^2} \tag{4.11}$$

where n_i is the number of observations per group i, r_{ij} is the rank of j from group i, N is the total number of observations, $\bar{r}_i = \frac{\sum_{j=1}^{n_i}}{n_i}$ and $\bar{r}_i = 1/2(N+1)$ is the average

Anderson - Darling test (null hypothesis: all samples come from a common population) (Anderson and Darling, 1954)

$$A^2 = -n - S \tag{4.12}$$

where A is the test statistic, $S = \sum_{k=1}^{n} \frac{2k-1}{n}[ln(F(Y_k)) + ln(1 - F(Y_{n+1-k}))]$, where F is the cumulative distribution.

Levene's test (null hypothesis: population variances are equal) (Brown and Forsythe, 1974)

$$W = \frac{(N-1)}{(k-1)}\frac{\sum_{i=1}^{k} n_i(\bar{z_i}. - \bar{z}..)^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(z_{ij} - \bar{z_i}.)^2} \tag{4.13}$$

where $\bar{z_i} = \sum z_{ij}/n_i$ and $\bar{z_i}. = \sum\sum z_{ij}/\sum_{n_i}$

Mann Whitney test (Null hypothesis: true location shift of means is equal to 0) (Mann and Whitney, 1947)

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2} \tag{4.14}$$

where $n1$ is the sample size for sample1, R_1 is the sum of the ranks in sample 1.

Sign test (Null hypothesis: true median difference equal to 0) (Mendenhall et al., 1989), which assumes that is the null hypothesis is true then the W (the test statistic, number of pairs for $y_i - x_1 > 0$ y where y and x the observation for each population respectively) follows a binomial distribution.

The tests have been applied on the subsets of parcels of different Euclidean distances (<100, 250, 500, 1000 meters and the whole area) from the light rail stations which is planned to be constructed in 2020. In all tests and for every distance from the stations, the results show that the distributions of the car accessibility are same. On the other hand, the null hypothesis that the public transport accessibility distributions are similar is rejected. The Mann-Whitney test measures the location shift of the distribution from 0.20, for the parcels up to 100 meters form the stations, to 0.15 for those 1000 meters around the stations. Moreover, the mean of the Theil index distribution in the public transport investment scenario is measured to be -0.001.

4.5.4 Conclusions

The evaluation of transport and land use policies is based on various sustainability, econometric, inequality and accessibility indicators, which are computed at an aggregate spatial level (usually city of country). Although the effects of the potential policies are usually simulated using complex Integrated Land-Use and Transport Models, the strength of microsimulation is not exploited, and the evaluation is based on single measurements. The aim of this research is to exploit the strength of microsimulation in three dimensions, space, time and agents, and analyze the variances of policy evaluation indicators within them. For the purpose of this research, the Land-Use and Transport Model UrbanSim integrated with the agent-based traffic simulation model MATSim is applied on Limmattal area, in Switzerland.

The analysis is performed in three steps. First of all, the socioeconomic characteristics of the area are forecasted for the years between 2001 and 2030 via two policy scenarios: a “Pure Dynamics” scenario, where the current trends are going on, and a “Charming Valley” scenario, which assumes the implementation of a new metro line in the year 2020. Using as input the simulated data that have been generated by 10 runs starting by different seed, the spatial distributions of the Theil inequality measurement, and the car and public transport accessibilities are generated. The results show that the variance between the different simulations is high and therefore it is essential that the mean values to be computed. The non-parametric Mann-Whitney test has been applied to compare the Theil distributions between the simulation years, concluding that the mean increases by 0.001 every five years at 99% confidence level. Having observed that the accessibility distributions can be approximated by a Weibull distribution, it is investigated the trend of the scale and shape parameters in time. The results show that the implementation of the new light rail line in the year 2020 has a direct effect on the scale parameter of the Weibull distribution in the Charming Valley scenario.

Finally, the location shift of the distributions between the two scenarios has been investigated. In order to do that, multiple non-parametric tests (because the distributions are non-normal) have been applied to the year 2030, such as the Kruskal-Wallis, Anderson-

Table 4.9: Inter-scenario analysis of variance

	<=100				<=250				<=500				<=1000				Whole area			
	test	p	CI	Dif.	test	p	CI	Dif.	test	p	CI	Dif.	test	p	CI	Dif.	test	p	CI	Dif.
THEIL																				
Anderson - Darling	0.19	0.66			3.90	0.01			3.91	0.01			3.60	0.01			7.69	0.00		
Levene's (ANOVA)	0.38	0.54			1.61	0.21			0.18	0.68			0.09	0.76			8.06	0.00		
Levene's (Kruskal - Wallis)	1.55	0.07			2.82	0.09			3.29	0.07			0.02	0.89			101.40	0.00		
Mann Whitney	11194	0.08	99%	-0.002	367157	0.02	99%	-0.001	1709528	0.02	99%	-0.001	4085536.00	0.02	99%	-0.001	15356218	0.00	99%	-0.001
Sign - test	52	0.00	99%	-0.001	309	0.00	99%	-0.001	723	0.02	99%	0.000	1022.00	0.00	99%	0.000	1690	0.00	99%	0.000
Car accessibility																				
Anderson - Darling	-1.18	0.68			-0.62	0.51			-0.57	0.50			-0.84	0.58			-1.06	0.64		
Levene's (ANOVA)	0.04	0.85			0.18	0.67			0.11	0.74			0.00	0.99			0.02	0.89		
Levene's (Kruskal - Wallis)	0.65	0.42			2.70	0.10			2.18	0.14			1.51	0.22			4.05	0.04		
Mann Whitney	43858	0.98	99%	0.00	1348866	0.79	99%	0.00	6509832	0.80	99%	0.00	17317121	0.00	99%	0.00	71756692	0.93	99%	0.00
Sign - test	86	0.48	99%	0.00	557	0.00	99%	0.00	1190	0.00	99%	0.00	1743	0.00	99%	0.00	2850	0.00	99%	0.00
PT accessibility																				
Anderson - Darling	37.24	0.00			228.02	0.00			420.08	0.00			316.29	0.00			158.46	0.00		
Levene's (ANOVA)	22.44	0.00			65.01	0.00			110.87	0.00			2.21	0.14			8.10	0.00		
Levene's (Kruskal - Wallis)	26.05	0.00			191.34	0.00			330.77	0.00			32.06	0.00			27.36	0.00		
Mann Whitney	57549	0.00	99%	0.20	1809191	0.00	99%	0.26	8489829	0.00	99%	0.20	20563444	0.00	99%	0.15	72838447	0.04	99%	0.02
Sign - test	234	0.00	99%	0.20	1254	0.00	99%	0.20	2484	0.00	99%	0.15	2975	0.00	99%	0.01	3596	0.00	99%	-0.03

Darling, Levene's, Mann-Whitney and Sign-test assuming different distances from the new light rail stations. These tests examine the null hypothesis that the observations of the two samples (scenarios) belong to the same population (their variance is 0). The results show the new light rail line has a direct effect on the Theil index, reducing its mean by 0.001 comparing to the Pure Dynamics scenario, when the whole area is examined. Concerning the public transport accessibility, the Mann-Whitney test shows, that the mean of the distribution increases by 0.20 to 0.15, 100 and 1000m around the new rail stations respectively. The car accessibility distribution remains the same. In future, the authors intend to extend this analysis in more indicators and scenarios.

By exploiting the strengths of microsimulation, this research offers a more complete policy evaluation methodology, contributing to the literature and creating new perspectives for future research.

4.6 Discussion

Practical additions to the current LUTI model framework, aiming to render them more flexible policy evaluation tools, have been proposed. A graph-theoretic solution for associations generation in synthetic population has been demonstrated. An UrbanSim & MATSim case study is them presented, detailing all stages of the development from estimation to validation.

Policy evaluation indicators are integrated with LUTI models for two-stage scenario analysis. First, the resulted from transport and land-use policies urban quality is measured, and second, indicators' distributions generated from multiple simulations are compared in three dimensions (agents, space and time). Multidimensional distribution analysis sheds more light to the side–effects of policy implementation, addressing equity considerations that are usually missed when single aggregate measure are used.

5 Sustainable transport policies in times of crisis

5.1 Research objective and chapter structure

This is a policy planning oriented Chapter. The aim is to investigate the use of alternative sustainable public transit policies in times of crisis, using advanced behavioral modeling methodologies such as hybrid choice and latent variable models. Section 5.2 analyzes the willingness of people to join potential car– and bike–sharing systems in Greece, and Section 5.3 describes the impact of recession on user satisfaction about public transit quality and demand.

5.2 Vehicle sharing systems: factors that affect the willingness to join

5.2.1 Introduction

Urban transport habits have been redefined during the last decade. The cost of purchasing and maintaining a car, the always increasing fuel prices and the restricted available parking space in urban areas, drive people to look for alternative ways of travel (Millard-Ball et al., 2005; Shaheen and Cohen, 2007). When car ownership becomes a luxury and public transport restricts the freedom and quality of travel, car and bike-sharing schemes seem to be the middle solution. The pressure on World Governments to reduce greenhouse gases (GHG) (Walsh, 1990; European Environment Agency, 2010) in order to fulfill their obligation toward the environmental protocols, a target that becomes more and more difficult to reach because of the increasing emissions by transport, in turn forces them to encourage the investment on alternative, sustainable, urban transport schemes, such as conventional or electrified carsharing and conventional or electrified bikesharing (Millard-Ball et al., 2005; Shaheen and Cohen, 2007; Barth and Shaheen, 2002; Shaheen et al., 2010a,b,c). These new ideas are emerging either supported by government funding, or promoted by ambitious entrepreneurs that exploit these new opportunities. Relevant initiatives like Getaround (www.getaround.com) offer peer-to-peer solutions through websites and smartphone applications. In these systems, car-owners can rent their vehicles for the time they do not use it.

The objective of this research is to provide some insight into the factors affecting the shared-vehicle systems' adoption. These factors can be related on the (a) system: e.g. the distance between the stations, the way of reservation, the type of vehicles, the restrictions of usage, and (b) user: e.g. age, income, current mode used for trips to different destinations, environmental consciousness. The survey was performed in Greece, where bikesharing is just emerging and carsharing is effectively unknown. An ordered logit model is used for the satisfaction of the respondents about their travel patterns, and factor analysis is performed on the advantages and disadvantages of car and bike-ownership, so as to reveal any latent correlation between the different variables. The factors affecting the adoption of carsharing and bikesharing schemes are analyzed descriptively. Ordered logit, and hybrid choice and latent variable models are developed to model the willingness of young Greeks to join these schemes.

5.2.2 Questionnaire design and survey results

The electronic questionnaire was implemented using the Google Forms tool and disseminated on-line. The objective was to examine the perception and attitude of people in Greece, toward car and bike-sharing services, as well as develop some insight into the influencing. Considering the survey involved vehicle sharing schemes that are not available in Greece, it was necessary to structure the questionnaire as a stated preference survey. Compared to revealed preference surveys, stated preferences surveys offer a practical way to collect data regarding potentially unavailable alternatives; on the other hand, since the information is not directly observed, they are susceptible to various sorts of response biases, such as anchoring bias (McFadden, 2001), inertia bias (Thaler and Sunstein, 2009), hypothetical bias (Murphy et al., 2005; Loomis, 2011), aggregation bias (Morrison, 2000). This research is also susceptible to this sort of biases and this should be considered when assessing the results presented. However, considering that such vehicle-sharing systems are not yet widely available in Greece, revealed preferences studies are not possible.

According to Dillman et al. (2009), compared with telephone surveys, internet surveys are advantageous because of the lower non-response rate. However, they are vulnerable to sampling error. They address a specific segment of the population (internet users), who usually are college students or certain professionals; 29% of people who live in USA do not have access to the Internet, while less educated people with lower income, people more than 65 years old and non-whites have less access. Dillman et al. (2009) also suggests that the sampling, non-response, coverage and measurement errors should all be considered in order to achieve better results from a survey. The implication of the uneven access to the questionnaire due to different socio-economic factors could lead to biased survey results; however, steps to reduce these limitations can be taken. In this study, the sampling and coverage errors were limited in two ways. First, by focusing in the target group that is more likely to join car/bike-sharing schemes (ages 18–35). Second, the non-response error was reduced by employing social media and by re-disseminating the questionnaire in the same group during the survey period. However, it should be acknowledged that measurement/satisfaction biases may still remain, since the respondents might overstate their willingness to join a proposed “green” scheme.

The survey is structured in four parts. The first part includes questions about the respondents’ travel patterns, such as modes used for daily trips, perception about advantages and disadvantages of car and bike ownership (costs, parking problems, pollution etc.), satisfaction about current travel patterns and familiarity with car and bike sharing. The second part provides a description of car and bike sharing systems and asks a number of questions about them. The third part includes questions about the perception of the importance of factors for bikesharing and carsharing adoption, while the questionnaire ends with questions about demographics.

The Google Forms offering was selected to develop and host our survey, because it provides the facilities for various types of questions and also offers a back-end that tabulates the responses into a spreadsheet. Furthermore, summary statistics of the results are also presented. The form was disseminated through various channels, including social media, such as facebook, and electronic mailing lists (such as University Transport Study Group (UTSG, www.utsg.net) and Travel Model Improvement Program (TMIP, tmip.fhwa.dot.gov), and three student lists at Imperial College and NTUA), as well as personal contacts of the authors. A message/e-mail was sent as reminder to the participants after every set of 100 responses that was received. The key demographic results (summarized in Table 5.1) include a balanced gender split, but – as expected for an internet-based survey – some statistics are not fully representative of the entire population. To demonstrate the areas in which the sample reflects correctly the population and the points where it could be improved, data from the last available full Census in Greece (2001; the final data from the 2011 census have not been fully processed and made available yet) are also presented in Table 5.1. Gender representation, employment status and household size are rather well represented in the sample, while marital status and education level are less representative. The sample includes more single respondents and a higher education level completed, which could be explained by the fact that these segments of the population may be more likely to respond to on-line questionnaires. Having said that, considering that people in Greece usually purchase their first car at the ages between 18 and 35 years old, but also considering the findings of Millard-Ball et al. (2005) that car-sharing members are usually between 25 and 35, the travel patterns of this younger age group and their willingness of joining carsharing or bikesharing are of great interest for further analysis.

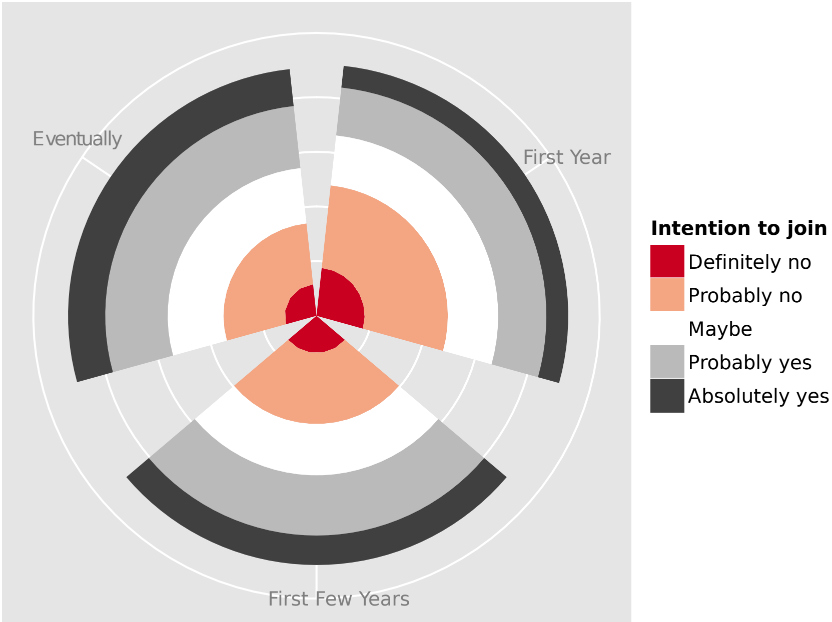
Table 5.1: Key demographic characteristics of age group 18–35 years old (N=233)

		Survey (ages 18–35)	**2001 Census (ages 20–35)**
Age	18–25	51.1%	7.6%
	26–35	48.9%	15.7%
Gender	Male	49.4%	51.5%
	Female	50.6%	48.5%
Marital status	Single	96.57%	67.0%
	Married	3.43%	29.6%
Employment status	Full-time	43.35%	61.6%
	Part-time	12.02%	-
	Student	33.05%	10.2%
	Unemployment-looking for a job	10.3%	4.5%
	Unemployment–not looking for a job	0.86%	-
	Homemaker	0.43%	12.7%
	Other	–	11.0%
Education level completed	High school	4.3%	45.4%
	College	2.2%	14.5%
	Higher education–University	48.5%	11.8%
	Masters	37.3%	1.1%
	Doctorate	7.7%	0.2%
	Other(less than high school)	–	26.8%
		Survey (ages 18–35)	2001 Census (all ages)
Household size	1	29.6%	7.0%
	2	17.6%	20.0%
	3	17.2%	22.5%
	4	28.3%	29.3%
	5 or more	7.3%	21.0%
Household income	<10	19.7%	N/A
	10–15	17.2%	N/A
	15–25	21.0%	N/A
	25–50	24.0%	N/A
	50–75	7.7%	N/A
	75–100	4.3%	N/A
	>100	6.0%	N/A

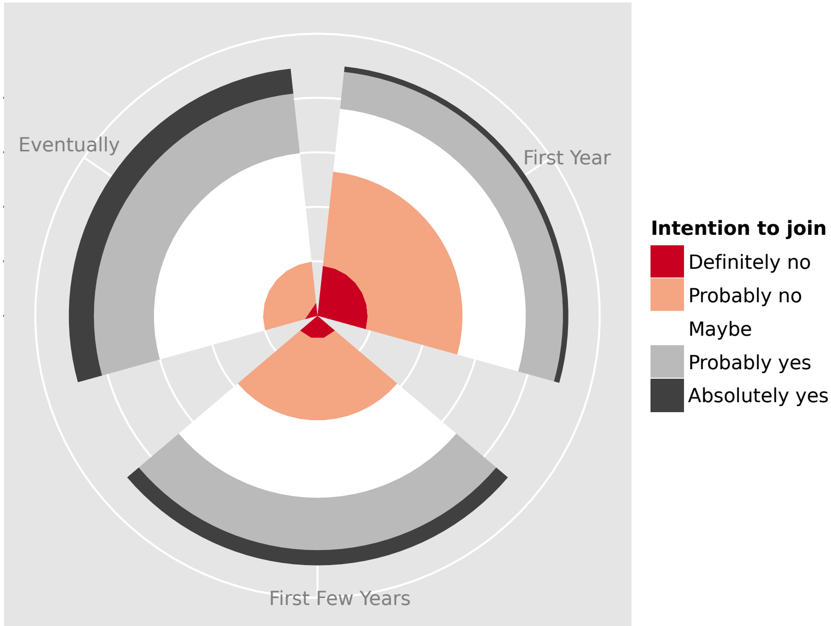
Notes: 1: percent of sample, 2: percent of full population, and N/A: not available

More than half (52.4%) of young Greek respondents indicated that they are not familiar with carsharing –the respective percentages reported by Ohta et al. (2009) and Shaheen and Martin (2007) in studies in Japan and China respectively were 40%. The corresponding percent for bikesharing rises to 57.5%. The respondents were asked if they would join the schemes (in a scale of “1–Definitely no”, to “5–Absolutely yes”) and (if yes) in which time horizon. The reference for the response levels was based on the time of initial implementation and operation of these schemes in Greece. Therefore, three levels have been selected: (i) “in the first year”, (ii) ‘in the first few years’ (up to 5 years) and (iii) “eventually (5 to 10 years)”. About 30% of the respondents answered that they will join carsharing; 15.9% will probably or definitely join the scheme in the 1st year, 8.6% in the first few years and 5.6% eventually. Concerning the bikesharing system, 42.1% indicated that they will join; 26.3% within the 1st year, 8.6% in the first few years and 5.2% eventually. Figure 5.1 shows the percentage of the responses for the three different time-horizons.

The majority (90.6%) of the respondents has a car driving license and almost half (47.6%) own a car, while many others use a family or company car. It is disappointing, considering the consequences on the environment caused by single-passenger car’s high carbon footprint, that more than half (51%) of them drive alone to work or school. About 38.3% of the subgroup that has planned to buy a car in the next 6 months answered that they will probably or definitely be affected by a carsharing scheme, while none categorically excluded this possibility. Despite the fact that 50% of the sample owns a bike, modal split shows that they don’t use them. Bikesharing could probably or absolutely affect 58.3% of the subgroup that planned to buy a bike in the next 6 months. The respondents indicate that they would more easily abandon the purchase of a bike than of a car if a sharing system were operated in their region of residence. About 63.2% and 77.7% of young Greeks declare that they are willing to spend less than 10 min to access a carsharing and bikesharing station, respectively. The higher percentage for bikesharing is attributed to the fact that bicycle is usually used for shorter distances.



(i) Bikeharing



(ii) Carsharing

Figure 5.1: Intention to join vehicle sharing systems

5.2.3 Data analysis and results

Satisfaction model

Respondents were asked to express their opinion in a rating scale [often called Likert scale, c.f. (Likert, 1932; Richardson, 2003)]. For example, they were asked to express their satisfaction regarding their travel patterns at a range from 1 (very dissatisfied) to 7 (very satisfied). An appropriate model for the analysis of such data is the ordered logit model (Train, 2009), which overcomes the violation of the independence between the errors and as a result the independence of the irrelevant alternatives (IIA) of a logit model (Ben-Akiva and Lerman, 1985) that is induced by the ordering of the alternatives.

The results of an ordered logit model for the respondents' satisfaction from their current travel patterns is shown in Table 5.2. Most of the estimated coefficients are significant at the 95% level, with the exception of one that is significant at the 90% level and one that is somehow below that. The signs and estimated magnitudes of the coefficients are reasonable and consistent with the a priori expectations. In particular, the main parameters that were retained in the final model as those that affect the satisfaction of the respondents with their current travel patterns include:

- Level of education. People with a doctorate degree are more satisfied with their current travel plans. Although this variable is not significant at a 90% or 95% confidence level, it has been retained in the model, as its sign and magnitude are consistent with expectations.
- Frequency of weekly trips to school. People commuting three times per week are the most satisfied, while the response decreases gradually for those who commute four, or five trips and more.
- Commute time to work or school. As total time spent for trips to work/school per day increases, the satisfaction of the commuters decreases. The time variables in the model are significant and negative, as expected.
- Mode of travel used for different trip purposes. People sharing a car or riding a bicycle for their trips to work or school are more satisfied, while those using public transport (bus/trolley/tram) for trips related to social activities are less satisfied than the others. Satisfaction is also high, when the commuters drive with fellow car passengers during their trips to groceries shopping.
- Age. The negative coefficient of the age variable reveals that the respondents in the age group 26–35 are less satisfied with their travel patterns than younger respondents.

The factors that affect the satisfaction are of interest to the policy makers, since their decisions should aim to increase, among others, the commuters' satisfaction. Moreover, this model helps to identify the factors that determine the satisfaction, which would then in turn be included as explanatory variables in the intention to join bar/bike-sharing models that follow, which are very relevant from a policy-making point of view.

Table 5.2: Model of satisfaction for current travel patterns

Ordered logit satisfaction model	**Variable**		
	Value	**Std. error**	**t-test**
Education: doctorate	0.92	0.60	1.52*
Trips per week to school: 3	1.59	0.57	2.79
Trips per week to school: 4	1.18	0.52	2.28
Trips per week to school: 5 or more	0.87	0.43	2.04
Mode for work/school trips: drive with others/Bicycle	1.02	0.56	1.83*
Mode for groceries shopping: drive with others	0.91	0.38	2.39
Mode for social activity: bus/trolley/tram	-1.33	0.52	-2.54
Time spent per day to work/school: 30–45 min	-1.52	0.38	-3.97
Time spent per day to work/school: more than 45 min	-2.62	0.56	-4.71
Employment status: working full/part–time	1.06	0.35	3.00
Age: 26–35 (reference level: 18–25)	-0.85	0.33	-2.60
Intercepts			
1|2	-4.56	0.81	-5.64
2|3	-2.28	0.46	-4.94
3|4	-0.75	0.41	-1.84*
4|5	0.82	0.40	2.03
5|6	2.07	0.43	4.81
6|7	4.29	0.56	7.71
Residual deviance: 532.09			
AIC: 566.09			

*Not significant at 0.05 level

Perception of car and bicycle ownership factors

Respondents were asked to indicate whether they consider a feature as an advantage or disadvantage of car/bike ownership (in a scale of absolutely no/probably no/maybe/probably yes/yes). A number of options were similar to those reported by Shaheen and Martin (2007) and a comparison is presented in Figure 5.2 (due to the different way the questions were phrased in the two studies, the options "yes" and "probably yes" were grouped for the Greek data). The graph shows that for young Greeks car-ownership is highly correlated with convenience (90% answered positively to the relevant questions). On the other hand, cost, pollution and parking difficulties are the predominant disadvantageous factors of car-ownership (more than 80%). The differences between the two studies can be attributed to various considerations, including: (i) the survey method followed was different (Shaheen and Martin, 2007) asked respondents to select which factors they consider as advantageous or disadvantageous and respondents could select between these five or indicate another); and (ii) socio-economic differences (Greeks between 18–35 years old versus Beijing residents of all ages).

Having said that, some interesting observations can be made from this comparison. The highest difference relates to the perception of safety, where 32% of young Greeks consider this an advantage of car-ownership versus 4% of Chinese respondents. This may be due to the very high perceived risk of two-wheelers in Greece (European Road Safety Observatory, 2007).

In order to reduce the number of variables, factor analysis (Sector 2.7; Washington et al., 2003) is performed (Table 5.3). Restrictions such as the factor rotation models are imposed to result in orthogonal or oblique models. Varimax rotation is a method that maximizes the sum of the variances of the loadings and is used for orthogonal rotation. Oblique models result in an interpretable factor structure. Different rotations are applied, and the one that results in factors closer to 1 or 0, is then selected. The larger the loading is the more influential to the respective variable the factor is. Loadings with values less than 0.3 are not shown, while those with values above 0.6 are in bold. The interpretation of the factors in relation to the variables is as follows:

- Car-ownership
  - Safety: Makes travel safer; is unsafe; causes driving stress.
  - Cost: Cost of purchase; and financial pressure.
  - Environment/parking: Parking difficulty; and environmental pollution.
  - Convenience: Travel convenience; comfort of travel; increased mobility and scope of activity.
- Bike-ownership
  - Personal/environmental benefits: Cost; environmentally friendly; and physical health.
  - Safety: Unsafe; and fatigue.
  - Flexibility: Flexibility and interoperability.
  - Infrastructure providence: Absence of parking stations and bike lanes.

The loadings can belong to more than one factors. For example, driving stress contributes both to the “Safety” and “Parking” factors.

Factor analysis reveals any latent correlation between the different variables. In car-ownership, for example, driving stress is correlated with safety to a factor (safety), but also with parking problems and environmental pollution to another (environment/parking). Concerning bike-ownership, respondents’ perception of environmental friendliness, low cost and physical healthy are highly related to a factor (personal benefits/environment). Moreover, there seems to be a relation between the perception of safety and fatigue.

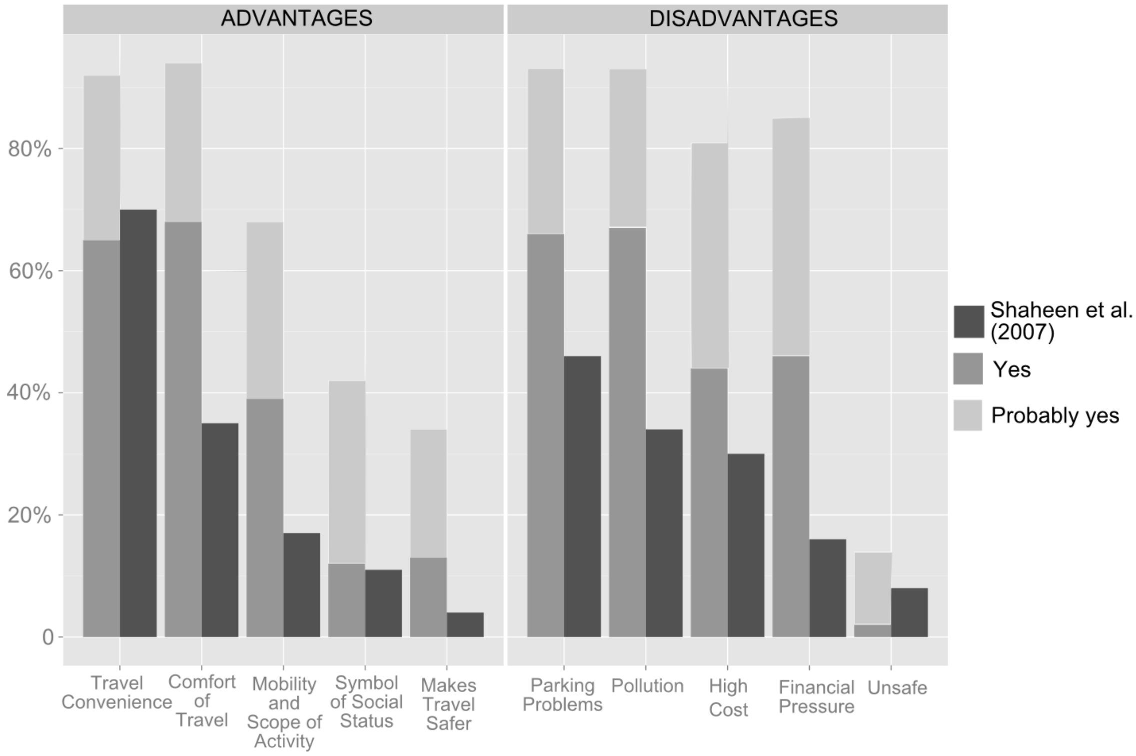


Figure 5.2: Perception of Car-ownership Factors

Figure 5.4 summarizes the perceived importance of the most important factors affecting carsharing and bikesharing adoption. In order of decreasing importance, the factors are distance of (bike- or car-sharing) station from house or job; ability to return the vehicle/bike in another station; ability to return the vehicle/bike without previously informing the center about the time and the station; time of day (e.g. day/night); reservation process; available type of vehicle/bike and availability of electric vehicles/bikes from the car/bikesharing company. The results indicate that the majority of the respondents consider as important factors those relevant with the stations' position both for carsharing and bikesharing, while time of day is more important for bikesharing usage. Almost half of the respondents are indifferent to the use of electric vehicles/bikes.

Table 5.3: Factor analysis for perceived car – and bike –ownership characteristics

Loadings	**Car ownership**			
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**
Advantages				
Travel convenience				0.602
Increase comfort of travel				0.588
Increase mobility and scope of activity				0.358
Makes travel safer	-0.405			
Disadvantages				
Parking problems			0.618	
Environmental pollution			0.672	
Driving stress	0.396		0.324	
High cost		0.976		
Financial pressure		0.481		
Unsafe	0.979			
Sum of square loadings	1.343	1.314	1.038	0.912
Proportion variance	0.134	0.131	0.104	0.091
Factor interpretation	**Safety**	**Cost**	**Environment/** parking	**Convenience**

Loadings	**Bike ownership**			
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**
Advantages				
Low cost	0.618			
Flexibility			0.411	
Environmental friendly	0.794			
Physical healthy	0.607			
Interoperability with other modes			0.894	
Disadvantages				
Absence of parking stations				0.578
Absence of bicycle lanes				0.562
Unsafe				
Fatigue		0.323		
Sum of square loadings	1.456	1.123	1.060	0.875
proportion variance	0.162	0.125	0.118	0.097
Factor interpretation	**Personal/environmental benefits**	**Safety**	**Flexibility**	**Infrastructure Provision**

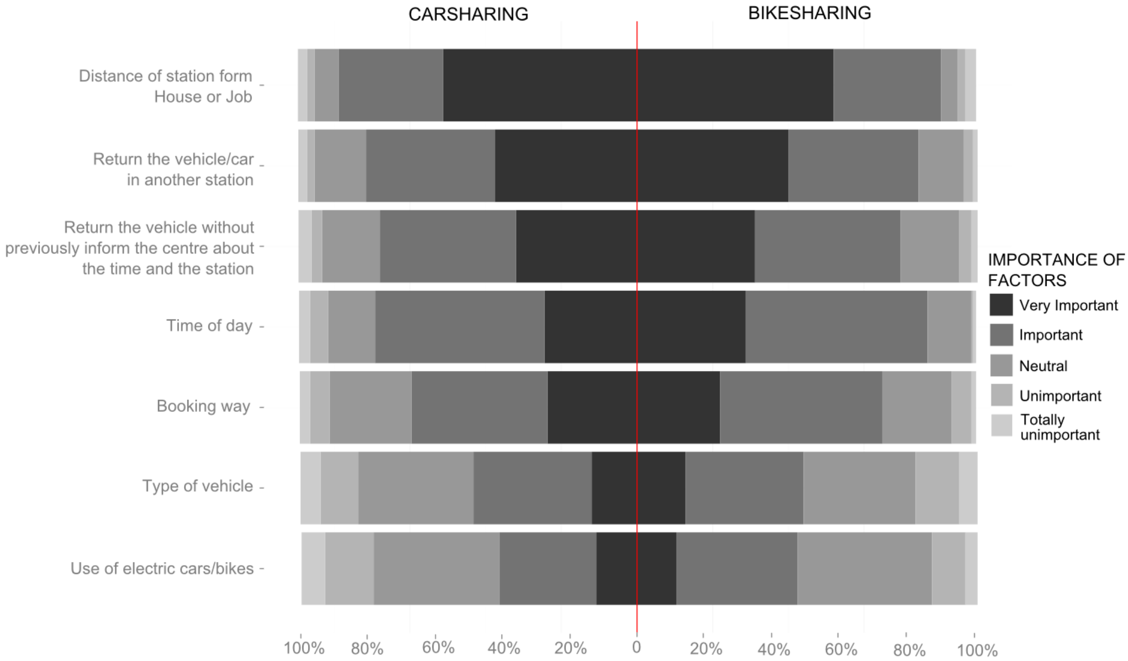


Figure 5.3: Importance of Carsharing and Bikesharing factors

Characteristics of prospective members

The survey examined the characteristics of the respondents that will join carsharing or bikesharing schemes. Three ordered logit models were estimated, one for each time period (within the first year, within the first few years and eventually) of joining carsharing and three for bikesharing. The explanatory variables of these models include the stated environmental consciousness, demographic characteristics, such as the household size and annual income, but also travel characteristics of the commuters. The results of these models are shown in Table 5.4.

Results show that those who indicate that they will join a carsharing scheme in the first year are less environmentally conscious than those that will do it in the next few years, which results in the highest coefficient for the three time-horizons. On the other hand, the most environmentally conscious will join bikesharing from the first year of operation. People with low household income (from 15,000 to 25,000 Euros) are more likely to become members, especially of a bikesharing scheme. Respondents who drove 100–150 km per day the week before the survey (the question was formulated in this way to obtain more reliable results, but it is really a proxy of the average distance traveled per week) and those who make their trips to social activity by taxi are more willing to join carsharing. On the contrary, those who drive 15–100 km per week or those who go to work or school by bus, trolley or tram are more likely to join bikesharing. According to these findings, bikesharing could be a substitute of public transport. Two independent variables of significant importance for the bikesharing models -but not for the carsharing-are age and education level. Respondents with a doctorate as well as those in the age group 26–35, are less willing to join the scheme than the less educated and younger respectively, despite the fact that those belonging of this age group are less satisfied of their current travel patterns.

Figure 5.4 provides a visual comparison of the estimations for the middle-time period (within the first few years). For each type of system (bike/car-sharing) the estimated coefficients and standard deviations of each variable are presented, along with their standard deviations (thicker lines) and twice the standard deviations (thinner lines).

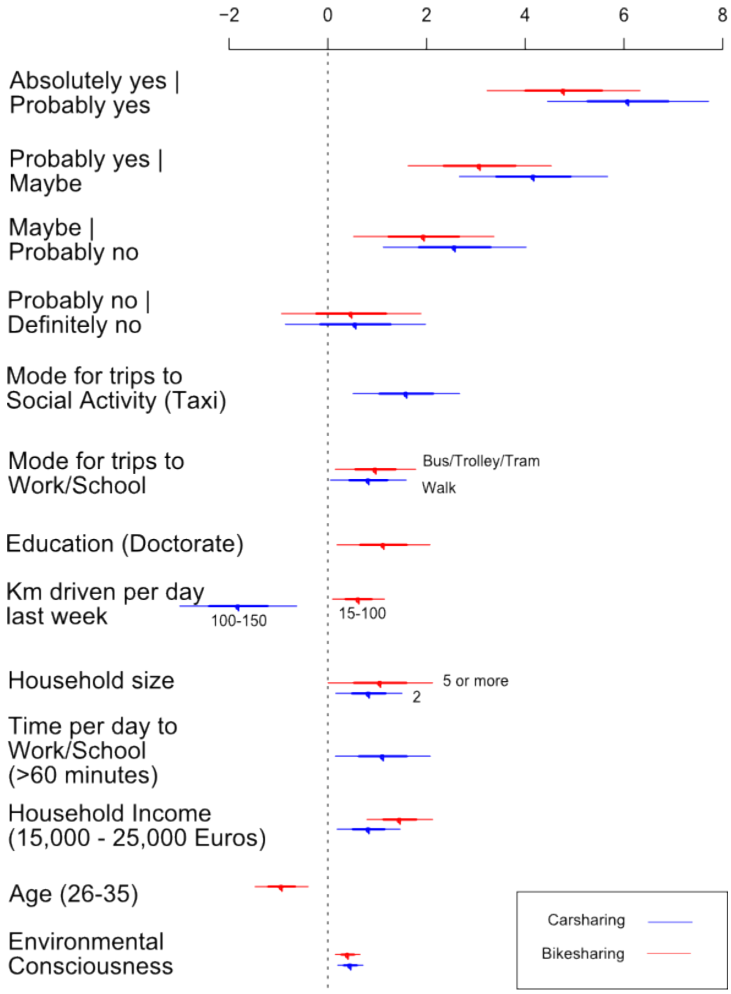


Figure 5.4: Comparative overview of estimated coefficients

Table 5.4: Intention to join car – and bike –sharing

Indicators	**Carsharing (N=233)**								
	Within the first year			**Within the first few years**			**Eventually (5-10 years)**		
	Value	Error	t-test	Value	Error	t-test	Value	Error	t-test
Environmental consciousness	0.33	0.12	2.70	0.46	0.13	3.58	0.38	0.12	3.08
Household income is 15–25E	0.62	0.31	2.02	0.93	0.32	2.92	0.67	0.30	2.22
100–150 km per day last week	-1.34	0.59	-2.26	-1.80	0.59	-3.04			
Take taxi for social activities	1.08	0.56	1.93	1.74	0.54	3.22	1.41	0.60	2.34
Walk to work/school				0.76	0.38	1.99			
Household size is equal to 2				0.69	0.33	2.07			
>60min per day to work/school				1.05	0.48	2.19			
Intercepts									
Definitely no/Probably no	0.45	0.67	0.67	0.47	0.71	0.66	-0.74	0.70	-1.05
Probably no/Maybe	2.29	0.69	3.34	2.66	0.72	3.68	0.91	0.67	1.35
Maybe/Probably yes	3.59	0.711	5.05	4.15	0.75	5.53	2.93	0.70	4.21
Probably yes/Absolutely yes	5.78	0.83	6.98	6.08	0.82	7.45	4.51	0.73	6.15
Residual deviance	**597.88**			**568.99**			**603.28**		
AIC	**613.88**			**590.99**			**617.28**		

Indicators	**Bikesharing (N=233)**								
	Within the first year			**Within the first few years**			**Eventually (5-10 years)**		
Environmental consciousness	0.47	0.12	3.82	0.40	0.16	3.21	0.40	0.12	3.28
Age is 26–35 years old	-0.73	0.27	-2.72	-0.88	0.27	-3.25	-1.19	0.26	-4.58
Household income is 15-25K	1.03	0.32	3.20	1.47	0.33	4.43	1.12	0.32	3.52
Mode for work/school trips is bus/trolley/tram	0.99	0.40	2.49	0.88	0.40	2.15	1.02	0.40	2.59
15–100km per day last week	0.62	0.26	2.37	0.58	0.26	2.19	1.12	0.32	3.52
Education: doctorate degree	-1.04	0.48	-2.19	-1.08	0.47	-2.29			
Household size is 5 or more				1.05	0.53	1.98			
Intercepts									
Definitely no/Probably no	1.07	0.68	1.56	0.38	0.71	0.55	0.11	0.69	0.16
Probably no/Maybe	2.77	0.70	3.97	2.03	0.71	2.86	1.67	0.69	2.42
Maybe/Probably yes	3.79	0.72	5.29	3.06	0.73	4.22	2.73	0.70	3.89
Probably yes/Absolutely yes	5.34	0.77	6.94	4.78	0.77	6.17	4.27	0.74	5.78
Residual deviance	**610.68**			**607.08**			**633.39**		
AIC	**630.68**			**629.08**			**651.39**		

5.2.4 Hybrid choice and latent variable model

Data

The dataset described in the previous section has been enriched with data collected by a paper survey (Derebouka, 2012). Individual interviews were performed at the metro stations, Malls, and local neighborhoods in Athens. These locations are suitable for data collection, because they are used by people of random demographic characteristics living in different areas around Attica, leading to a good classified sample. In a period of three months (April to July 2012) 194 completed paper questionnaires were collected (118 of which are used in this research, due to the age subgroup 18-35 that has been used). Although this survey led to a more representative sample of the population -in terms of age, education level and household size distribution- than with the on-line, in this paper we focus on the young respondents (18 to 35 years old), in order to be consistent with both datasets.

One difference between the two datasets was the range of 'satisfaction about the current travel patterns' question, which on the paper survey was taking values from 1 to 7, while on the on-line questionnaire from 1 to 5. Therefore, in order to make the responses compatible, the paper responses 2, 3 and 5, 6 were grouped (in two categories). The results of the surveys are expected to differ in terms of systematically different behavior, understanding of the subject, and honesty of the participants when responding. It is expected that the respondents of the on-line questionnaires tend to be more positive in their responses, while it is also possible that some of them skip reading the introductory paragraph of the carsharing systems, and were probably confusing it with carpooling. This bias was avoided during the paper survey, because the interviewer explained the carsharing scheme in detail, making sure that the respondent was not confusing it with carpooling before completing.

Model development

In this study, an ordered logit model is developed, to model the willingness of young Greeks (18 to 35 years old) to join a possible, future carsharing scheme in the mid-term future (1 to 5 years). The dependent variable takes values from 1 (absolutely no) to 5 (definitely yes). Moreover, the model contains a latent variable with ordered responses as nests, aiming to capture the latent satisfaction about the current travel patterns of the respondents. The two different datasets, on-line and paper responses, were used jointly. The method used is similar to the combination of SP and RP data, introduced by Ben-Akiva and Morikawa (1990) in order to overcome the shortcoming of the one to the other. Louviere et al. (1999) makes a review of studies combining different datasets for model estimation.

The hybrid model is composed by three utility functions: 1) the willingness to join using the on-line dataset, where a common scale is applied to every explanatory variable; 2) the willingness to join using the paper dataset and 3) the satisfaction. The explanatory variables of the utilities to join the carsharing scheme for both datasets are the same. The authors tried to retain the model consistent with the ordered logic model presented in the previous section, during the development. The estimation begun by including all the available variables, and then gradually eliminate the insignificant (one by one) until reaching the final specification, without considering the latent classes at first step. Then, the latent class of the satisfaction was developed, similar to the previous satisfaction model. However, not all the variables remained significant and as a result many of them were eliminated. The coefficients of the utility functions were estimated simultaneously using the maximum likelihood estimation in BIOGEME (Bierlaire, 2003).

The variables of the main model (willingness to join) are:

- $social_{taxi}$: the respondent uses taxi for trips to social activity; the model examines this case over the alternatives, which are the car, bicycle, public transport, and other
- $income_3$: a dummy variable that takes the value 1 for household income between 15 to 25 thousand euros, else 0; the alternatives are incomes <10, 10-15, 25-75, 75-100 and >100 thousand euros
- $envcon$: respondents perception of how environmental conscious they think they are; takes values from 1 to 7
- $satisfaction$: satisfaction of the respondents about the current travel patterns; it is the latent variable

The variables of the satisfaction latent class model are:

- age: a dummy variable that takes the value 1 if the respondent is within the age 26 and 35, else 0
- $time$: a dummy variable that takes the value 1 if the respondent spends more than 45 minutes travelling to work/school per day, else 0
- $social_{bus}$: a dummy variable that takes the value 1 if the respondent uses the bus for trips to social activities, else
- $married$: a dummy variable that takes the value 1 if the respondent is married, else 0
- $carown$: a dummy variable that takes the value 1 if the respondent owns a car, else 0

- *omega*: the error term of the latent variable

The utility functions of the main models and the latent variables are presented below.

V_1 is the utility function applied on the on-line dataset
V_2 is the utility function applied on the paper dataset
Sat is the structural equation of the latent satisfaction latent

$$V_1 = \beta_{scale} \times (\beta_{envcon} \times envcon + \beta_{sat} \times satisfaction + \beta_{inc_3} \times income_3 + \beta_{staxi} \times social_{taxi}) \tag{5.1}$$

$$V_2 = \beta_{envcon} \times envcon + \beta_{sat} \times satisfaction + \beta_{inc_3} \times income_3 + \beta_{staxi} \times social_{taxi} \tag{5.2}$$

$$Sat = \beta_{age_{2635}} \times age + \beta_{time} \times time + \beta_{sbus} \times social_{bus} + \beta_{married} \times married + \beta_{carown} \times carown + \beta_{sigma} \times omega \tag{5.3}$$

Figure 5.5 shows the full path diagram of the Willingness to Join model with Latent Variables, in a form that is similar to Walker (2001). X_{1-8} are the explanatory variables of the structure and the latent models, $U_1^{on-line}$ is the utility of the on-line survey respondents, U_2^{paper} is the utility of the paper survey respondents, and y is the Propensity to Join. Tables 5.5 and 5.6 present the model estimation results (Table 5.6 separates the ordered logit models' threshold estimates for clarity). Comparing the model with Table 5.4, it is observed that household size, high education level (doctorate), using public transport to work/school, and age are not significant in the new models. On the other hand, the income (€15,000 - €25,000) and the environmental consciousness are significant in both models. The more environmental conscious the respondent is, the more likely it is that he will join the scheme, while the more satisfied he is with his current travel patterns, the less likely it is (to join the new scheme). Respondents with middle income (between 15 and 25 thousand euros per year) are more likely to join. This probably indicates that people with lower income expect carsharing to be expensive, and they will still prefer to travel by public transit or foot. On the other hand, those with higher incomes will still find more attractive the choice of using their own vehicles. Moreover, employees, and those who use taxi for their trips to social activities (taxi is very popular for social trips in Greece) are more likely to join.

Concerning the satisfaction model, variables that were found significant in Table 5.2, where the model was estimated separately and are not significant here, are: high education level (doctorate), number of trips to work/school, mode for food shopping drive with others or bike and travel time to work/school in the range of 30-40 minutes. On the other hand the car ownership became significant. More specifically: young respondents between 26 to 35 years old are less satisfied; the more time they spend for trips to work/school the less satisfied they are; those using public transport for trips to work/school are less satisfied; satisfaction is higher for married respondents and those that own a car.

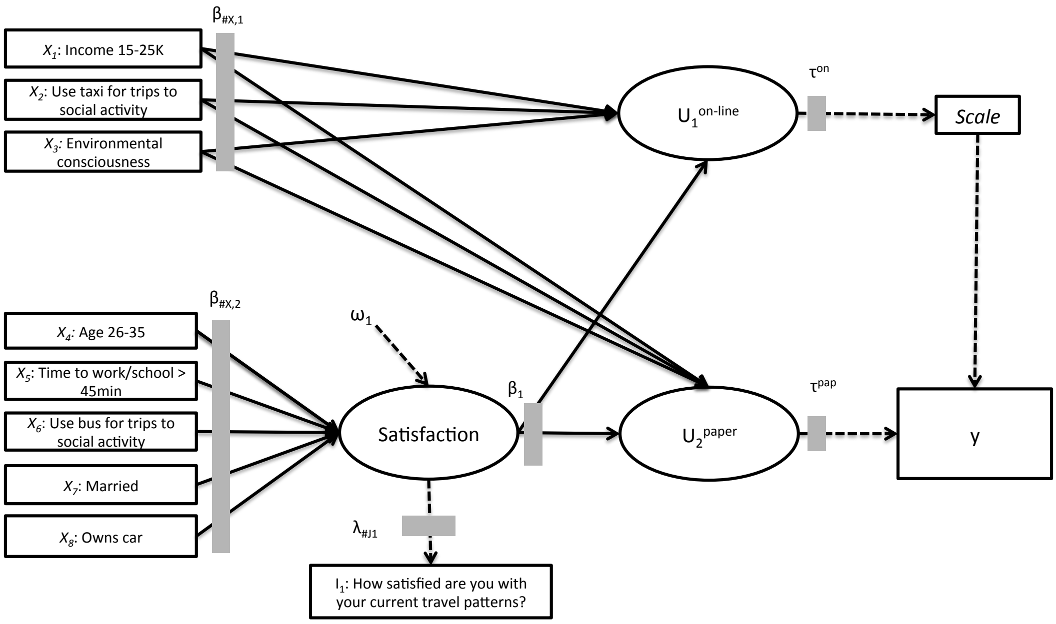


Figure 5.5: Comparative overview of estimated coefficients

Table 5.5: Hybrid choice and latent variable model

Variable	Value	t-test
β_{scale}	0.797	2.17
β_{envcon}	0.340	2.84
β_{inc3}	0.630	2.04
β_{sataxi}	1.77	2.27
β_{sat}	-0.456	-1.80*
β_{2635}	-0.696	-2.67
β_{time}	-0.959	-3.29
β_{sabus}	-1.46	-3.07
$\beta_{married}$	0.751	1.49*
β_{carown}	0.850	3.05
σ	1.02	2.18
Data sample n = 351		
Log-likelihood = -908.459		
$\bar{\rho} = 0.638$		

(·) for p $\lesssim 0.1$, (*) for p $\lesssim 0.05$, (**) for p $\lesssim 0.01$, (***) for p $\lesssim 0.001$

Table 5.6: Intercepts of hybrid choice and latent variable model

Model	Variable	Value	t-test
Main (on-line)	1|2	-1.02	-1.40*
	2|3	1.05	9.02
	3|4	2.54	9.46
	4|5	4.38	6.95
Main (paper)	1|2	-0.76	-1.14*
	2|3	1.03	5.49
	3|4	2.78	6.75
	4|5	5.18	5.19
Satisfaction latent	1|2	-4.40	-7.07
	2|3	-1.29	6.24
	3|4	0.54	6.39
	4|5	4.00	6.92

(·) for p $\lesssim 0.1$, (*) for p $\lesssim 0.05$, (**) for p $\lesssim 0.01$, (***) for p $\lesssim 0.001$

5.2.5 Conclusions

In this research, the characteristics of the shared mobility systems bikesharing and car-sharing – electrified or not – are presented and compared and the results of an survey are quantitatively analyzed. A questionnaire was first disseminated electronically – via social media and mailing lists – mainly in Greece, where both schemes are still somewhat unknown. The dataset was later enriched with paper survey data. The aim is to identify the satisfaction of young Greeks about their current travel patterns and their perception toward car – and bike –sharing schemes.

A factor analysis is performed to reveal possible hidden correlation between the car or bike-ownership advantages and disadvantages, while the factors affecting the future adoption of carsharing and bikesharing schemes are analyzed quantitatively. Moreover, ordered logit models are estimated for the satisfaction of young Greeks about their current travel patterns, and their willingness to join vehicle-sharing schemes in different time horizons (immediately, in the next couple of years and eventually). The satisfaction model reveals that people sharing a car with others or using bicycle for trips to work or school are more satisfied, while satisfaction decreases when the time spent to the destination increases. In addition, Greeks of ages between 26 and 35 are less satisfied than those younger than 26.

The estimations of the "join the scheme" models using the on-line data, show that income affects significantly the probability of joining these schemes, with those belonging to the low-mid income class (€15,000) - €25,000) to be more likely. The models suggest that carsharing may attract people that currently use bus, tram or trolley for trips to work or school, while bikesharing those who go on foot. Age is also a significant determinant of the willingness in joining bikesharing, with Greeks from 26 to 35 years old being more reluctant than younger, despite the fact that they are less satisfied of their current travel patterns. Furthermore, carsharing is shown to be more popular among those that currently use taxi for their trips to social activity. Moreover, the more environmental conscious the respondent declares that he is, the more possible it is that he will join one of the two schemes.

Thereafter, the dataset in enriched with data collected with paper face–to–face survey. The propensity to join carsharing is then modeled using hybrid choice and latent variable models and mixed paper and on-line data. Satisfaction of the respondents about their current travel patterns is included in the form of a latent variable, structured by one discrete-response indicator. Two different utility functions were developed for each dataset, and a unique scale parameter was applied on the variables of the on-line utility. The reason behind this scale is to measure the distance between the responses of the participants of the different datasets. The structural and latent models were estimated simultaneously.

The results show that, as expected, the Internet survey respondents are prone to give more positive answers, probably in order to satisfy the interviewer. This is verified by the positive scale variable into the model. Young Greeks with income €15,000 - €25,000 and those who use taxi for their trips for social activity, are more willing to join a possible carsharing system in the future. The more satisfied people are with their current travel patterns, the less possible it is that they will join a scheme. Young respondents between 18 and 25 years old, married and car owners are more satisfied with their current travel patterns. On the other hand, those who use bus for their trips to social activity and those who spend much time traveling, are less satisfied. Comparing these findings with the ordered logic model it is observed that variables that were significant in the Internet-data model are not significant in the hybrid-mixed, and vise versa. For example, the hybrid-mixed model shows that the household size, the education level and the age, do not determine the willingness to join. Moreover, the results show that the responses of the on-line data are over-rated, probably leading to biases.

The results might differ from one study to the other, but the aim is to show that when applying different modeling methodologies and/or using different data, this is possible. Researchers and policy makers should be aware of these potential differences during transport policy planning.

The current financial hardships can offer a unique opportunity for the deployment of vehicle-sharing systems. Already in Greece there is a spike in the number of people using bicycles for their daily travel, while many more abandon their cars and start using public transport. Vehicle-sharing systems need critical mass to attract a lot of users (and therefore be effective and profitable), but users will not join them in large numbers until they are widely available. The larger cities in Greece (Athens, Thessaloniki) and the rest of Europe are fertile ground for such initiatives. The small steps that have been taken to date need to be made bolder and move toward the generation of new ideas, in order to revitalize the economy and provide synergies with different sectors.

Modeling the willingness of the commuters to join future vehicle-sharing schemes, constitutes a valuable tool for policy makers and potential investors in estimating the expected demand of the system. The results of these models should be taken into consideration at the planning phase. The spatial allocation of the demographic characteristics of the prospective users, should determine the selection of the optimal location of stations. The initial succeed to provide the service to those needed more, will lead to minimisation of the investment risks.

Future research may help to demonstrate how the sustainable development of urban regions could be affected by such systems. Furthermore, the information presented in this Section could be useful toward the integration of these vehicle-sharing schemes into traffic simulation software and – perhaps more importantly – in LUTI models.

5.3 Public transportation: the impact of recession on user satisfaction and demand

5.3.1 Introduction

Public transport is traditionally the most effective weapon against traffic congestion, especially in metropolitan areas. The quality and performance of transit services has attracted the interest of many organizations, researchers and scientists the recent years. According to Tyrinopoulos and Antoniou (2008), several approaches have been used as quality measurements, such as traveller s satisfaction, loyalty, and benchmarks. In the US and Europe several manuals have been published aiming to build a framework for quality measuring, such as: the TRB Handbook for Measuring Customer Satisfaction and Service Quality (Transportation Research Board, 1999), the TRB Transit Capacity and Quality of Service Manual (Transportation Research Board, 2004), and the EC-CEN Transportation-Logistics and Services-Public Passenger Transport-Service Quality Definition, Targeting and Measurement (European Committee for Standardization , 2002).

The quality of transit services has been examined by different perspectives the last few years. Srinivasan et al. (2007) examined the impact of quality changes in public transport domain of the developing countries, by applying multivariate models for the current and previous mode choice decisions. They concluded that there is imminent need to improve the capacity of urban public transportation systems in cities of developing countries. Eboli and Mazzulla (2011) introduced a quality indicator that is based on direct and indirect measurements, such as information provided at the stations (e.g. maps), stations comfort, frequency and punctuality of the service, but also price. Cirilli et al. (2011) found that one of the most important factors that determines the quality of the public transport service is the punctuality. About one third of the respondents answered that they would be willing to pay more for an on-time service.

Taylor et al. (2008) analyzed the data collected by operators in 265 urban areas in USA and found that there is a high spatial differentiation of the quality, namely the regional geography, the metropolitan economy, the characteristics of the population and the characteristics of the road network. Moreover they found that the reduction/increase of ticketing cost and frequency could double (or halve) the demand for the systems in an urbanized area. They conclude that high frequency attracts more passengers, while high price keeps them away.

The above are only few of the numerous examples of studies and researches on the subject. However, the analysis of the quality of public transport services in times of economic crisis is gaining an increasing interest. How does the economic crisis affect commuters modal choices? In times of crisis, which are the main factors that determine public transport satisfaction? How have these factors changed compared to the pre-crisis period? And finally, what can public transport organizations do to adapt to the new situation? Greece has entered a long period of economic crisis with adverse effects on various aspects of daily life, including mobility. Many urban areas need to adapt to new conditions and challenges driven by the economic recession. Public transportation in Greece is controlled by the State. As such, public transport is considered a social good and therefore it can and must play a key role in this new situation.

The primary objective of this research is to measure the effects of the economic crisis on the satisfaction of the public transport passengers. Moreover, the factors that affect the increase of the individual demand for public transit trips are examined. Athens and two of the existing transportation systems (bus and metro) are used. The results of the analysis are compared with the findings of Tyrinopoulos and Antoniou (2008), who performed a similar research in the same area in 2008, when the country s economy was still developing.

5.3.2 Case study and previous research

The transportation infrastructure in Athens is described in Section 3.2.1.

In this research, the satisfaction of the users of two public transit operators: AMEL, the Athens metro operating company, and ETHEL, the company of thermal buses is investigated.

AMEL

The objective of AMEL is to manage and operate the underground metro network in Attica, but also the additional infrastructure such as the facilities, stations, vehicles and media. The metro network is composed by two lines (23 stations), which are integrated with the urban light rail line operated by ISAP, the characteristics of which are not analyzed in this research. These two lines serve 650,000 passengers per day. The frequency of the trains is 3 min in peak hour and 5-10 min at other times (Hellenic Institute of Transport, 2003).

ETHEL

The thermal bus service is the system with the highest demand because of the extended service area coverage. According to the Hellenic Ministry of Public Works (Hellenic Ministry of Public Works, 2009), the average bus speed in the city is 7km/h, while on the suburbs 18km/h, mainly because of the limited length of the bus lanes that are usually violated.

ETHEL manages: a) 40 lines that connect the core centers Athens and Piraeus, with the smaller municipalities in Attica; b) 20 peripheral inter-municipal lines that do not cross the core centers; c) 123 municipal-level lines that support the main lines; d) 19 express lines; e) seven school lines. The company operates 16,000 trips daily, owns 2099 buses and operates 1822 of them (Hellenic Institute of Transport, 2003).

Data collection

The data for this research were collected by face-to-face interviews with public transport passengers. Based on the requirement that the results should be comparable with Tyrinopoulos and Antoniou (2008), to allow for direct comparisons between the two datasets, the margins for change were limited. Minor modifications and additions were made, mainly including questions about 1) the degree of criticality of the characteristics, and 2) addition of questions regarding the current economic crisis.

The questionnaire is composed by the following three parts: 1) demographic questions, such as age, gender and occupation; 2) travel patterns, such as number of metro/bus lines daily used, frequency of usage, trip purpose and main reason of using public transport, and finally, 3) quality assessment of the service. The last part contains questions regarding the "importance" and the "satisfaction" of the respondents with respect to several quality factors. Interviewees were asked to rate how important or satisfied they are concerning the following factors Tyrinopoulos and Antoniou (2008):

1. General characteristics of the public transit system

    - Service frequency: frequency of the service in the lines of the transit systems.
    - On-time performance: accuracy of the departure times of the vehicles at terminal stations in relation to the predefined schedule.
    - Service provision hours: operating hours of the service provision on a given day.
    - Network coverage: spatial coverage of the area under consideration with public transit services.
    - General information provision: sufficiency of the information provided to the passengers about the general characteristics of the transit services, (e.g. terminals and stops points).
    - Types of tickets: sufficiency of the various available types of tickets with respect to the coverage of the needs of the public.
    - Prices of tickets: price-structure of the various types of tickets available.
    - Tickets selling network: sufficiency of the tickets selling network and the ease to purchase tickets from the various selling points.
    - Personnel behavior: behavior of the various types of personnel of the transport operator (e.g., drivers), when communicating and transacting with the passengers.

- Existence of bus lanes: sufficiency and performance of the bus lanes to facilitate the efficiency of the transit service
- Measures for environmentally friendly public transit: contribution of public transit in the protection of the environment and the adequacy of the relevant actions and measures taken by the relevant authorities.

2. Terminals and stops

- Walking distance: distance that passengers have to walk from the origin point to the closest terminal and stop.
- Information provision: sufficiency of the information available to the passengers about the services provided at the terminals and stops
- Conditions: conditions of the terminals and stops concerning shelter, visibility, seating capacity, etc.
- Safety: perceived sense of safety of the passengers when waiting at the terminals and stops to use the public transit service.

3. Vehicles

- Onboard conditions: conditions inside the vehicle during the execution of a journey (e.g., seats and air-conditioning).
- Vehicles cleanliness: level of cleanliness of the vehicles from various standpoints (seats, handles, etc.).
- Driving behavior: driving performance of the vehicle s driver.
- Onboard information provision: provision of information inside the vehicle during the trip.
- Accessibility to disabled and mobility impaired people: provision of facilities by the transit operator to facilitate the accessibility of transit services by disabled and mobility impaired people.
- Safety: perceived sense of safety of the passengers when on-board.

4. Transfer points

- Distance between transfer points: distance that passengers have to walk between transfer points in order to continue their trip.
- Waiting time: time that passengers have to wait at transfer points in order to continue their trip.
- Information provision: provision of information to passengers at the transfer points about the combination of the various lines and modes, and their time schedules.

The questionnaire concludes asking the overall satisfaction of the respondents about the transit quality in Athens, and how much more or less they use it compared to five years ago. The data were collected by personal, face-to-face interviews at the metro and bus stations, mainly during the rush hour. In total, 400 interviews were performed; 213 with metro users and 187 with bus users. Most of the analysis was performed using the statistical software R (R Development Core Team, 2014), while for the estimation of the hybrid choice and latent variable model BIOGEME (Bierlaire, 2003) was used.

5.3.3 Results and comparison

Factor analysis

Similar to the previous study (Tyrinopoulos and Antoniou, 2008), factor analysis has been performed on the importance of the quality attributes. The number of factors has been chosen based on the criterion to explain more than 10% of the variance. The resulted factors differ from 2008, since they are composed by different quality attributes. The unobserved factors revealed are: 1) service quality, 2) information/courtesy, 3) service production, and 4) transfer quality.

Tables 5.7 and 5.8 demonstrate the factor analysis results of 2008 and 2013 for metro (AMEL) and buses (ETHEL). Figure 5.6 shows the satisfaction s distribution comparison between 2008 and 2013, and Figure 2 the ternary plots in order to assist the comparison of the factors according to their total proportion variance. As mentioned in Tyrinopoulos and Antoniou (2008), the triangle/ternary plot is a convenient way to visualize the results of a three-factor analysis. The proportion variances of the factors have been normalized to 1. Each edge of the triangle represents an axis for each factor. The projection of the points to the sides is made parallel to the next edge, following the clock-wise direction. Concerning AMEL passengers, Information/courtesy is currently considered as more important (38%) than in 2008 (34%). Quality of service is perceived as a combined factor with service production, unlike five years ago, and it is considered important by 20%, while service provision and quality of service capture a 30%. ETHEL passengers today consider as more important service production (28%) than in 2008 (19%). Quality of transfer and service compose a single factor, the most important for 2013 (40%), but lower than in 2008 when both factors combined were perceived important by 67% (27% and 40% respectively).

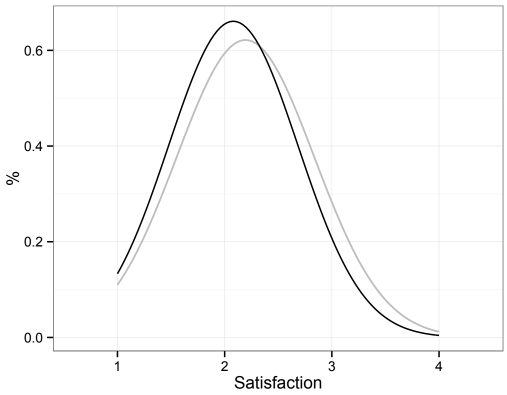
Ordered logit model

In the next step of the analysis, ordered logit models of the passengers satisfaction from public transport operators were estimated. The dependent variable of these models is overall stated satisfaction, while all possible satisfaction attributes were included as explanatory variables in the first estimation, and then insignificant factors were eliminated one by one until reaching the final specification; however, it was decided to maintain some parameters even if not significant at 95% confidence interval, if the signs and magnitudes were reasonable, as this may be caused by the small number of observations. The results are compared with the models of 2008, revealing changes in the factors that affect the passengers satisfaction.

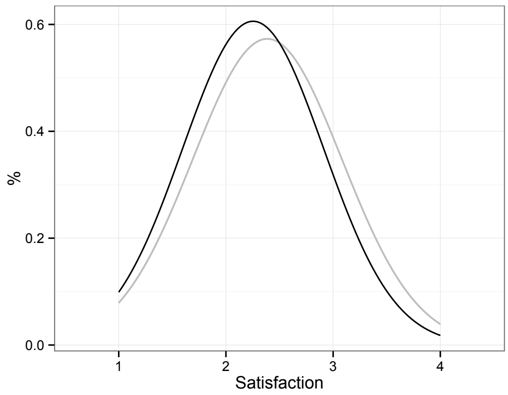
The satisfaction explanatory variables are ordered in a 5-level Likert scale, while the response in 4-level. The respondents were asked to select between being “very dissatisfied” (value 1) and “very satisfied” (value 5 or 4 respectively). The results are presented in Tables 5.11 - 5.10. Unlike the findings of the 2008 study, the attributes considered as more ‘luxurious’, such as the behavior of the driver and non-driver staff, do not determine the satisfaction of the AMEL and ETHEL services. Passengers of both operators are currently more satisfied when the information provided regarding transfers increases, while in 2008 this attribute was not significant in the model. This trend could be a consequence of the improvement of the travel information provided, after the development of the travel-planner platform Google Transit for Athens in 2011. This platform provides detailed information about the times and transfers of the public transit in the area. However, this remains a personal estimation of the authors, since there was not a question related with this platform in the questionnaire.

The respondents perceive that AMEL service is characterized by punctuality and frequency, a finding that was implicitly confirmed in 2008, although the variables per se were not significant in the model. The European Performance Satisfaction Index (European Performance Satisfaction Index Rating Institute, 2005) makes a similar conclusion, namely that the AMEL passengers have the highest satisfaction among other public transit passengers in Europe. In contrast to 2008, AMEL passengers satisfaction is not affected by the ticketing systems (addressing fares, tickets types and availability) provided, despite the fact that the system remains the same as five years before. This could imply that either they have become familiar with it, or they expect a better system (e.g. electronic ticket/card).

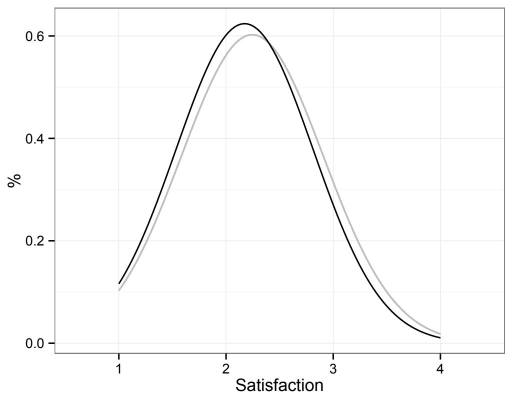
ETHEL passengers seem to realize and appreciate more today than in 2008 that the network coverage of the system is high and there are stops near their origins and destinations. On the other hand, in 2013, they are not satisfied with the punctuality. This may be the result of the frequent closures of the city center due to protests, which lead many service lines to be interrupted regularly.



(i) ETHEL



(ii) AMEL



(iii) Both

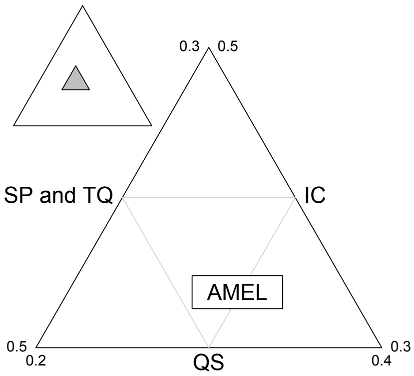
Figure 5.6: Satisfaction comparison between 2008 (grey) and 2013 (black)

Table 5.7: Importance factor analysis results - AMEL

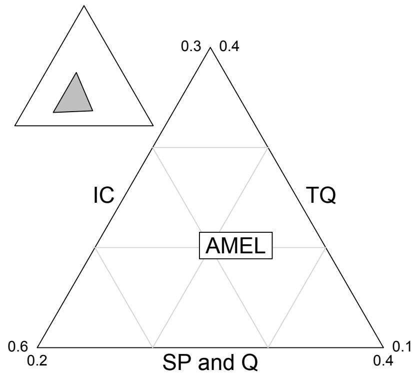
	AMEL					
	2008			2013		
	Factor 1	**Factor 2**	**Factor 3**	**Factor 1**	**Factor 2**	**Factor 3**
Service frequency	**0.921**				**0.666**	
On-time performance	**0.875**				**0.747**	
Service hours	0.447				**0.624**	
Timetable information			0.361	0.485	0.400	
Price		0.467			0.401	0.314
Behavior of personnel			0.542	**0.634**		
Existence of bus lanes		0.469		0.487	0.307	
Distance/time to access stop		0.412			0.391	
Timetable information at stop			**0.611**	0.508		0.332
Waiting conditions at stop		**0.724**		0.501		0.393
Condition in-vehicle		**0.990**		0.421	0.440	0.414
Driver behavior			0.426	**0.709**		0.316
Information in-vehicle			**1.065**	**0.677**		
Accessibility (w.r.t. disabilities)	0.351			0.569		0.420
Transfer distance	0.575					**0.678**
Transfer waiting time	0.599				0.419	**0.624**
Information regarding transfers			0.524	0.489		0.470
Sum of square loadings	**2.861**	**2.336**	**2.630**	**3.393**	**2.638**	**2.092**
Proportion variance	**0.168**	**0.137**	**0.155**	**0.200**	**0.155**	**0.123**
Factor interpretation	Service production/ transfer quality	Quality of service	Information/ courtesy	Information/ courtesy	Service production and quality	Transfer quality

Table 5.8: Importance factor analysis results - ETHEL

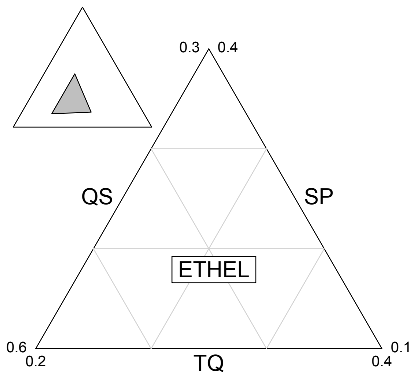
	ETHEL					
	2008			2013		
	Factor 1	**Factor 2**	**Factor 3**	**Factor 1**	**Factor 2**	**Factor 3**
Service frequency			**0.768**	0.388		**0.637**
On-time performance			**0.739**			**0.947**
Service hours				0.343		0.325
Timetable information	**0.608**			0.401		0.231
Price	0.444			0.419		0.338
Behavior of personnel	**0.783**				**0.746**	
Existence of bus lanes				0.334	0.370	
Distance/time to access stop			0.354	0.500		
Timetable information at stop	**0.659**			0.484		0.428
Waiting conditions at stop	0.468				0.488	0.361
Condition in-vehicle	0.578		0.380	0.307	**0.662**	
Driver behavior	**0.771**				**0.711**	
Information in-vehicle	0.454			0.338	0.560	
Accessibility (w.r.t. disabilities)	0.460		0.351	0.415	0.404	
Transfer distance		**0.920**	0.575	**0.780**		
Transfer waiting time		**0.822**	0.599	**0.650**	0.327	
Information regarding transfers	0.504	0.533		0.496	0.398	0.311
Sum of square loadings	**3.504**	**2.298**	**1.597**	**3.049**	**2.986**	**2.347**
Proportion variance	**0.197**	**0.135**	**0.094**	**0.176**	**0.176**	**0.138**
Factor interpretation	Quality of service	Transfer quality	Service production	Transfer and service quality	Information/ courtesy	Service production



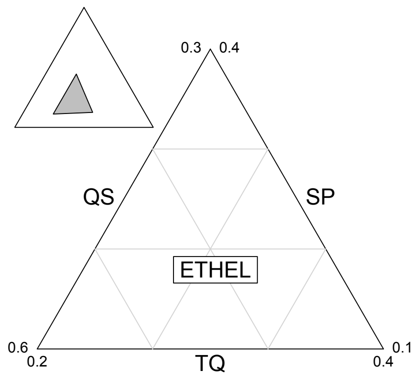
(i) AMEL - 2008



(ii) AMEL - 2013



(iii) ETHEL - 2008



(iv) ETHEL - 2013

Figure 5.7: Ternary plot comparison

Table 5.9: Satisfaction ordered logit model estimations - AMEL

	AMEL					
	2008			2013		
	Value	**S.E.**	**t-test**	**Value**	**S.E.**	**t-test**
Punctuality	-	-	-	0.506	0.184	2.757
Transfer distance	-	-	-	-	-	-
Network coverage	-	-	-	-	-	-
Vehicle cleanliness	0.660	0.152	4.338	0.252	0.151	1.674
Driver behavior	-	-	-	-	-	-
Waiting conditions	-	-	-	-	-	-
Ticketing systems	0.297	0.140	2.118	-	-	-
Behavior of non-driver staff	0.404	0.145	2.781	-	-	-
Service frequency	-	-	-	0.615	0.218	2.829
Information regarding transfers	-	-	-	0.445	0.157	2.838
Time of operation	-	-	-	-	-	-
Distance/Time to station	-	-	-	-	-	-
Intercepts						
1|2	8.24	0.85	9.69	2.37	0.60	3.65
2|3	4.66	0.66	7.12	5.99	0.74	8.15
3|4	1.54	0.58	2.67	10.29	1.06	9.66
Residual deviance	363.6			347.3		
AIC	376.6			361.3		

Table 5.10: Satisfaction ordered logit model estimations - ETHEL

	ETHEL					
	2008			2013		
	Value	**S.E.**	**t-test**	**Value**	**S.E.**	**t-test**
Punctuality	0.247	0.118	2.097	-	-	-
Transfer distance	0.208	0.099	2.096	-	-	-
Network coverage	0.282	0.087	3.229	0.602	0.185	3.257
Vehicle cleanliness	0.361	0.105	3.450	-	-	-
Driver behavior	0.181	0.096	1.877	-	-	-
Waiting conditions	0.270	0.101	2.681	-	-	-
Ticketing systems	-	-	-	-	-	-
Behavior of non-driver staff	-	-	-	-	-	-
Service frequency	0.380	0.126	3.005	0.576	0.221	2.604
Information regarding transfers	-	-	-	0.612	0.211	2.897
Time of operation	-	-	-	0.586	0.208	2.818
Distance/Time to station	-	-	-	0.434	0.196	2.214
Intercepts						
1|2	10.37	0.66	15.61	5.26	0.86	6.12
2|3	6.69	0.51	13.04	10.29	1.21	8.54
3|4	2.64	0.41	6.46	15.00	1.69	8.89
Residual deviance	868.0			237.8		
AIC	888.0			253.8		

Table 5.11: Satisfaction ordered logit model estimations - all operators

	All operators					
	2008			2013		
	Value	S.E.	t-test	Value	S.E.	t-test
Punctuality	0.189	0.068	2.793	0.187	0.141	1.314
Transfer distance	0.237	0.060	3.929	0.265	0.147	1.807
Network coverage	0.233	0.053	4.412	0.394	0.109	3.611
Vehicle cleanliness	0.288	0.060	4.765	0.262	0.114	2.307
Driver behavior	0.161	0.066	2.451	-	-	-
Waiting conditions	0.260	0.060	4.367	-	-	-
Ticketing systems	0.141	0.051	2.738	-	-	-
Behavior of non-driver staff	0.143	0.059	2.416	-	-	-
Service frequency	0.416	0.072	5.744	0.626	0.165	3.798
Information regarding transfers	-	-	-	0.452	0.130	3.470
Time of operation	-	-	-	-	-	-
Distance/Time to station	-	-	-	-	-	-
Intercepts						
1|2	10.48	0.40	26.00	3.591	0.514	6.983
2|3	7.04	0.33	21.56	7.611	0.649	11.722
3|4	3.32	0.27	12.17	11.937	0.903	13.214
Residual deviance	2457.2			609.6		
AIC	2481.2			627.6		

Hybrid choice and latent variable model

Hybrid choice and latent variable models are state–of–the–art discrete choice analysis models. They integrate the latent behaviors into the model, measuring the error between the different questions. Atasoy et al. (2012) showed that they have higher prediction accuracy, which implies better forecasting, the objective in policy analysis. In this research, a hybrid choice and latent variable model has been developed to estimate the overall satisfaction of public transit passengers in Athens. The use of this type of model was considered essential due to the large number of satisfaction variables that refer to similar attributes.

The latent variables are constructed by more than two ordered discrete indicators: 1) Information awareness, 2) quality of service, 3) service production and 4) transfer quality. A factor analysis using the joint AMEL and ETHEL datasets was first performed, in order to identify any common patterns between the different attributes and select the indicators for each latent variable. Figure 5.8 shows the full path diagram of this model, and Table 5.12 the results. The formulation of this model is:

$$V_{satisfaction} = \beta_{2529} age_{2529} + \beta_{st} student + \beta_{eco} economy + \beta_{nocar} nocar + \beta_{sat} satisfaction + \beta_{SP} SP + \beta_{IN} I + \beta_{QS} QoS + \beta_{TQ} TQ \quad (5.4)$$

During estimation, some demographic variables were maintained even if not significant at 95% confidence interval, as suggested by Ziliak et al. (2008). The results show that passengers between 25 and 29 years old and students are more satisfied, while those who use public transport by necessity, because they are more financially constrained or they do not own a car, are less satisfied. The service production, information awareness, quality of service and transfer quality latent variables are all positive and significant. This model allows the inclusion of more satisfaction attributes into the model (13 instead of 6 in Table 5.12), as indicators of the latent variables, and as a result the better exploitation of the available information.

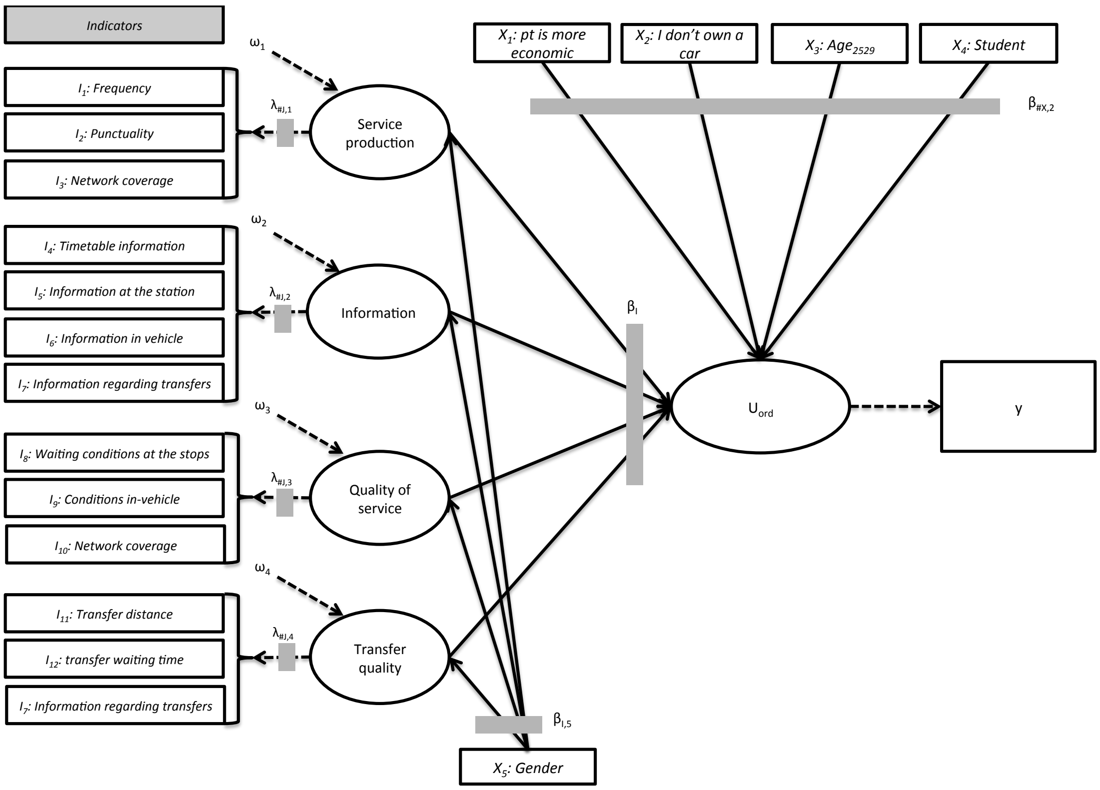


Figure 5.8: Full path diagram of hybrid choice and latent variable model of overall satisfaction

Table 5.12: Satisfaction hybrid choice and latent variable model

Variable	Specification	Value	t-test
β_{2529}	Dummy, 1 if between 25-29 y.o.	0.740	2.70
$\beta_{student}$	Dummy, 1 if student	0.693	2.37
β	Dummy, 1 if he/she does not own a car	-0.815	-2.13
β	Dummy, 1 if main reason for using public transport is economy	-0.811	-2.29
β_{SP}	Latent service provision	0.786	5.00
β_{IN}	Latent information awareness	0.236	1.79
β_{QS}	Latent quality of service	0.146	2.23
β_{TQ}	Latent transfer quality	0.325	2.46
σ_{SP}	Error for β_{SP}	0.389	3.96
σ_{IN}	Error for β_{IN}	0.524	5.59
σ_{QS}	Error for β_{QS}	0.969	14.50
σ_{TQ}	Error for β_{TQ}	0.466	5.06
$\beta_{femaleSP}$	Dummy, 1 if female	0.367	0.61
$\beta_{femaleIN}$	Dummy, 1 if female	-0.343	-1.53
$\beta_{femaleQS}$	Dummy, 1 if female	-0.807	-2.10
$\beta_{femaleTQ}$	Dummy, 1 if female	-0.731	-3.62
1|2		-2.85	-9.39
2|3		4.45	-13.47
3|4		4.55	8.28
ρ		0.391	

Increase of demand

In this step, the identification of the attributes that may help determine the increase of public transport demand between the last five years is attempted. The respondents were asked to rate 'How much more or less they use public transit now than 5 years ago', responding in a 5-level Likert-scale. Using this as dependent variable, and demographic and travel attributes of the individuals as explanatory, we estimated an ordered logit model. Table 5.13 shows the results of the models.

The specification of the utility function is:

$$\begin{aligned} V_{more/less} = \beta_{female} female + \beta_{2529} age_{2529} + \beta_{3039} age_{3039} + \beta_{work} work + \\ \beta_{education} education + \beta_{economy} economy + \beta_{nodrive} nodrive + \\ \beta_{bikefoot} bikefoot + \beta_{usecar} usecar + \beta_{sat} satisfaction \end{aligned} \tag{5.5}$$

The results indicate that women are using bus more than before, while respondents between 25 and 39 years old are using public transport more than the rest age categories. The employees and self-employed respondents are using the metro less, which is probably related with the previous employment status (e.g. if they were students 5 years ago). Those whose main trip purpose for using public transport is work or education indicate higher usage of metro and bus.

Table 5.13: Demand ordered logit model

		All		**AMEL**		**ETHEL**	
Variable	**Specification**	**Value**	**t-test**	**Value**	**t-test**	**Value**	**t-test**
β_{female}	Dummy, 1 if female	0.297	1.62	-	-	0.778	2.57
β_{2529}	Dummy, 1 if between 25-29 y.o.	-0.683	-3.25	-0.766	-2.82	-0.704	-2.01
β_{3039}	Dummy, 1 if between 30-39 y.o.	-0.913	-3.84	-0.756	-2.17	-	-
$\beta_{selfempl}$	Dummy, 1 if self employed	-	-	-0.970	-2.21	-	-
β_{empl}	Dummy, 1 if employed	-	-	-1.14	-2.64	-	-
β_{work}	Dummy, 1 if main trip purpose is work	1.32	5.83	1.47	4.70	0.880	2.08
$\beta_{education}$	Dummy, 1 if main trip purpose is education	1.41	5.54	0.878	2.19	1.25	2.69
$\beta_{economy}$	Dummy, 1 if main reason for using public transport is economy	0.425	1.83	-	-	-	-
$\beta_{nodrive}$	Dummy, 1 if main reason is that he/she does not drive	0.431	1.90	-	-	-	-
β_{nocar}	Dummy, 1 if main reason is that he/she does not own a car	-	-	0.562	1.46	-	-
$\beta_{bikefoot}$	Dummy, 1 if main reason for using PT less than before, is that he/she uses bike/walk more	-3.32	-6.62	-3.24	-4.65	-	-
β_{usecar}	Dummy, 1 if main reason for using PT less than before, is that he/she uses car more	-1.86	-5.96	-1.60	-3.47	-1.27	-4.19
β_{daily}		-	-	-	-	1.61	5.09
β_{sat}	Ordered 1 to 4	0.441	2.98	0.758	3.21	0.493	1.88
1|2		-2.44	-3.08	-2.81	-2.52	0.064	0.10
2|3		1.27	6.47	1.27	4.91	1.19	6.62
3|4		1.95	8.29	2.18	6.89	1.20	6.47
4|5		1.88	10.50	1.90	7.56	1.63	7.40
Data sample		399		213		186	
ρ		0.801		0.808		777	

5.3.4 Conclusions

The shift of commuters and travelers from private car to public transport and other soft modes, like bicycle and walking, is one logical effect of the economic crisis. The last three years of the deep economic recession in Greece have shown that people gradually abandon their car due to high fuel rates, increased taxes etc., and they turn to alternative modal choices. In this respect, public transport operators should take advantage of this timing and not only attract new market segments (e.g. low income individuals), but also maintain these new passengers by providing high quality transit services. Therefore, understanding the passenger's preferences and factors that drive modal choices due to the economic crisis, and determine public transport satisfaction is of vital importance.

The aim of this research is to present the results of an extensive analysis of the passengers' satisfaction variation, as a consequence of the economic crisis, using two transit systems in Athens: bus and metro. The results of the analysis have been compared with the findings of a respective research performed in 2008 in the same area (Athens) and the changes in the passengers' behavior have been further investigated. The data used in this research have been collected with face-to-face surveys, using a questionnaire similar to the corresponding research in 2008. The respondents were asked to rate how important and significant a number of transit attributes are.

Appropriate statistical methods have been applied. More particularly, factor analysis has been first performed, in order to uncover important transit attributes according to the opinions of passengers, to reduce the number of attributes and to reveal any indirect relations between them. The satisfaction of the respondents has then been modeled using ordered logit models, and hybrid choice and latent variables model.

The factor analysis revealed that metro users (AMEL) consider information provision and courtesy more important now than in 2008, while less emphasis is given to the quality of service in relation to five years ago. This is a rather anticipated finding, since metro is considered as the highest quality transport system in Athens and as such the quality of service is taken for granted, while users place more emphasis on attributes that will facilitate more their daily trips like transfer quality and information provision. Moreover, the absence of other high quality transport systems imposes the shift of passengers' focus to other factors.

On the other hand, currently service production (mainly addressing service frequency and on-time performance) is by far the most significant factor for the bus users (ETHEL) compared to the previous survey. Traditionally, service characteristics like on-time performance have been a serious burden for the Athens bus transport due to the severe traffic congestion and in effect the inability of this mode to keep its schedule, and the frequent violation of the bus lanes. Nowadays, with the emergence of the economic crisis, these service characteristics become even more important. It is noted that since the launch of metro in Athens, bus transport plays a feeder role to the metro system. The economic crisis forced many commuters living in remote urban areas (suburbs) to use bus more often for local trips in order to access a metro station and therefore, the service production plays a decisive role in their mobility. The ordered logit models revealed that metro service punctuality and frequency is highly perceived by the respondents in both surveys. Again, this finding confirms that the Athens metro remains at the highest position in the satisfaction rates that users give to the metro system in Athens.

Another interesting finding is that fares and other ticketing attributes received low satisfaction rates by the metro users in the present survey, contrary to 2008. The economic crisis has certainly played a key role in this finding. Despite the fact that the value for money for the Athens metro is high (the fare for using the metro is relatively low for the quality of the service; 1.4 euro is the fare for using all transit modes in a time period of 1.5 hours), the economic pressure especially in the low-income citizens imposes a decrease of the fares according to the opinions of the passengers.

As for the bus users, they highly appreciate the network coverage of the bus lines, even more than in 2008. This is the main advantage of the bus system in Athens and, in combination with the coordination with the metro network, both modes provide an excellent alternative for travellers especially for long distance trips. On the other hand, the main disadvantage of the bus service in Athens is punctuality; also highlighted by the present survey. Taking into account the numerous demonstrations that took place in the Athens city center in the last three years, then an already existing problem has been deteriorated, and this has been reflected in the results of the present survey.

The hybrid choice and latent variables model provided further useful insight into the latent behaviors of the respondents and proved to be particularly helpful for market segmentation analysis. According to the respective analysis, users between 25 and 39 years old use public transportation more than five years ago. This can be explained by the fact that the current unemployment rate in Athens in those ages has been increased drastically due to the economic recession; therefore those users cannot afford using their private cars and they turn to public transport. The same analysis showed that students are more satisfied with the transit services. This could be due to the fact that the students have fewer requirements from the public transport services than the rest users.

These models can be employed to predict the impact of public transport improvement policies – such us upgrade of service production, information provision, quality of service or transfer quality – on users satisfaction and public transport demand. The increase of demand, in combination with the investment costs, could be analyzed in a LUTI model to evaluate different policies.

Concluding, this research has been the first attempt to understand the impact of the economic crisis to the quality of the public transport services in Athens from the users point of view. The analysis performed employing appropriate statistical methods derived quite interesting results and findings about the influence of the economic recession to the factors affecting the choices of transit users and their satisfaction compared to the period before the emergence of the crisis. Public transport operators and governmental agencies may use these results in order to adapt to the new situation. For example, AMEL and ETHEL should place more emphasis on attributes like information provision that will facilitate the transfer between metro and bus. The passengers prefer using a combined mode of transport, especially for long trips, and any action that will facilitate this choice would be beneficial not only for them abut also for the transport operators in order to attract more customers. Furthermore, the government should examine the possibility to give incentives, such as reduced fares, especially to low-income or unemployed citizens, aiming to use more public transport.

5.4 Discussion

Public transit systems form an integral component of the urban structure. Factors such as network coverage and location of stations, quality of service and information provision, determine the demand and as a consequence the system's ability to maximise the benefits for the society and economy. Leveraging the current economic crisis, which creates new prospects in public transit system improvement, to our benefit, is essential.

At first, the potential implementation of emerging public transit systems such as carsharing and bikesharing in Greece has been explored. Then, the impact of the crisis on public transit users' perception of service quality has been investigated. In both cases, advanced behavioral modeling methodologies are employed. Public transport investment schemes aiming to improve the current systems' service provision, but also the implementation of alternative, sustainable transportation modes such as vehicle–sharing schemes, should be further analysed in LUTI models.

6 Opportunistic data sources

6.1 Research objective and chapter structure

Large data collection projects usually fall within the remit of the public authorities, non-profit organizations and private companies. Nevertheless, emerging Internet technologies, and more specifically Web 2.0, social media and smartphones, that currently compose the technological state–of–the–art, support the development of initiatives that promote alternative ways of collecting information. Under these circumstances, the public can participate intentionally –or not– in data collection, a procedure that is widely known with the term crowdsourcing (Howe, 2006). As many such initiatives operate without being covered by an umbrella of a funded scheme, their aim is to collect data of high quality with the minimum cost, benefiting from the power of the crowds. Social media serve as crowdsourcing platforms in many ways. The variety of their members' characteristics (nationality, age, level of education, interests) renders them powerful tools, suitable for different applications. In the past, people have used social media for marketing and political campaigns, for news updates by information networks or people, for professional advertisement and recruitment and, most recently, for organization of protests revolutions. The data collected from these events are sought for the research community, which has successfully used them accordingly.

The motivation for this research arose from the need to find the various datasets required for ongoing research activities related with transportation in general, and LUTI models in particular. The requirements of LUTI models in data, remains one of the main concerns of the researchers, because of the multitude and the level of detail that is needed; LUTI models become increasingly disaggregate, in order to respond to the more recent methodological developments (e.g. agent-based modeling), while the combination of different sub-models (e.g. household, job, building location choice) requires demographic, building and transport data (Waddell, 2002). In this research, the potential of using opportunistic data collected from the web, for the development of Integrated Land-Use and Transport Models is investigated. Various possible data sources are presented and their suitability is evaluated, with particular reference to applications that have been developed for the purpose of this doctoral research.

The rest of this chapter is structured as follows: The Section following the introduction describes potential opportunistic data sources, from crowdsourcing data of social media, to formal, centrally controlled public and private databases. Sections 6.3 to 6.5 provide information about the accessible web data sources that are suitable for each LUTI dataset (households, transport, buildings, jobs), accompanied by a detailed description of the particular applications that have been developed for the purposes of this dissertation. More specifically: 1) a real estate data collection tool from publicly available web-sites, 2) the use of crowdsourcing data for network quality assessment and 3) the use of social media (twitter, facebook and Google) for transport and demographic data collection. Section 6.6 makes conclusions and recommendations for the future.

6.2 Data sources

6.2.1 Crowdsourcing

The term crowdsourcing was introduced by Howe (2006) in an article published at the Wired magazine, as the act of a company to outsource a function that was performed by employees, to a large network of individuals. Estellés - Arolas and González Ladrón-de-Guevara (2012), review the various definitions given since then, and suggest the most representative until now. More succinctly, crowdsourcing here is defined as the procedure where volunteers perform the work that used to be made by professionals. Well-known examples of web sites that successfully use “crowds” to collect data are the free on-line encyclopedia Wikipedia (wikipedia.org), Open Street Maps (openstreetmaps.org) but also any kind of Open Source software. Moreover, one of the most known crowdsourcing data collection applications are the crises maps, such as that created after the earthquake in Haiti in 2010 (Zook et al., 2010). The advantages of using crowd-sourced data include their being: 1) free for use by the public, professionals and researchers, 2) easily, instantly accessible, 3) usually up-to-date, and 4) often collected by the latest means of technology. The quality of the resulting datasets differs among platforms, and depends on the amount (and expertise) of participants, administrators and the time of the data-collection process.

Doan et al. (2011) distinguish crowdsourcing into the following categories: 1) explicit systems, where users collaborate explicitly, meaning that they can evaluate (e.g. star rating systems), share (e.g. pictures, knowledge), network (e.g. social media), build artifacts (e.g. Wikipedia) and execute tasks (e.g. demand media), and 2) implicit systems, where users collaborate to solve problems implicitly, either standalone (e.g. bet, play) or piggyback on other systems (e.g. search engines, buy products). Goodchild and Glennon (2010) use the term Volunteered Geographic Information (VGI) to describe the spatial crowdsourcing data. According to the same research, the contribution of numerous users in gathering geographic information, results in systems sometimes of higher quality than those developed by a single user, even if (s)he is specialized. The term Volunteered Geographical Systems (VGS) (Savelyev et al., 2011), is used for the integration of social networks and VGI. Savelyev et al. (2011) give an example of using smartphone applications that integrate VGI and social media for emergency response situations. VGS are based on the willingness of people, not only to report the cases, but also to actively respond to other requests (e.g. www.avego.com).

Social Media, a relatively new technological way of social networking, serve as crowdsourcing platforms. Many social media have been developed but some of them never became popular. These platforms are addressed either to users with common interests, or qualifications, or to the general public. Social media users are of different nationality, age, education level, employment status and they have different interests. The wide variation of its members' characteristics renders social media suitable for data-collection. According to the statistics published by facebook on January 2011 (http://www.facebook.com/press /info.php?statistics), the most popular social network had more than 500 million active users, 50% of whom log on every day. On average, each member has 130 friends and in total, they spend more than 500 billion minutes every month logged on. Facebook is available on more than 70 languages and 70% of its users are outside the USA. Madden (2010) found that in 2010, 61% of the Internet users in USA utilize social networks, up from 46% the year before, while the 86% of those are between 18 to 29 years old.

The use of social media for marketing and political campaigns, for news updates by information networks or people, for professional advertisement and recruitment and, most recently, for organization of protests revolutions, has become common. The wealth of the data, hidden in this information, can offer valuable insight regarding attitudes and perceptions of the general public, as well as information on emerging incidents as they happen. Different twitter applications have been developed until now, some of which create a connection between twitter and the statistical software R (R Development Core Team, 2014) or tools providing similar functionality. The integration of these platforms offers a powerful tool for statistical analysis. The thoughts or actions of twitter' s users that are publicly available via the "tweets" can be scrapped, parsed and analyzed in R using the appropriate packages such as XML (Leung and Newsam, 2012).

Social media have already been used for transportation-related applications (Carvalho et al., 2010; Amey et al., 2011; Bregman, 2011). More recently, Crooks et al. (2013) analyzed spatial geo-referenced tweets, to show how twitter can be used as a distributed sensor system for earthquakes. Moreover, flickr and facebook include functions for geo-referenced image uploading on the web, creating new prospects for their usage as tools for road quality information collection. Crowdsourcing applications became widely available since the introduction of Web 2.0. While in the conventional web the used data are prespecified, Web 2.0 enables users to add data themselves. In this framework, the collection of data has become a very easy and promising procedure (Smith et al., 2009). Another factor that contributed to facilitating the data-collection practices is the widespread use of smartphones. Smartphones are equipped with a number of cheap but very powerful tools used for data collection, such as accelerometers, gyroscopes, Bluetooth, GNSS sensors and cameras, while their ability to go online and connect to social media allows the transfer of this data to Web 2.0 applications (Yang et al., 2012).

6.2.2 Centrally controlled databases

Public

Many States around the World have made efforts to organize and maintain on-line, centralized, formally controlled databases. European Union established the Infrastructure for Spatial Information, generally known as INSPIRE Directive, in order to gather the spatial data in a common European platform. The aim of this action is to provide directly the necessary data for the evaluation of policies that may have an impact on the environment of the Community (INSPIRE, 2007). The INSPIRE Directive is a framework that provides the technical implementing rules that is needed to be followed by each country of the European Union. These rules are common and are adopted in a number of areas, such as Metadata, Data Specifications, Network Services, Data and Service Sharing and Monitoring and Reporting (http://inspire-geoportal.ec.europa.eu).

For example, in Greece the INSPIRE data are uploaded on the website http://geodata.gov.gr. In this portal urban spatial data such as the following are stored: 1) geographical, (e.g. cadastral parcels, urban zones and Municipality borders, in different levels); 2) environmental, agricultural and hydrographical (e.g. location and size of NATURA areas, green belts, forests, rivers); 3) geological; 4) land cover (CORINE) and 5) transport information (e.g. locations of bus/metro stations, transit schedules).

Other open national data websites include: 1) http://data.gov (US), 2) http://data.gov.uk (in UK), 3) http://data.gov.it (Italy), 4) http://datos.gob.es (Spain), 5) http://data.gouv.fr (France), 6) http://data.norge.no (Norway), 7) http://data.belgium.be (Belgium), 8) http://data.gc.ca (Canada), 9) http://data.overheid.nl (Netherlands), 10) http://data.govt.nz (New Zealand), 11) http://opengovdata.ru (Russia), 12) http://data.gov.au (Australia), 13) http://datos.gob.cl (Chile), 14) http://dados.gov.br (Brazil), 15) http://opendata.ee (Estonia), 16) http://dados.gov.pt (Portugal), 17) http://data.gov.in (India), http://data.gv.at (Austria), 18) http://open-data.europa.eu (EU). More information can be found in http://census.okfn.org/, where all the open data catalogs around the world are listed. This platform is being operated by the open knowledge foundation (OKF), a non-profit organization that promotes the open data around the world.

Commercial

Private entities also own and manage large amounts of information, especially information that requires heavy capital investment, such as satellite imagery. The most well known geo-spatial database managed by a private company is probably Google Earth/-Google Maps. Many of the data available on this platform have not been exploited for transportation research, yet. More specifically, the street-view application can be used to reduce the need for direct on-site visits, when possible, thus considerably reducing the cost of many tasks. For example, on-site surveys of land-use patterns and building intensity in an area (such as those undertaken in Antoniou et al., 2011) could be replaced (at least partially) by walkthroughs in Google Street View. This tool provides an unprecedented level of detail and resolution in large parts of the world and can supplement data available in other databases, due to its level of detail; theoretically one can either conduct parking surveys using Google Street View data.

6.3 Land-use data

One of the primary datasets required for the development of an Integrated Land-Use and Transport Model, is the land-uses/buildings. This dataset contains information about the land-use, number of residential or not residential units per building, surface, price of property, year of construction and other building attributes. The structural characteristics determine the attractiveness of the building/land-use to households and jobs, and as a result affect the household and employment location choice by increasing or decreasing the utility of the agents. These data are usually collected and provided by the authorities that conduct the building census, or the National Departments of Economics. In Greece, the buildings and their structural attributes are available by the Hellenic Statistical Authority (ELSTAT). Although ELSTAT provides them tabulated without charge, they are useless without the Geographical Information System that assigns each building in space, which costs more than 20,000 euros. Moreover, the Ministry of Finance provides -in hard copies- the average dwelling prices per square meter per zone (defined descriptively by the names of the surrounding streets).

6.3.1 Real estate prices and structural attributes

The unavailability of credible centrally managed databases of real estate prices and building attributes, often redirects researchers into alternative sources, such as the real estate websites. These websites are social networks that connect the property owners, the customers and the real estate agents. Despite the limitations of the listed prices and their difference with real transaction values, they remain a good proxy (Chapter 3). Similar to Löchl and Axhausen (2010) who developed an application in python, an application has been developed in R (R Development Core Team, 2014) for the purposes of this doctoral research, to parse these data from the web, aiming to estimate dwelling purchase and rent price models. In this Section, the deployment of this tool is described in detail. The collected data were used for the estimation of real estate price models for Athens (Section 3.2) and Thessaloniki (Section 3.3) in Greece.

The Hellenic Ministry of Finance and Economics provides the average property prices per zone (defined descriptively by the names of the surrounding streets), only in hard copies. In order to avoid the time – and resource–intensive procedure of converting them into electronic media, which would have probably cost many person-months, while the prices would be outdated, it was decided to follow Löchl and Axhausen (2010) and read the prices from a real estate website. For this purpose, a tool was developed in R, using the packages XML (Duncan, 2014b) and RCurl (Duncan, 2014a) that provide the necessary functionality. Within a period of 5 months (from September 2011 to January 2012), the following house attributes were parsed: 1) price, 2) type of house (i.e. dwelling), 3) square meters, 4) number of bedrooms, 5) floor, 6) year of construction, 7) availability of independent heating, 8) air-conditioning, 9) garden, 10) fire place, 11) parking availability, 12) type of parking, 13) type of view, 14) orientation and the 15) geo-location (X,Y coordinates) on the Coordinate Reference System (CRS) ellipsoid WGS84. Duplicate data, houses of unusual sizes and prices, and data with missing geographical information, price, or year of construction, were removed. However, several factors that might cause biases remained: 1) not all the dwellings for sale and rent are advertised in the internet, 2) wrong attributes could be used in the classifieds and 3) most important, the final transaction price is often different than the listed (Horowitz, 1992). However, list prices remain a good proxy of transaction prices. Lyons (2013) performed a research in Dublin and found that these two values are highly correlated in space and time. The outcome of this process was the construction of dwelling databases for the two biggest cities in Greece: Athens (8066 properties for sale and 8400 properties for rent), (Sections 3.2, 3.4 and 3.5), and Thessaloniki (2130 for sale and 2483 for rent) (Section 3.3).

6.3.2 Points of interest and transport infrastructure

Transportation infrastructure locations are widely available in centrally controlled web databases. In Section 3.2 the locations of transportation stations in Athens, which the Athens Urban Transportation Organization (OASA) made publicly available on http://www.geodata.gov.gr, are used. This is a centrally-controlled public web-site where data are listed according to the Infrastructure for Spatial Information in Europe (INSPIRE) Directive. For the same research, the position of the airport, ports, marinas, Attica Tollway and the inner-ring, were downloaded from the public website http://www.index.pois.gr. in the end of 2012, indicating practically how the open information can effectively be exploited for research purposes once made freely accessible.

The locations and characteristics of non-residential land-uses are provided in an informal way by crowdsourced portals. Social media users can "check" in places such as restaurants, bars, gas stations, pharmacies, health centers, garages and sports centers, using facebook (http://www.facebook), foursquare (http://www.foursquare.com) or tripadvisor (http://www.tripadvisor.com), composing databases of points of interest, location, and characteristics but also the demand of the users (times voted, score, demographic characteristics of the voters). The value of such data in developing an Integrated land-use and transport model, could be proved substantial in many ways: 1) estimation of social activity location choice model, 2) development of indicators measuring the attractiveness of the facilities, 3) measurement of their effect on real estate prices, 4) estimation of household and employment location choices and 5) estimation of residential and non-residential development project location choice models, when these data are recorded in time. In Section 3.2 we use the locations of the non-residential buildings (point of interest) and the boundary of the inner-ring (traffic control policy in Athens), available at http://www.index.pois.gr, for the estimation of the real estate price models. The findings of this research are analyzed in Chapter 3 of this dissertation.

6.3.3 Land uses from land cover

Several applications have been developed for the collection of land-use information from crowdsourcing data, in conjunction with publicly accessible satellite images (remote sensing). Leung and Newsam (2012) suggested a methodology to extract land-uses using geo-referenced images, available in social media. More precisely, they downloaded from Flickr –using its application programming interface (API)– images of two University campuses in US, and they assigned them to their real location on the campus map, while at the same time they classified the images using the bag–of–words (BoW) model, widely applied in computer-vision (Jiang et al., 2007).

Fritz et al. (2009) have developed an on-line platform (www.geo-wiki.org) that uses the crowds to validate the land-cover, by employing highly resolution remote sensing images from Google, in order to improve the quality of the analysis. In the future, land-cover images will be proved useful for the development of LUTI models. Siora (2011) developed a remote sensing methodology based on object-based image analysis to extract land cover from satellite images, and applied econometric models, to model the land-use change (urbanization) in time. Finally, remote-sensing methodologies could be applied to extract the size of the buildings, which indicates their capacity. The data collected from this method are used for the development of spatial econometric models with discrete response, for the parametrization of land-use/cover (Chapter 3).

6.3.4 Environmental data

Real estate prices and household location choices depend on environmental indicators such as the availability of green spaces and CO_2 emissions. Environmental data are usually available through publicly managed databases. An ongoing European FP7 co-funded project named COBWEB (http://cobwebproject.eu/), aims to enable people to participate as 'sensors' in environmental information collection via crowd-sourcing, in order to enhance the European centrally managed databases. Moreover, the recently developed crowed-sourced platform AthensTreeMap (http://www.athenstreemap.gr) enables the stakeholders to import the location of trees and their type, in Athens.

6.4 Transport data

6.4.1 Mobility data

Transportation researchers further exploit the potential of employing widely used social media, such as facebook twitter and flickr, for mobility data collection. Efthymiou and Antoniou (2013b) describe the development of a preliminary twitter application that can retrieve information about the number of the tweets containing the words "carsharing", "bikesharing" and "electric vehicles", and the geographical location of the users, coded in R (R Development Core Team, 2014). The script then reads the time and location of the users from the html page and prints graphs and maps using the "ggplot2" (Wickham, 2014) package of R.

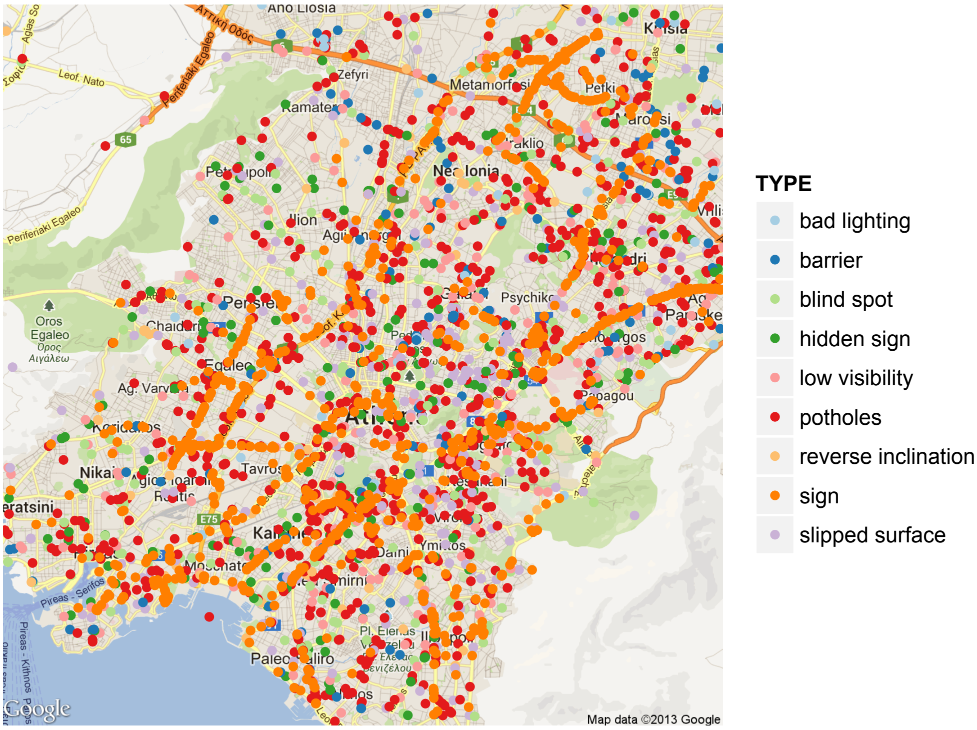
Normally, travel data are collected via surveys/travel-diaries, a time and money consuming practice that is susceptible to biases. The emergence of new technologies that use smartphones has opened new horizons in travel data collection. Vautin and Walker (2011) distinguish three main advantages of the smartphones over the traditional travel-diaries: 1) they provide real-time data, 2) personalization of travel information and 3) advanced sensor technologies. Smartphones can be tracked continuously, indicating the exact location of the user. A massive number of applications are currently available; from route planners to maps whose databases are continuously enriched from daily inquires by the users. The Stanford model determination app (Nham et al., 2008) has been developed as a methodology to predict the mode used by acceleration and speed measurements. Bierlaire et al. (2013) have developed a similar methodology using Bluetooth data. Microsoft research center in China has recently released, under the framework of the GeoLife project (Zheng et al., 2009), GPS trajectories of 182 users collected in a period of three years. This dataset contains information about the location (latitude, longtitide, altitude) but also the scope of travel activity (go to home, work, for shopping, sightseeing, drinking, hiking and cycling). The trajectories were recorded by GPS equipped phones and GPS loggers, and are very detailed (locations every 1-5 seconds or 5-10 meters). Moreover, data of location based social networks (LBSN), such as facebook and foursquare, that make use of the "check in" function, could be used if available. The aforementioned data can be used for the OD matrix generation, needed for the transport component of the Integrated Land-Use and Transport Model. Walk Score (http://www.walkscore.com) provides walking, bike and transit accessibilities per neighborhood. The computation of the accessibility score is based on an algorithm utilizing the distance of a house to the closer amenity. These scores can be used as variables for the estimation of a real estate price model, or household location choice models. It has been proved that the higher walking scores add about 4,000 to 34,000$ per dwelling (Cortright, 2009).

6.4.2 Network and quality assessment

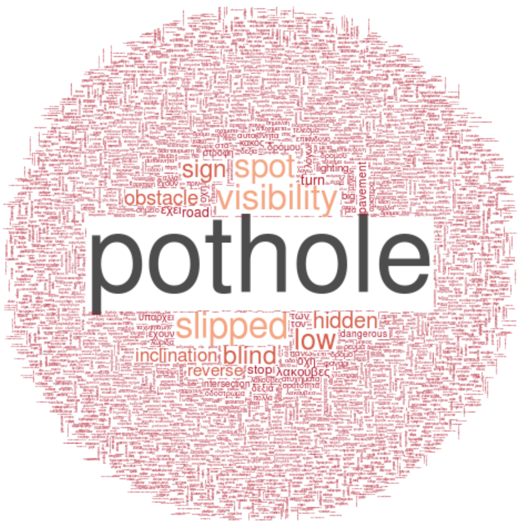
The most known example of a platform hosting VGI is the OpenStreetMaps (OSM), which actually is a map of the Earth, freely accessible and editable by the users. The OSM database (composed by the road/rail/waterway network, the images but also by other spatial variables) is widely used by transportation researchers and consultants, who use the crowdsourcing information for professional reasons.

Organizations have developed applications for assessing the quality of the road network. SeeClickFix (http://www.seeclickfix.com) is an on-line crowdsourcing tool where the public can report neighbor issues and see them fixed. Its role is intermediary and links the public with the policy makers. Citizens, media organizations, community groups and governments, use SeeClickFix for different purposes. Once a volunteer reports an issue, another can access, vote, edit the description and add information about it. The more the contribution from the users, the more possible it is that the local governments will 'fix' the deficiency. SeeClickFix uses all the characteristics needed to improve life of citizens: Transparency, collaboration, scale, efficiency, and simplicity (http://www.seeclickfix.com).

The Road Safety Institute Panos Mylonas (IOAS) launched the "Pin-Project" in 2007 (http://www.msfree.gr/pin). The aim of this initiative is to motivate the users of the Greek road network to report voluntarily, by adding a "pin" followed by a short description on a map, any constructional defects (e.g. slippery surface, blind spots, bad lighting, barrier on the street, potholes) that increase the risk of road accidents across the Hellenic motorways. The application is unavailable in English, meaning that its use is restricted only to Greek language speakers. The primary goal of this project was to create a spatial database with the Greek streets defects. Then, the regional volunteering groups undertake to curry out autopsies at the locations indicated in the map, in order to determine the size of the problem, take pictures, and write a detailed description. After that, a report is written and submitted to the Mayor of each Municipality. Finally, the pin on the map should be updated by changing change shape and color, in order to enable the user to know the stage of the process (however, this step was not finally implemented). Pin-Project accepted up to 650,000 visitors until May 2009, 12,421 points had been“pinned” and according to the users, 7,468 could be fixed immediately. From these points, 1,081 were characterized as dangerous for serious traffic accidents, while the majority of the pins (69%) were about the existence of dangerous potholes on the road surface (Danelli - Mylonas, 2009). The Geographical Information System (GIS) created, offers a valuable spatial database, which has so far been exploited only by the Road Safety Institute. Figure 6.1i shows the locations of the Pin-Project defects per type and seriousness in Attica, and Figure 6.1ii depicts the results of the word cloud that has been applied on the comments of the Pin-Project volunteers, making clear the problem from which the Greek road network suffers. The web page www.illegalsigns.gov.gr offers a platform where volunteers can report the existence of illegal marketing signs across the Greek street network. The volunteers indicate the exact location, attach a photo and make comments, and then are removed by the local authorities (Municipalities, Department of Transport and sometimes the office of Archaeological Athenian sites).



(i) Road defects



(ii) Word cloud

Figure 6.1: PIN Project

The aforementioned crowdsourced data can be used for the development of indicators to measure the road network quality, and should be considered when implementing an Integrated Land-Use and Transport Model.

6.5 Demographic data

6.5.1 Population sampling

The household location choice model (HLCM) is one of the main components of the LUTI models. The location of the households determines the prices, the building development and the locations of jobs. In order to estimate this model, the modeler requires either the whole census, or 1) a sample of the real population, and 2) aggregate demographic data. Using these two, the researcher can generate a synthetic population. One can obtain a sample of the population using on-line surveys. Many different applications for on-line survey are now available, such as the commercial SurveyMonkey (suveymonkey.com), the open source LimeSurvey (limesurvey.org) and the Google Forms capability within Google Documents (google docs). Despite the number of features that the paid on-line platforms offer, their high cost and the existence of free alternatives makes them less attractive. LimeSurvey and Google docs offer integrated platforms for design and analysis of the results. Furthermore, Google docs offer more than 7GB of free storage space. The main disadvantage of the Google offering is the restrictions in the number of the question types; however, most of these restrictions can be overcome through creative use of the available types. In this section is presented the design and dissemination stages of an on-line survey that was organized for the collection of the data used in Section 5.2. While the purpose of the research was to ask Greek people about their perception on car and bike sharing schemes, and their willingness to join them in the future, the procedure is recommended for the collection of household data. The questionnaire was structured in Google forms, because it offers an integrated, simple, flexible and manageable environment. Its design was based on two aspects. The first aspect was to create a questionnaire that will be addressed to everyone at an international level, and for that reason it was internet-based and composed in the English language. The second aspect was to achieve more accurate (less biased) results with the minimum cost and time. As every internet-based survey, this is vulnerable to biases, but being concerned about that they made an effort to eliminate them.

The questionnaire was disseminated via social media and mailing lists. An "event" was created on facebook and the survey was hosted there for two weeks. With this event, a group of candidate respondents is created, to whom the creator can send mass messages. This helped to keep the interest in high levels. The strategy for reminders was such that a reminder message was sent every (about) 100 completed questionnaires, so as to remind it to the others and achieve a higher response rate. The results showed that a high reply level was achieved in a few days. The use of twitter and Linkedin was also investigated.

However, they were excluded from the analysis, as they do not offer a similar tool for dissemination. In addition, the questionnaire was disseminated via e-mail to mailing lists and personal contacts. The results of the survey partially indicate the synthesis of the social media users. The majority (88.9%) of the Greek sample is between 18 and 35 years old, which is correlated with the high percentage of singles (88.5%) while only 4.5% is more than 45 years old. This is because people of younger ages are more usual to have access to the Internet and be members of social media. A recent research in Greece (www.observatory.gr), found that only 40.7% of people between 45 and 54 years old and 15.5% of those more than 55 have access to the internet, while 80.95% of those that have access belongs at the age class 18-35. Even though the percentage of those who have access has been doubled over the last four years (www.observatory.gr), it is still too low and approximately 1/3 below the average of the European Union. The percentage of those who use social media is higher at the ages between 18-34. Concerning the gender, an almost balanced 50/50 response sample was obtained (49.5% male and 50.5% female). Results for the income biases cannot be extracted, because a high percentage of households declares (even in official statements) less money than they earn. Despite the biases, this methodology offers a practical way of estimating and analyzing young people perception (about carsharing and bikesharing in that case). Different models were estimated for the age subgroup of 18-35 years old, where the sample is more compact (Section 5.2).

6.5.2 Synthetic population

Individual demographics, as well as home and work locations, are commonly collected in a census, but this data is typically not available for microsimulation due to privacy concerns. This has led to work on population synthesis, which aims to generate virtual people with the same demographics as the real population. Generally two data sets are needed for synthesis: a microsample (which typically does not contain any location information) and aggregate measures for the smallest available unit (the name of which varies by country) Williamson et al. (1998). The idea is that this synthetic population would behave similarly to the real population, allowing for its use in microsimulation and other applications without any of the data confidentiality concerns that arise when real census data is used. Synthetic population generation has applications across many fields of research Fotheringham and Rogerson (2009); Zaidi et al. (2009); Birkin and Clarke (2011).

Beckman et al. (1996a) was the first who introduced the idea to combine the census and property data to create a synthetic population. Bowman (2009) summarizes the available techniques (up to that time). Frick and Axhausen (2004) used a 5% representative sample (public used Sample, PUS) of Zurich and added socio-demographic characteristics, car availability, driving license etc. to create a synthetic population. More recently, an alternative techniques has been suggested by Farooq et al. (2013a). Synthetic population generators are available on the web (e.g. PopGen, http://cc.oulu.fi/~jaspi/popgen/popgen.htm). In Chapter 4.2 of this dissertation, a graph-theoretic approach for association generations in synthetic populations is present.

6.6 Conclusions and discussion

The development of LUTI models requires plenty of demographic, building and transport data, which increases dramatically the incurred costs. On the other hand, Web 2.0, social media and smartphones, which form the technological state-of-the-art, create many prospects in this direction. Contributing user-generated data to the web has become an easy procedure, in which the users are enjoyably involved.

In this chapter, the benefits arising from crowd-sourced data availability are investigated, and their use in every stage of development of an LUTI model is suggested. The analysis is supported with the presentation of applications that have been developed for the purposes of this dissertation. More specifically, a tool to read and store real estate price data from the web has been developed. The resulted database, together with non-residential data provided by centrally managed services, were used for the estimation of purchase and rent price modes of dwellings with great success. Moreover, on-line questionnaires to ask people about their perception to join carsharing and bikesharing schemes were constructed and disseminated, using free on-line survey tools. It is suggested that this process should be followed for the collection of demographic data and in conjunction with aggregated data, be used to generate synthetic population. Finally, mobility databases are becoming increasingly available on-line, while one could extract them indirectly from social media. Table 6.1 presents examples of potential, open, on-line sources for LUTI model data.

In conclusion, given the current economic recession, researchers should seek for new data sources. Data acquisition constitutes a large part of the research projects' pie, which could be used more effectively elsewhere. Social media are part of these opportunistic data sources. The potential data sources presented in this chapter consist a small part of the total and the aim is to activate the reader to seek for alternative sources. However, many practical limitations and concerns that need to be further examined still remain, before these data become more credible and applicable. Issues such as privacy, data handling and storage and so on need to be considered, but in a way that does not restrict the use of these valuable data sources.

Table 6.1: Open data for LUTI models

Dataset	Attributes	Examples of sources
Land Uses	Points of Interest / Non-residential buildings/ Transport infrastructure	Central Databases INSPIRE Directive Twitter + Bag-of-Words Google earth census.okfn.org tripadvisor.com foursquare.com index.pois.gr walkscore.com
	Building attributes / Property Prices	Real Estate websites
	Land cover	geo-wiki.org Google earth
Transport	Mobility data	Smartphone data Twitter International Traffic Database (ITDb) PeMs: http://pems.dot.ca.gov/ Census Transportation Planning Products: http://ctpp.transportation.org/ Microsoft research center in China foursquare.com
	Network/ Quality assessment	openstreetmaps.org seeclickfix.com illegalsigns.gov.gr msfree.gr/pin
Households/ Jobs	Sample of population / Demographic characteristics	web survey tools
	Jobs / Education level Synthetic population	linkedin.com facebook.com National statistical authorities / Census
Spatial Units	Municipalities / Zones/ Parcels / Grid	National Databases inspire-geoportal.ec.europa.eu census.okfn.org Google earth

7 Conclusion

7.1 Summary and findings

Improvements and extensions of the current Integrated Land-Use and Transport model framework have been demonstrated. Covering almost every step of the development, the suggested approaches complement the existing framework, from data collection and construction, model estimation and validation, to output interpretation and policy evaluation. The proposed additions are applicable to –disaggregate– agent-based, but also aggregate and/or activity-based LUTI models.

LUTI models are used to forecast the impact of policies on urban development. Due to the number and nature of the interacted factors in urban areas, there is significant complexity that creates many challenges in LUTI model development. Despite the progress in advancing the methodological component of the LUTI models within the last decade, is still far from reality to consider them flexible tools for policy evaluation. An additional factor that enhances this limitation, is the number of time and money required for their development and implementation.

Living in times of big data –where an increasing number of governments and private companies collect, store and manage them– using alternative data sources of –preferably– real data is essential to decrease the budget requirements and increase the forecasting accuracy of LUTI models. To a large extend, the current doctoral research has been based on public, freely available, real data, which helped to overcome the financial constraints that were faced due to the crisis in Greece.

Spatial econometric models have been developed to measure the impact of transportation infrastructure and policies on real estate purchase prices and rents, emphasising in particular to the effects of the economic crisis. The results of different spatial econometric and linear regression models are then compared, verifying that spatial models, and especially the spatial error model, solve the spatial autocorrelation and have better fit.

The results show that the impact of transportation infrastructure locations on real estate prices differs form system to system. For example, dwellings located closer to metro stations have higher prices, while to ISAP have lower.

Aiming to support the methodological improvement of synthetic population generation, a graph-theoretic approach to generate associations has been proposed, giving satisfactory results. The steps followed for the development of a complete LUTI model in UrbanSim & MATSim are presented, where the agent-based traffic simulation model (MATSim) is used for the first time jointly with UrbanSim, revealing the potential generated form their integration. Urban quality indicators such as accessibility, population segregation and density, available open space, economic viability and housing affordability, land-use and social mix, and building density, have been used integrated with LUTI model for scenario analysis and policy evaluation, manifesting the benefits of trans-disciplinary collaboration in urban planning. Urban designer, planners, analysts and policy makers, can effectively cooperate, using LUTI models for visualising the result of their strategy. Moreover, a methodology for indicator policy evaluation based on distributions rather than single aggregate measure, has been presented. Until now, the strengths of microsimulation is three dimensions, space, time and agents, were not effectively exploited. The proposed methodology detects potential trade-offs resulting from the implementation of a policy.

This doctoral research has been conducted in times of crisis, which makes some results particularly interesting. People were asked to rate how probable they believed it would be to join a potential carsharing and bikesharing scheme in Greece, when such schemes were still absent from the area. The willingness of people was modelled and the results show that those of mid income, the more environmental consciousness and the less satisfied with their current travel patterns would be more likely to join. Furthermore, the satisfaction of public transit users about public transit quality has been modelled, using hybrid choice and latent variable models. An ordered logic model shows how the demand for public transit travel has been increased within the last five years of the crisis, and which factors have contributed to this direction.

7.2 Research contributions

This research makes several contributions to the state–of–the–art of Integrated Land-Use and Transport models:

- The prospect of using crowdsourced data in LUTI model development is explored. On-line real estate data are collected and used for estimation of real estate price models. Land-use/cover data extracted from satellite images by remote sensing techniques, are used for the development of land-use/cover change model. Aiming to reduce the dependence on expensive demographic databases, a graph-theoretic approach for associations generation of synthetic population is proposed.

- A framework that integrates spatial econometric models in LUTI is proposed. In particular:
  - Real estate prices are modeled using spatial econometric techniques. These models result in better fit, comparing to the traditional Ordinary Least Squares (OLS) framework, and effectively remove the spatial autocorrelation. Aiming to measure the impact of transportation infrastructure and policies on real estate purchase prices and rents, they are applied in two case studies in Greece (Athens and Thessaloniki).
  - Land-use/cover change is modeled using discrete spatial probit, resulting in better model fit than the generalized linear regression. The impact of large-scale transportation infrastructure developments on land-use change in Athens, is examined.
  - A demand-oriented approach for spatial allocation of transportation facilities (stations of electric vehicles, carsharing etc.) based on spatial econometric models and multi-criteria analysis (MCA), is proposed.
- A methodology for qualitative and quantitative policy evaluation indicators analysis, based on the strengths of microsimulation in three dimensions (agents, space and time) is developed. Agent-based indicators for accessibility (e.g. utility based logsum), inequality (e.g. Theil index), economy/investments (e.g. price-to-rent ratio) and social quality (e.g. age mix) are investigated. The proposed methodology is based on spatial distributions instead of single aggregate measurements.
- The impact of sustainable public transport policies on urban development, using hybrid choice and latent variable models is explored.
- The effects of the financial crisis (derivative of various policies) on transportation infrastructure/policies and real estate prices are investigated.

7.3 Directions for future research

Methodological, algorithmic and application-related directions for further research are outlined in this section.

- **Spatial model simulation** Spatial econometrics should be integrated in a current LUTI model –such as UrbanSim– to be used for the real estate price model estimation, and price prediction. A concern that are could be raised about this integration is, what number of nearest neighbor points should be used for the construction of the weight matrixes? Beginning with the implementation of a simple spatial model –such as the spatial error– and then re-estimate it at repeatedly every year of simulation, using different weight matrices depending on AIC minimization, could probably be the answer. The use of price–to–rent ratio in LUTI models should be further investigated, since it could affect the location choice of the agents.

- **Further crisis impact analysis** In Sections 3.5 and 5.3 the impact of the crisis on real estate prices and public transit users' perception of quality, has been measured. However, the field is fertile for further research, since the exploration of the effects of similar extreme, unexpected situations on people's behavior and urban development, would be of great interest. From the experience that has been gained and the data that have been collected, the crisis conditions can be simulated in a LUTI model environment, and then validated, improving its forecasting accuracy. Knowing the reaction of urban areas in such cases, would be useful to deal with similar situation in the future.

- **Social welfare function** In this research, the distributions of inequality and accessibility indicators resulted from various simulations have been analyzed. However, an economic indicator that gains particular interest is the Social Welfare Function. Within the framework of the Sustaincity project, SWF has been extended and adjusted to the European cities, and has been implemented in UrbanSim. Further investigation of its disaggregate spatial distribution resulted from multiple simulations is required.

- **Psychometric indicators** The development of advanced behavioral modeling methodologies such as the hybrid discrete choice and latent variable models, create new prospects for behavioral modeling even when data related with lantent bahaviors are not available. Such behaviors should be taken into consideration at the agent location choice models, since they could affect the decision of the agent on where to be located. Estimating the models using stated/revealed preference data and then using the estimation parameters in the LUTI models simulation is recommended.

- **Land-cover data** Forecasting of future urbanization is an interesting problem that could be examined if more dense sequences of remote-sensing images were analyzed.

- **Associations generation** Matching people into other household positions is outside the scope of this work, but is planned to be considered in future research.

A List of publications

Chapter	Title	Type*
All	Efthymiou D. (2014), "Exploration of Potential Improvements for the Integrated Land-Use and transport models", *93rd Annual Meeting of the Transportation Research Board*	P
All	Efthymiou D. (2012), "Exploration of Potential Improvements for the Integrated Land-Use and transport models", *Seminar at 'Ecole Polytechnique Fédéral de Lausanne*	P
2	*All*	
3.2	Efthymiou D. and C. Antoniou (2013), "How do Transport Infrastructure and Policies Affect House Prices and Rents? Evidence from Athens", *Transportation Research Part A: Policy and Practice*, Vol. 52, pp 1–22	JA
	Efthymiou D. and C. Antoniou (2012), "How does Transport Infrastructure Affect Dwelling Prices in Athens?", *1st Symposium of the European Association for Research in Transportation (hEART).* Lausanne, September	P
3.3	Efthymiou D. and C. Antoniou (2013), "Measuring the Effects of Transportation Infrastructure Location on Real Estate Prices and Rents. Investigating the Potential Current Impact of a Planned Metro Line", *EURO Journal on Transportation and Logistics*	JA
3.4	Efthymiou D. and C. Antoniou. "Transport Policies and Property Investments: Modeling the price–to–rent ratio"	JUR
3.5	Efthymiou D. and C. Antoniou. "Investigating the impact of recession on transportation cost capitalization and housing demand: a spatial analysis"	JUR
3.6	Siora E., Argialas D., Efthymiou D. and C. Antoniou. "Modeling land-use changes and transportation interaction using object-based image analysis and spatial econometric models"	JUR
3.7	Efthymiou D., C. Antoniou and I. Tyrinopoulos (2012), "Spatially Aware Model for Optimal Site Selection. Method and Application in Mobility Center in Greece", *Transportation Research Record, Journal of the Transportation Research Board*, No. 2276, pp. 146–155	JA
	Efthymiou D., C. Antoniou, I. Tyrinopoulos and E. Mitsakis (2012), "Spatial Exploration of Effective Electric Vehicle Infrastructure Location", *In Procedia - Social and Behavioral Sciences, Proceedings of the Transportation Research Arena, Athens, Vol. 48, pp. 765–774*	CP
	Efthymiou D., C. Antoniou and I. Tyrinopoulos (2012), "Spatially-Aware Optimal Site Selection. Method and Application in Mobility Center in Greece", *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, January, Washington D.C.	CP
4.2	Anderson P., Farooq B., Efthymiou D. and M. Bierlaire (2014) "Association Generation in Synthetic Population for Transportation Applications: Graph-Theoretic Solution", *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, January, Washington, D.C.	CP
	Anderson P., Farooq B., Efthymiou D. and M. Bierlaire (2014) "Association Generation in Synthetic Population for Transportation Applications: Graph-Theoretic Solution", *Transportation Research Record, Journal of the Transportation Research Board*	JA

4.3	Cabrita I., Gayda S., Hurtubia R., Efthymiou D., Thomas I., Peeters D., Jones J., Nagel K., Nicolai T., and D. Roder, "Brussels Case Study", *SustainCity book* (to appear)	BC
	Efthymiou D., Hurtubia R., Bierlaire M. and C. Antoniou (2013), "The Integrated Land-Use and Transport Model of Brussels", *Swiss Transport Research Conference*, Ascona, April	CP
4.4	Wissen Hayek U., Efthymiou D., Farooq B., von Wirth T., Teich M., Neuenschwander N., Gret-Regamey A. "Scale Depending Evidence of Urban Quality Indicators for Fulfilling Human Needs"	JUR
	Efthymiou D. and U. Wissen. "Sustainable Urban Patterns (SUPat)", *Workshop of National Research Program NRP 65 "New Urban Quality"*. November, Zurich	W
4.5	Efthymiou D., Farooq B., Bierlaire M. and C. Antoniou (2014), "Multidimensional Indicators Analysis for Transport Policy Evaluation", *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, January, Washington, D.C.	CP
	Efthymiou D., Farooq B., Bierlaire M. and C. Antoniou (2014), "Multidimensional Indicators Analysis for Transport Policy Evaluation", *Transportation Research Record, Journal of the Transportation Research Board*	JA
	Efthymiou D., Farooq B., Bierlaire M. and C. Antoniou (2013), "A Multidimentional Indicators Analysis for Policy Evaluation", *2nd Symposium of the European Association for Research in Transportation (hEART)*, Stockholm	P
	Efthymiou D., Farooq B., Bierlaire M. and C. Antoniou (2013), "Agent-Based Indicators Analysis in the Context of Policy Evaluation: Preliminary Findings", *Workshop Land-Use Transport Interaction Models*, Paris, France	W
	Efthymiou D., Farooq B., Bierlaire M. and C. Antoniou (2013), "Agent-Based Indicators Analysis in the Context of Policy Evaluation", *Swiss Transport Research Conference*, Ascona, April	CP
	Proost S. and Van der Loo S. and Antoniou C. and D. Efthymiou (2013). "Indicators of sustainable development for microsimulation models", *SustainCity book* (to appear)	BC
5.2	Efthymiou D. and C. Antoniou (2014), "Modeling the Propensity to Join Carsharing Using Hybrid Choice and Latent Variable Models and Mixed Internet-Paper Survey Data", *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, January, Washington, D.C.	CP
	Efthymiou D., C. Antoniou and P. Waddell (2013), "Factors affecting the adoption of vehicle sharing systems by young drivers", *Transport Policy*, Vol. 29, pp. 64-73	JA
	Efthymiou D., Derevouka E. and C. Antoniou (2013), "Factors affecting the Acceptance of Carsharing Systems by Greek Commuters", *International Conference of Transportation Research*, Volos, Greece, October	CP
	Efthymiou D (2013), "Modeling the willingness to join carsharing using latent class discrete choice models and mixed internet/paper survey data", *Workshop on Discrete Choice Models*, Lausanne, Switzerland	W
	Efthymiou D., C. Antoniou and P. Waddell (2012), "Which Factors Affect Willingness to Join Vehicle Sharing Systems? Evidence from Young Greek Drivers", *Proceedings of the 91st Annual Meeting, of the Transportation Research Board*, January, Washington D.C.	CP
	Efthymiou D. and C. Antoniou (2011), "Transport Telematics Applications for Carsharing and Bikesharing Systems", *International Conference on Intelligent Transportation Systems (ITS)*. Patra, Greece	CP
5.3	Efthymiou D., Kaziales M., Antoniou C. and Y. Tyrinopoulos (2014) "Measuring the Effects of Economic Crisis on User Perception of Public Transport Quality", *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, January, Washington, D.C.	CP
	Efthymiou D., Kaziales M., Antoniou C. and Y. Tyrinopoulos (2014) "Measuring the Effects of Economic Crisis on User Perception of Public Transport Quality", *Transportation Research Record, Journal of the Transportation Research Board*	JA
6	Efthymiou D. and C. Antoniou "Leveraging Opportunistic Data for Integrated Land Use and Transport Models"	JUR
	Efthymiou D. and C. Antoniou (2012), "Leveraging Crowd-Sourced Road Defect Information for Road Quality Assessment", *Hellenic Conference in Road Safety*, Volos	CP
	Efthymiou D. and C. Antoniou (2012), "Opportunistic Transport Data Collection", *92nd Annual Meeting of the Transportation Research Board*, January 2013, Washington D.C.	P

* JA: Journal article; CP: Conference proceedings; BC: Book chapter; JUR: Journal Under Review

W: Workshop; S: Seminar; P: Presentation;

Bibliography

Al-Mosaind, M. A., Dueker, K. J., Strathman, J. G., 1993. Light-rail transit stations and property values: A hedonic price approach. Transportation Research Record: Journal of the Transportation Research Board 1400, 90–94.

Alberti, M., 1996. Measuring urban sustainability. Environmental Impact Assessment Review 16 (4), 381–424.

Alberti, M., 2005. The effects of urban patterns on ecosystem function. International Regional Science Review 28 (2), 168–192.

Albrechts, L., Healey, P., Kunzmann, K. R., 2003. Strategic spatial planning and regional governance in Europe. Journal of the American Planning Association 69 (2), 113–129.

Alonso, W., 1964. Location and Land Use. Harvard University Press, Cambridge MA.

AMETRO, 2013. Attiko Metro S.A. http://www.ametro.gr/page/default.asp?la=1&id=20, [Online; accessed 25-March-2013].

Amey, A., Attanucci, J., Mishalani, R., January 2011. Real-time ridesharing – the opportunities and challenges of utilizing mobile phone technology to improve rideshare services. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board. Washington D.C.

Anas, A., 1998. NYMTC transportation models and data initiative: The NYMTC land use model. Tech. rep., Alex Anas and Associates, Williamsville, NY.

Anas, A., Liu, Y., 2007. A regional economy, land use and transportation model (RELU-TRAN): Formulation, algorithm design, and testing. Journal of Regional Science 47 (3), 415–455.

Anderson, T. W., Darling, D. A., 1954. A test of goodness-of-fit. Journal of the American Statistical Association 49 (268), 765–769.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Springer.

Anselin, L., 1995. Local indicators of spatial association - LISA. Geographical Analysis 27 (2), 93–115.

Anselin, L., March 2010. Thirty years of spatial econometrics. Papers in Regional Science 89 (1), 3–25.

Anselin, L., Ibnu, S., Youngihn, K., 2006. GeoDa: An introduction to spatial data analysis. Geographical Analysis 38 (1), 5–22.

Anselin, L., Le Gallo, J., 2006. Interpolation of air quality measures in hedonic house price models: Spatial aspects. Spatial Economic Analysis 1 (1), 31–52.

Anselin, L., Lozano-Garcia, N., 2008. Errors in variables and spatial effects in hedonic house price models of ambient air quality. Empirical Economics 34 (1), 5–34.

Antoniou, C., Psarianos, B., Brilon, W., 2011. Induced traffic prediction inaccuracies as a source of traffic forecasting failure. Transportation Letters: The International Journal of Transportation Research 3 (4), 253–264.

Arentze, T., van den Berg, P., Timmermans, H., 2012. Modeling social networks in geographic space: approach and empirical application. Environment and Planning-Part A 44 (5), 1101–1120.

Arrow, K. J., 1953. Social Choice and Individual Values. Cowles Foundation Monographs; New York: Wiley 1964.

Atasoy, B., Glerum, A., Bierlaire, M., 2012. Attitudes towards mode choice in switzerland. disP - The Planning Review.

Athens Urban Transport Organization, 2009. Transport studies for the development of a transport master plan for the athens metropolitan area. (in Greek).

Atkinson, R., 2009. Commentary: Gentrification, segregation and the vocabulary of affluent residential choice. Urban Studies 45 (12), 2626–2636.

Baatz, M., Schape, A., 2000. Multiresolution segmentation-an optimization approach for high quality multi-scale image segmentation. In: Angewandte Geographische Informationsverarbeitung XII. Beitrge zum AGIT-Symposium.

Bae, C.-H., Jun, M.-J., Park, H., April 2003. The impact of Seoul's subway line 5 on residential property values. Transport Policy 10 (2), 85–94.

Bajari, P., Benkard, C. L., 2001. Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. Stanford University.

Bajari, P., Kahn, M. E., 2005. Estimating housing demand with an application to explaining racial segregation in cities. Journal of Business and Economic Statistics 23 (1).

Bajic, V., 1983. The efects of a new subway line on housing prices in metropolitan Toronto. Urban Studies 20 (2), 147–158.

Ballas, D., 2005. Geography matters: simulating the local impacts of national social policies. Joseph Rowntree Foundation.

Ballas, D., Clarke, G., 1999. Regional versus local multipliers of economic change? a microsimulation approach. In: 39th European Regional Science Association Congress, University College Dublin, Dublin, Ireland. pp. 23–27.

Ballas, D., Clarke, G., Dorling, D., Rossiter, D., 2007. Using SimBritain to model the geographical impact of national government policies. Geographical Analysis 39 (1), 44–77.

Ballas, D., Clarke, G., Turton, I., 1999. Exploring microsimulation methodologies for the estimation of household attributes. In: 4th International Conference on GeoComputation, Mary Washington College, Virginia, USA.

Baltagi, B. H., Egger, P., Pfaffermayr, M., 2008. Estimating regional trade agreement effects on FDI in an interdependent world. Journal of Econometrics 145 (1-2), 194–208.

Banister, D., Berechman, Y., 2001. Transport investment and the promotion of economic growth. Journal of Transport Geography 9 (3), 209–218.

Bank of Greece, 2012a. Results of real estate agents research. Economic Research Department, Real Estate Market Analysis Section.

Bank of Greece, 2012b. Short-term indicators for the real estate market. Economic Research Department, Real Estate Market Analysis Section.

Baradaran, S., Ramjerdi, F., 2001. Performance of accessibility measures in Europe. Journal of Transportation and Statistics 4 (2-3), 31–48.

Baraldi, A., Puzzolo, V., Blonda, P., Bruzzone, L., Tarantino, C., 2006. Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images. IEEE, Transactions on Geoscience and Remote Sensing 44 (9), 2563–2586.

Barth, M., Shaheen, S., 2002. Shared vehicle systems: framework for classifying carsharing, station cars, and combined approaches. Transportation Research Record: Journal of the Transportation Research Board 1791, 105–112.

Barthelemy, J., Toint, P. L., 2013. Synthetic population generation without a sample. Transportation Science 47 (2), 266–279.

Bartik, T. J., 1987. The estimation of demand parameters in hedonic price models. The Journal of Political Economy 95 (1), 81–88.

Bates, J., 2003. Economic evaluation and transport modelling: theory and practice. In: 10th International Conference on Travel Behaviour Research. Luzern.

Beckman, R., Baggerly, K., Keith, A., McKay, M. D., 1996a. Creating synthetic baseline populations. Transportation Research Part A: Policy and Practice 30 (6), 415–429.

Beckman, R. J., Baggerly, K. A., McKay, M. D., 1996b. Creating synthetic baseline populations. Transportation Research Part A: Policy and Practice 30 (6), 415–429.

Ben-Akiva, M., de Palma, A., McFadden, D., Abou-Zeid, M., Chiappori, P.-A., de Lapparent, M., Durlauf, S. N., Fosgerau, M., Fukuda, D., Hess, S., Manski, C., Pakes, A., Picard, N., Walker, J., 2012. Process and context in choice models. Marketing Letters 23 (2), 439–456.

Ben-Akiva, M., Lerman, S. R., 1985. Discrete Choice Analysis. MIT Press, Cambridge MA.

Ben-Akiva, M., Morikawa, T., 1990. Estimation of travel demand models from multiple data sources. In: Koshi, M. (Ed.), 11th International Symposium on Transportation and Traffic Theory. Yokohama, Japan.

Benjamin, J. D., Sirmans, G. S., 1996. Mass transportation, apartment rent and property values. Journal of Real Estate Research 12 (1), 1–8.

Benser, C. A., 2002. Spatial autoregressive specification with a comparable sales weighting scheme. Journal of Real Estate Research 24 (2), 193–211.

Benz, U., Hoffman, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. Journal of Photogrammetry and Remote Sensing 58 (3), 239–258.

Bhat, C. R., Guo, J. Y., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. Transportation Research Part B 38 (2), 147–168.

Bierlaire, M., 2003. BIOGEME: A free package for the estimation of discrete choice models. In: Proceedings of the 3rd Swiss Transportation Research Conference. Ascona, Switzerland.

Bierlaire, M., Cheng, J., Newman, J. P., 2013. A probabilistic map matching method for smartphone GPS data. Transportation Research Part C: Emerging Technologies 26, 78–98.

Bierlaire, M., Fetiarison, M., 2009. Estimation of discrete choice models: extending biogeme. In: 9th Swiss Transport Research Conference. Ascona, Switzerland.

Birkin, M., Clarke, M., 2011. Spatial microsimulation models: A review and a glimpse into the future. In: Population Dynamics and Projection Methods. Springer, pp. 193–208.

Bitter, C., Mulligan, G., Dall'erba, S., 2007. Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. Journal of Geographical Systems 9 (1), 7–27.

Bivard, R. S., Pedesma, E. J., Gomez, R. V., 2008. Applied Spatial Data Analysis with R. Use R! Series.

Bjorn, A., 2009. Essays on examining the impacts of forest cover on housing prices using bayesian model averging and geographically weighted regression. Ph.D. thesis, University of Washington.

Boarnet, M. G., Chalempong, S., 2001. New highways, house prices, and urban development: A case study of toll roads in Orange Country, CA. Housing Policy Debate 12 (3), 575–605.

Bowes, D., Ihlanfeldt, K., 2001. Identifying the impacts of rail transit stations on residential property values. Journal of Urban Economics 50 (1), 1–25.

Bowman, J. L., 2009. A comparison of population synthesizers used in microsimulation models of activity and travel demand. Unpublished working paper.

Box, G. E. P., Cox, D. R., 1946. An analysis of transformations. Journal of the Royal Statistical Society. Series B (methodological) 26 (2), 211–252.

Brandeau, M. L., Chiu, S. S., 1989. An overview of representative problems in location research. Management Science 35 (6), 645–674.

Brandy, M., Irwin, E., 2011. Accounting for spatial effects in economic models of land use: Recent developments and challenges ahead. Environmental and Resource Economics 48 (3), 487–509.

Brasington, D. M., Hite, D., 2005. Demand for environmental quality: A spatial hedonic analysis. Regional Science and Urban Economics 35 (1), 57–82.

Bregman, S., January 2011. What's the worst that can happen? How to stop worrying and love social media. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board. Washington D.C.

Brennan, T. P., Cannaday, R., Colwell, P. F., 1984. Office rent in the Chicago CBD. Real Estate Economics 12 (3), 243–260.

Brons, M., Nijkamp, P., Pels, E., Rietveld, P., 2003. Railroad noice: Economic valuation and policy. Transportation Research: Part D: Transport and Environment 8 (3), 169–184.

Brown, A. L., 2003. Increasing the utility of urban environmental quality information. Landscape and Urban Planning 65 (1), 85–93.

Brown, J. N., Rosen, H., 1982. On the estimation of structural hedonic price models. Econometrica 50 (3), 765–768.

Brown, M. B., Forsythe, A. B., 1974. Robust tests for the equality of variances. Journal of the American Statistical Association 69 (346), 364–267.

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression-modeling spatial non-stationarity. The Statistician 47 (3), 341–443.

Burkard, R. E., Dell'Amico, M., Martello, S., 2009. Assignment problems. Siam.

Burkhardt, J., Millard-Ball, A., 2006. Who's attracted to car-sharing? Transportation Research Record: Journal of the Transportation Research Board 1986, 98–105.

Büttner, J., 2010. Planning a new bike sharing scheme. making decisions based on previous experience. In: Proceedings of the Velo-City Global. Copenhagen.

Cameron, S., Field, A., 2000. Community, ethnicity and neighbourhood. Housing Studies 15 (6), 826–843.

Candau, J., Rasmussen, S., Clarke, K. C., 2002. A coupled cellular automaton model for land use/land cover dynamics. In: Proceedings of the 4th International Conference on Integrating GIS and Environmental Modeling. Banff, Canada.

Carvalho, S., Sarmento, L., Rosetti, R., September 2010. Real-time sensing of traffic information in twitter messages. In: Proceedings of the 4th Workshop on Artificial Transportation Systems and Simulation ATSS at IEEE ITSC 2010. madeira, Portugal.

Carver, S., 1991. Integrating multi-criteria evaluation with geographical information systems. International Journal of Geographical Information Science 5 (3), 321–339.

Cervero, R., Duncan, M., 2002. Transit's value-added effects: Light and commuter rail services and commercial land values. Transportation Research Record: Journal of the Transportation Research Board 1805, 8–15.

Cervero, R., Landis, J., 1993. Assessing the impacts of urban rail transit on local real estate markets using quasi-exparimental compatisons. Transportation Research Part A: Policy and Practice 27 (1), 13–22.

Charlton, M., Fotheringham, A. S., March 3 2009. Geographically weighted regression white paper.

Charypar, D., Nagel, K., 2005. Generating complete all-day activity plans with generic algorithms. Transportation 32 (4), 369–397.

Chen, H., Rufolo, A., Dueker, K. J., July 1997. Measuring the impact of light rail systems on single family home values: A hedonic approach with GIS application. Discusssion paper 97-3, Portland State University.

Cherchi, E., Polak, J., Hyman, G., 2004. The impact of income, tastes and substitution effects on the assessment of user benefits using discrete choice models. In: European Transport Conference. Strasbourg.

Chhetri, P., Han, J. H., Chandra, S., Corcoran, J., 2013. Mapping urban residential density patterns: Compact city model in Melbourne, Australia. City, Culture and Society 4, 77–85.

Cho, S.-H., Bowker, J. M., Park, W. M., 2006. Measuring the contribution of water and green space amenities to housing values: An application and comparison of spatially weighted hedonic models. Journal of Agricultural and Resource Economics 31 (3), 485–507.

Chorianopoulos, I., Pagonis, T., Koukoulas, S., Drymoniti, S., 2010. Planning, competitiveness and sprawl in the Mediterranean city: The case of Athens. Cities 4, 249–259.

Chow, L.-F., Chi, H., Zhao, F., Chen, Z., January 2010. Subregional transit ridership models based on geographically weighted regression. In: Proceedings of the 89th Annual Meeting of the Transportation Research Board. Transportation Research Board, Washington D.C.

Church, 2002. Geographical information systems and location science. Computers and Operations Research 29 (6), 541–562.

Ciari, F., Balmer, M., W., A. K., January 2009. Concepts for large scale car-sharing system: modelling and evaluation with an agent-based approach. In: Proceedings of the 88th Annual Meeting of the Transportation Research Board. Washington D.C.

Ciari, F., Löchl, M., Axhausen, K. W., July 2008. Location choice of retailers - an agent-based approach. In: 15th International Conference on Recent Advances in Retailing and Services Science. Zagreb.

Cirilli, C., Eboli, L., Mazzulla, G., 2011. On the asymetric user perception of transit service quality. International Journal of Transportation 5 (4), 216–232.

Clarke, A., 2012. Housing need in the United Kingdom. International Encyclopedia of Housing and Home, 538–543.

Clower, T. L., Weinstein, B. L., 2002. The impact of Dallas (Texas) area rapid transit light rail stations on taxable property valuations. Australasian Journal of Regional Studies 8 (3), 389–402.

Cohen, J., Coughlin, C., 2009. Changing noice levels and housing prices near the Atlanda airport. Growth and Change 40 (2), 287–313.

Cohen, J. P., Coughlin, C. C., 2008. Spatial hedonic models of airport noise, proximity, and housing prices. Journal of Regional Science 48 (5), 859–878.

Cohen, J. P., Paul, C. M., 2007. The impacts of transportation infrustructure on property values: A higher-order spatial econometrics approach. Journal of Regional Science 47 (3), 457–478.

Collins, A., Evans, A., May 1996. Aircraft noise and residential property values: An artificial neural network approach. Journal of Transport Economics and Policy 28 (2), 175–197.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E., 2004. Digital change detection methods in ecosystem monitoring: a review. International Journal of Remote Sensing 25 (9), 1565–1596.

Cortright, J., 2009. Walking the walk: How walkability raises home values in U.S. cities. Tech. rep., Impresa, Inc.

Coughlin, P., Nitzan, S., 1981. Electoral outcomes with probabilistic voting and Nash social welfare maxima. Journal of Public Economics 15 (1), 113–121.

Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J., 2013. # Earthquake: Twitter as a distributed sensor system. Transactions in GIS.

Crowley, R. W., 1973. A case study of the effects of and airport on land values. Journal of Transport Economics and Policy 7 (2), 144–152.

Current, J., Min, H., Schilling, D., 1990. Multiobjective analysis of facility location decisions. European Journal of Operational Research 49 (3), 295–307.

Curtis, C., March 2011. Integrating land use with public transport: The use of a discursive accessibility tool to inform metropolitan spatial planning in Perth. Transport Reviews 31 (2), 179–197.

Damn, D., Lerman, S. R., Lerner-Lam, E., Young, J., September 1980. Response of urban real estate values in anticipation of the washington metro. Journal of Transport Economics and Policy 14 (3), 315–336.

Danelli - Mylonas, 2009. Road safety: Actions and perspectives. In: Swedish Trade Council Conference.

Davis, C., Jha, M. K., 2011. A dynamic modeling approach to investigate impacts to protected and low-income populations in highway planning. Transportation Research Part A: Policy and Practice 45 (7), 598–610.

De Jong, G., Pieters, M., Daly, A., Graafland, I., Kroes, E., Koopmans, C., May 2005. Using the logsim as an evaluation measure: Literature and case study. Tech. rep., RAND Corporation.

De Jong, P., Sprenger, C., van Veen, F., 1984. On extreme values of Moran's I and Geary's c. Geographic Analysis 16 (1), 17–24.

De Palma, A., Marchal, F., Nesterov, Y., 1997. METROPOLIS: A modular system for dynamic traffic simulation. Transportation Research Record: Journal of the Transportation Research Board 1607, 178–184.

De Palma, A., Picard, N., 2005. Route choice decision under travel time uncertainty. Transportation Research Part A: Policy and Practice 39 (4), 295–324.

De Palma, A., Proost, S., van der Loo, S., 2010. Assessing transport investments – Towards a multi-purpose tool. Transportation Research Part B 44 (De Palma, A., Proost, S., van der Loo, S., 2010. Assessing transport investments - to- wards a multi-purpose tool. Transportation Research Part B 44, 834–849. 7), 834–849.

Debrezion, G., Pels, E., Rietveld, P., 2007. The impact of railway stations on residential and commercial property value: A meta-analysis. Journal of Real Estate and Financial Economics 35 (2), 161–180.

Debrezion, G., Pels, E., Rietveld, R., March 2006. The impact of rail transport on the real estate prices: An empirical analysis of the Dutch housing market. Tinbergen institute discussion paper, Tinbergen Institute, Free University, Department of Spatial Economics.

Definiens, A. G., 2006. Definiens Professional 5 - User Guide (Document Version 5.0.6.2) and Reference Book (Document Version 5.0.6.1).

DeMaio, P., 2009. Bike-sharing: its history, models of provision, and future. Journal of Public Transportation 12 (4), 41–56.

Deming, W. E., Stephan, F. F., 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. The Annals of Mathematical Statistics 11 (4), 427–444.

Dempsey, N., Brown, C., Bramley, G., 2012. The key to sustainable urban development in UK cities? the influence of density on social sustainability. Progress in Planning 77 (3), 89–141.

Densham, P. J., Ruston, G., 1992. A more efficient heuristic for solving large p-median problems. Papers in Regional Science 71 (3), 307–329.

Derebouka, E., 2012. Investigation of greek travellers perception for emerging transport systems using econometric models. Master's thesis, Rural and Surveying Engineering, National Technical University of Athens (in Greek).

Dillman, D., Smyth, J. D., Christian, M., 2009. Internet, Mail and Mixed-Mode Surveys: The Tailired Design Method. John Wiley and Sons, Inc., New York.

DiPasquale, D., Wheaton, W. C., 1996. Urban economics and real estate markets. Prentice Hall.

Doan, A., Ramakrishnan, R., Halevy, Y., 2011. Crowdsourcing systems on the world-wide web. Communications of the ACM 54 (4), 86–96.

Dobson, J. E., 1979. A regional screening procedure for land use suitability analysis. Geographical Review 69 (2), 224–234.

Dorantes, L. M., Paez, A., Vassallo, J. M., 2011. Analysis of house prices to assess the economic impacts of new public transport infrastructure: Madrid metro line 12. Transportation Research Record: Journal of the Transportation Research Board 2245.

Dray, S., Legendre, P., Peres-Neto, P. R., 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (pgnm). Ecological Modelling 196 (3), 483–493.

Dufour, D., Ligtermoet and Partners, 2010. Presto cycling policy guide general framework, intelligent energy. Tech. rep., Europe.

Dugundji, E. R., Walker, J. L., 2005. Discrete choice with social and spatial network interdependencies: An empirical example using mixed generalized extreme value models with field and panel effects. Transportation Research Record: Journal of the Transportation Research Board 1921 (1), 70–78.

Duncan, T. L., 2014a. RCurl: General network (HTTP/FTP/...) client interface for R. http://CRAN.R-project.org/package=RCurl. R package version 1.91-1.

Duncan, T. L., 2014b. XML: Tools for parsing and generating XML within R and S-Plus. http://CRAN.R-project.org/package=XML. R package version 3.9-4.

Eboli, L., Mazzulla, G., 2011. A methodology for evaluating transit service quality based on subjective and objective measures from the passengers point of view. Transport Policy 18 (1), 172–181.

Echenique, M. H., 1985. The use of integrated land use and transport models: the cases of sao paulo, the use of integrated land use and transport models: The cases of sao paulo, brazil and bilbao, spain. The Practice of Transportation Planning, M. Florian (ed.), The Hague: Elsevier, 263–286.

Edwards, K. L., Clarke, G. P., 2009. The design and validation of a spatial microsimulation model of obesogenic environments for children in Leeds, UK: SimObesity. Social Science & Medicine 69 (7), 1127–1134.

Efthymiou, D., Antoniou, C., 2013a. How do transport infrastructure and policies affect house prices and rents? evidence from Athens, Greece. Transportation Research Part A: Policy and Practice 52, 1–22.

Efthymiou, D., Antoniou, C., 2013b. Use of social media for transport data collection. In: Procedia - Social and Behavioral Sciences, Proceedings of the Transportation Research Arena. Vol. 48. Athens, pp. 775–785.

Elbers, C., lanjouw, P., Mistiaen, J. A., Ozer, B., 2005. Re-interpreting sub-group inequality decompositions. Tech. Rep. 3687, World Bank Policy Research Working Paper.

Elhorst, J. P., 2003. Specification and estimation of spatial panel data models. International Regional Science Review 26 (3), 244–268.

Epple, D., Sieg, H., 1999. Estimating equilibrium models of local jurisdictions. Journal of Political Economy 107 (4), 645–681.

Espay, M., Lopez, H., Summer 2000. The impact of airport noise and proximity on residential property values. Growth and Change 31 (3), 408–419.

Espino, R., Roman, C., Ortuzar, J., 2006. Analyzing demand for suburban trips: A mixed RP/SP model with latent variables and interaction effects. Transportation 33 (3), 241–261.

ESRI, 2011. ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands, CA.

Estellés - Arolas, E., González Ladrón-de-Guevara, F., 2012. Towards an integrated crowdsourcing definition. Journal of Information Science 38 (2), 189–200.

European Committee for Standardization , 2002. Transportation - logistics and services - public passenger transport - service quality definition, targeting and measurement. Tech. rep., CEN.

European Environment Agency, . E., 2010. Tracking progress towards kyoto and 2020 targets in europe. Tech. Rep. 7, EEA.

European Performance Satisfaction Index Rating Institute, 2005. Customer satisfaction index. Tech. rep., EPSI.

European Road Safety Observatory, 2007. Traffic safety basic facts: Motorcycles and mopeds. Tech. rep., ERSO.

Farooq, B., Bierlaire, M., Hurtubia, R., Flötteröd, G., 2013a. Simulation based population synthesis. Transportation Research Part B: Methodological.

Farooq, B., Miller, E., 2012. Towards integrated land use and transportation: A dynamic disequilibrium based microsimulation framework for built space markets. Transportation Research Part A: Policy and Practice 46 (7), 1030–1053.

Farooq, B., Miller, E. J., Chingcuanco, F., Cook, M., 2013b. Microsimulation framework for urban price-taker markets. Journal of Transport and Land Use 6 (1), 41–51.

Farooq, B., Miller, E. J., Haider, M., 2010. Hedonic analysis of office space rent. Transportation Research Record: Journal of the Transportation Research Board 2174, 118–127.

Feitelson, E., 1989. Transportation noise, property rights, and institutional structure: The Israeli experience in perspective. Transportation Research Part A: Policy and Practice 23 (5), 349–358.

Fisher, K., Egerton, M., Gershuny, J. I., Robinson, J. P., 2007. Gender convergence in the american heritage time use study (AHTUS). Social Indicators Research 82 (1), 1–33.

Fleisher, A., 1965. Review of a model of metropolis by Ira S. Lowry. Journal of the American Institute of Planners 31, 175–6.

Forkenbrock, D. J., 2001. Comparison of external costs of rail and truck freight transportation. Transportation Research Part A: Policy and Practice 35 (4), 321–337.

Fotheringham, A. S., Charlton, M. E., Brunson, C., 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. Environment and Planning A 30 (11), 1905–1927.

Fotheringham, A. S., Rogerson, P. A., 2009. The SAGE handbook of spatial analysis. SAGE Publications Limited.

Franklin, J. P., August 2006. The equity effects of roadway tolls: An application of hicksian wealfare meassure with income effects. In: The Expanding Sphere of Travel Behaviour Research, 11th International Conference on Travel Behaviour Research. Kyoto.

Franklin, J. P., Waddell, P., January 2003. A hedonic regression of home prices in king county, washington, using activity-specific accessibility measures. In: Proceedings of the 82nd Annual Meeting of the Transportation Research Board. Washington D.C.

Frazier, C., Kockelman, K. M., 2005. Spatial econometric models for panel data: Incorporating spatial and temporal data. Transportation Research Record: Journal of Transportation Research Board 1902, 80–90.

Frick, M., Axhausen, K., 2004. Generating synthetic populations using ipf and monte carlo techniques: Some new results. In: Swiss Transport Research Conference. Ascona, Switzerland.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F., Obrsteiner, M., 2009. Geo-wiki.org: The use of crowdsourcing to improve global land cover. Remote Sensing.

Gargiulo, F., Ternes, S., Huet, S., Deffuant, G., 2010. An iterative approach for generating statistically realistic populations of households. PloS one 5 (1), e8828.

Getis, A., Griffith, D. A., 2002. Comparative spatial filtering in regression analysis. Geographical analysis 34 (2), 130–140.

Getis, A., Ord, J. K., 1992. The analysis of spatial association by use of distance statistics. Geographical Analysis 24 (3), 189–206.

Geurs, K. T., Wee, B., 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. Journal of Transport Geography 12 (2), 127–140.

Giaoutzi, M., Damianides, L., 1990. Transport Policy and the Environment: Six Case Studies. Earthscan, London, Ch. The Greek transport system and environment.

Goldner, W., 1971. The Lowry model heritage. Journal of the American Institute of Planners 37 (2), 100–110.

Golob, T. F., 2003. Structural equation modeling for travel behavior research. Transportation Research Part B: Methodological 37 (1), 1–25.

Goodchild, M., Glennon, J. A., 2010. Crowdsourcing geographical information for disaster response: A research frontier. International Journal of Digital Earth 3 (3), 231–241.

Gordon, A., Simondson, D., White, M., Moilanen, A., Bekessy, S., 2009. Integrating conservation planning and landuse planning in urban areas. Landscape and Urban Planning 91 (4), 183–194.

Grazi, F., van den Bergh, J., Rietveld, P., 2007. Spatial welfare economics versus ecological footprint: modeling agglomeration, externalities and trade. Environmental and Resource Economics 38 (1), 135–153.

Griffith, D. A., 2008. Spatial-filtering based contributions to a critique of geographically weighted regression (GWR). Environment and Planning A 40 (11), 2751–2769.

Grün Stadt Zürich, 2006. Das grünbuch der stadt zürich. Tech. rep., Grün Stadt Zürich.

Gutiérrez, J., Urbano, P., 1996. Accessibility in the European Union: the impact of the trans-European road network. Journal of Transport Geography 4 (1), 15–25.

Haider, M., Miller, E. J., 2000. Effects of transportation infrastructure and location on residential real estate values: Application of spatial autoregressive techniques. Transportation Research Record: Journal of the Transportation Research Board 1722, 1–18.

Hakimi, S. L., 1964. Optimum locations of switching centers and the absolute centers and medians of a graph. Operations Research 12 (3), 450–459.

Halkias, B., Tyrogianni, H., 2008. PPP projects in Greece: The case of Attica Tollway.

Hansen, H. S., 2005. GIS multi-criteria analysis of wind farm development. In: ScanGis 2005 : Proceedings of the 10th Scandinavian Research Conference on Geographical Information Science. pp. 75–87.

Harding, A., Lloyd, R., Bill, A., King, A., 2004. Assessing poverty and inequality at a detailed regional level: New advances in spatial microsimulation. No. 26. Research Paper, UNU-WIDER, United Nations University (UNU).

Harland, K., Heppenstall, A., Smith, D., Birkin, M., 2012. Creating realistic synthetic populations at varying spatial scales: a comparative critique of population synthesis techniques. Journal of Artifical Societies and Social Simulation 15 (1), 1–15.

Healey, P., 2007. Urban complexity and spatial strategies. Towards a relational planning for our times. Routledge, Taylor & Francis, London and New York.

Hellenic Institute of Transport, 2003. Handbook for the implementation of the quality control system of oasa. stage 2 final report of the project: An integrated system for the quality assessment of the oasa passenger services. Tech. rep., HIT.

Hellenic Institute of Transport Engineers, December 2008. Roadmap for mobility in athens within the next decade: from congestion to sustainable mobility.

Hellenic Ministry of Public Works, 2009. Strategic engironmental impacts study and update of the regulatory master plan for athens. in Greek.

Hellenic Statistical Authority, 2011. Announcement of temporary results of census. Tech. rep., EL.STAT.

Hendriks, H., 2010. Idea contest: a method to capture the interest of professionals, local decision makers and to start the public discussion. In: Proceedings of the Velo-City Global. Copenhagen.

Henneberry, J., 1998. Transport investment and house prices. Journal of Property Valuation and Investment 16 (2), 144–158.

Hermes, K., Poulsen, M., 2012. A review of current methods to generate synthetic spatial microdata using reweighting and future directions. Computers, Environment and Urban Systems 36 (4), 281–290.

Hess, D. B., Almeida, T. M., May 2007. Impact of proximity to light rail rapid transit on station-area property values in buffalo, new york. Urban Studies 44 (5-6), 1041–1068.

Hopcroft, J. E., Karp, R. M., 1973. An n^5/2 algorithm for maximum matchings in bipartite graphs. SIAM Journal on computing 2 (4), 225–231.

Horowitz, J. L., 1992. The role of the list price in housing markets: Theory and an econometric model. Journal of Applied Econometrics 7 (2), 115–129.

Hoshino, T., Kuriyama, K., 2009. Measuring the benefits of neighbourhood park amenities: Application and comparison of spatial hedonic approaches. Environmental and Resource Economics 45 (3), 429–444.

Howe, 2006. The rise of crowdsourcing, Wired 14.06. http://crowdsourcing.typepad.com/.

Huang, Z., Williamson, P., 2001. A comparison of synthetic reconstruction and combinatorial optimisation approaches to the creation of small-area microdata. Department of Geography, University of Liverpool.

Hubert, J. P., Toint, P. L., 2002. La Mobilité Quotidienne des Belges. Press Universitaires de Namur.

Hunt, J. D., Kriger, D. S., Miller, E. J., May 2005. Current operational urban land-use-transport modelling frameworks: A review. Transport Reviews 25 (3), 329–376.

Hurtubia, R., Bierlaire, M., 2012. Estimation of bid functions for location choice and price modeling with a latent variable approach. Networks and Spatial Economics, 1–19.

Hurtubia, R., Bierlaire, M., Martinez, F., 2012. Dynamic microsimulation of location choices with a quasi-equilibrium auction approach. In: Proceedings of the 12th Swiss Transport Research Conference. Ascona, Switzerland.

Hynes, S., Morrissey, K., O'Donoghue, C., Clarke, G., 2009. Building a static farm level spatial microsimulation model for rural development and agricultural policy analysis in ireland. International Journal of Agricultural Resources, Governance and Ecology 8 (2), 282–299.

Ibeas, Á., Vordera, R., dell'Olio, L., Coppola, P., Dominiguez, A., 2012. Modelling transport and real-estate values interactions in urban systems. Journal of Transport Geography 24, 370–382.

INSPIRE, 2007. Directive 2007/2/ec of the European Parliament and of the council of 14 march 2007 establishing an infrastructure for spatial information in the European community (inspire). Tech. rep., The European Parliament and the Council of the European Union.

Ion, L., Curu, T., Boussier, J. M., Teng, F., Breuil, D., 2009. Site selection for electric cars of a car-sharing service. In: Proceedings of EVS24: International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium & Exhibition.

Jenks, M., Jones, C. (Eds.), 2010a. Dimensions of the Sustainable City. Future City 2. Springer, Doderecht, Heidelberg, London, New York, Ch. Issues and Concepts, pp. 1–19.

Jenks, M., Jones, C. (Eds.), 2010b. Dimensions of the Sustainable City. Future City 2. Springer, Doderecht, Heidelberg, London, New York, Ch. Complementaries and Contradictions, pp. 239–256.

Jiang, Y. G., Ngo, C. W., Yang, J., 2007. Towards optimal bag-of-features for object categorization and semantic video retrieval. In: 6th ACM International Conference on Image and Video Retrieval.

Johnson, R., Wickern, D., 1992. Applied Multivariate Statistical Analysis, 3rd Edition. Prentice- Hall, Englewood Cliffs, NJ.

Jones, P. M., Dix, M. C., Clarke, M. I., Heggie, I. G., 1983. Understanding travel behaviour. Monograph.

Jonsson, D., Berglund, S., Almstrom, P., Algers, S., 2011. The usefulness of transport models in Swedish planning practice. Transport Reviews 31 (2), 251–265.

Kakamu, K., Wago, H., 2005. Bayesian spatial panel probit model with an application to business cycle in Japan. In: International Congress on Modelling and Simulation.

Kalmanje, S., Kockelman, K. M., 2004. Credit-based congestion pricing: Travel, land value and welfare impacts. Transportation Research Record: Journal of the Transportation Research Board 1864, 45–53.

Kapoor, M., Kelejiam, H. H., Prucha, I. R., 2007. Panel data models with spatially correlated error components. Journal of Econometrics 140 (1), 97–130.

Karanikolas, N., Anastasiadou, E., April 2012. The effect of location of metro stations on real estate values of commercial properties. a case study of Thessaloniki, greece. Journal of Economics and Engineering 3 (1), 4–11.

Karathanassi, V., Iossifidis, C., Rokos, D., 2000. A texture-based classification method for classifying built areas according to their density. International Journal of Remote Sensing 21 (9), 1807–1823.

Karlström, A., 2000. Non-linear value functions in random utility econometrics. In: International Conference on Travel Behaviour Research, Gold Coast, Australia.

Kawamura, K., Mahajan, S., 2006. Hedonic analysis of impacts of traffic volumes on property values. Transportation Research Record: Journal of the Transportation Research Board 1924, 69–75.

Keddi, J., Tonkiss, F., 2010. The market and the plan: Housing, urban renewal and socio-economic change in London. City, Culture and Society 1, 57–67.

Kelejian, H., Prucha, I., May 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40 (2), 509–533.

Kemel, E., Yoon, S. Y., Goulias, K. G., January 2008. An inequality assessment tool for urban development using gis-based accessibility measures and a fractal version of Theil's index. In: Proceedings of the 87th Annual Meeting of the Transportation Research Board. Washington D.C.

Kim, C. W., Philips, T. T., Anselin, L., 2003. Measuring the benefits of air quality improvement: a spatial hedonic approach. Journal of Environmental Economics and Management 45, 24–39.

Kissling, W. D., Gudrun, C., 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. Journal of Global Ecology and Biogeography 17 (1), 59–71.

Kitamura, R., Pas, E. I., Lula, C. V., Lawton, T. K., Benson, P. E., 1996. The sequenced activity mobility simulator (SAMS): An integrated approach to modeling transportation, land use and air quality. Transportation 23 (3), 267–291.

Klaiber, H. A., Phaneuf, D. J., 2010. Valuing open space in a residential sorting model of the twin cities. Journal of Environmental Economics and Management 60 (2), 55–57.

Klier, T., McMillen, D. P., 2008. Clustering of auto supplier plants in the United States. Journal of Business and Economic Statistics 26 (4), 460–471.

Knowles, R. D., 2006. Transport shaping space: differential collapse in time-space. Journal of Transport Geography 14 (6), 407–425.

Kockelman, K. M., 1990. The effects of location elements on home purchase prices and rents: Evidence from the san francisco bay area. Transportation Research Record: Journal of the Transportation Research Board 1606, 40–50.

Kraier, J., Chishen, W., 2004. House prices and fundamental value. Tech. rep., Federal Reserve Bank of San Francisco.

Kruskal, Wallis, 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 47 (260), 583–621.

Kuhn, H. W., 1955. The hungarian method for the assignment problem. Naval Research Logistics Quarterly 2 (1-2), 83–97.

Lancaster, K., 1966. A new approach to consumer theory. Journal of Political Economy 74 (2), 132–157.

Lane, C., 2005. Phillycarshare: First-year social and mobility impacts of carsharing in philadelphia, pennsylvania. Transportation Research Record: Journal of the Transportation Research Board 1927, 158–166.

Langley, J. C., 1981. Highways and property values: The Washington beltway revised. Transportation Research Record: Journal of the Transportation Research Board 812, 16–21.

Leggett, C. G., Bockstael, N. E., 2000. Evidence of the effects of water quality on residential land prices. Journal of Environmental Economics and Management 39 (2), 121–144.

Lenormand, M., Deffuant, G., 2012. Generating a synthetic population of individuals in households: Sample-free vs sample-based methods. arXiv preprint arXiv:1208.6403.

LeSage, J., Pace, R. K., 2009. Introduction to Spatial Econometrics. CRC Press, Taylor and Francis Group.

Leung, D., Newsam, S., 2012. Exploring geotagged images for land-use classification. In: Proceedings of the ACM multimedia 2012 workshop on Geotagging and its applications in multimedia. pp. 3–8.

Leung, Y., Mei, C.-L., Zhang, W.-X., 2000. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. Environment and Planning A 32 (1), 9–32.

Levinson, D., 2010. Equity effects of road pricing: A review. Transport Reviews 30 (1), 33–57.

Li, H., Calder, C. A., Cressie, N. A. C., 2007. Beyond Moran's i: testing for spatial dependence based on the spatial autoregressive model. Geographical Analysis 39 (4), 357–375.

Likert, R., 1932. A technique for the measurement of attitudes. Archives of Psychologie.

Linneker, B., Spence, N., 1996. Road transport infrastructure and regional economic development: The regional development effects of the M25 london orbital motorway. Journal of Transport Geography 4 (2), 77–92.

Lipscomb, C., 2003. Small cities matter, too: The impacts of an airport and local infrastructure on housing prices in a small urban city. Review of Urban & Regional Development Studies 15 (3), 255–273.

Litman, T., 2011a. Developing indicators for sustainable and livable transport planning. Victoria Transport Policy Institute.

Litman, T. A., 2011b. Pay-as-you-drive pricing for insurance affordability. Tech. Rep. 8, Victoria Transport Policy Institute.

Löchl, M., Axhausen, K., 2010. Modeling hedonic residential rents for land use and transport simulation while considering spatial effects. Journal of Transport and Land Use 2 (2), 39–63.

Löchl, M., Bürgle, M., Axhausen, K. W., 2007. Implementierung des integrierten flächennutzungsmodells urbansim für den grossraum zürich – ein erfahrungsbericht. Tech. rep., ETH.

Loomis, J. B., 2011. What's to know about hypothetical bias in stated preference valuation studies? Journal of Economic Surveys 25 (2), 363–370.

Louviere, J. J., Meyer, R. J., Bunch, D. S., Carson, R., Dellaert, B., Hanemann, W. M., Hensher, D. A., Irwin, J., 1999. Combining sources of preference data for modeling complex decision processes. Marketing Letters 10 (3), 205–217.

Lowry, I. S., August 1964. A model of metropolis. Memorandum RM-4035-RC, RAND Corporation, 1700 Main St. Santa Monica, California.

Lozani-Gracia, N., Anselin, L., 2011. Is the price right? Assessing estimates of cadastral values for Bogota, Colombia. Tech. rep., Arizona State University.

Lu, D., Mausel, P., Brondizios, E., Moran, E., 2004. Change detection techniques. International Journal of Remote Sensing 25 (12), 2365–2407.

Luce, R. D., 1959. Individual Choice Behavior a Theoretical Analysis. John Wiley and sons.

Lyons, R. C., 2013. Price signals and bid-ask spreads in an illiquid market: The case of residential property in Ireland, 2006-2011. Available at SSRN 2205742.

Ma, L., 2011. Generating disaggregate population characteristics for input to travel-demand models. Ph.D. thesis, University of Florida.

Macket, R. L., 1983. The Leeds integrated land use transport (LILT) model. Repot sr 805, Crowthorne: UK Transport and Road Research Laboratory.

Madden, M., 2010. Older adults and social media. Tech. rep., Pew Internet & American Life Project, Pew Research Center, Washington D.C.

Mann, H. B., Whitney, D. R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 18 (1), 50–60.

Marks, A., Thrall, G., Arno, M., 1991. Siting hospitals to provide cost-effective health care. Geo Info Systems 2 (8), 58–66.

Marschak, J., 1960. Binary-choice constraints and random utility indicators. In: Proceedings of a Symposium on Mathematical Methods in the Social Sciences.

Martin, E., Shaheen, S., Lidicker, J., 2010. Carsharings's impact on household vehicle holdings: results from a North American shared-use vehicle survey. Transportation Research Record: Journal of the Transportation Research Board 2143, 150–158.

Martinez, F., 1995. Access: the transport-land use economic link. Transportation Research Part B 29 (6), 457–470.

Martinez, F., 1996. MUSSA: Land use model for santiago city. Transportation Research Record: Journal of Transportation Research Board 1552.

Martinez, F., Donoso, P., 2010. Residential Location Choice. Springer Berlin Heidelberg, Ch. The MUSSA II Land Use Auction Equilibrium Model, pp. 99–113.

Martinez, L. M., Viegas, J. M., 2009. Effects of transportation accessibility on residential property values. Transportation Research Record: Journal of the Transportation Research Board 2115, 127–137.

Martinez, L. M., Viegas, J. M., 2013. A new approach to modelling distance-decay functions for accessibility assessment in transport studies. Journal of Transport Geography 26, 87–96.

Matsuoka, R. H., Kaplan, R., 2008. People needs in the urban landscape: Analysis of landscape and urban planning contributions. Landscape and Urban Planning 84 (1), 7–19.

McCool, S. F., Stankey, G. H., 2004. Indicators of sustainability: Challenges and opportunities at the interface of science and policy. Environmental Management 33 (3), 294–305.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In Frontiers in Economics.

McFadden, D., 1981. Econometric Models of Probabilistic Choice. In Structural Analysis of Discrete Data with Econometric Applications. Vol. 198-272. MIT Press, Cambridge MA.

McFadden, D. L., 2001. Economic choices (revised version of nobel prize lecture). The American Economic Review 91 (3), 351–378.

McMillen, D., 2014. McSpatial Nonparametric spatial data analysis.

McMillen, D. P., 1992. Probit with spatial autocorrelation. Journal of Regional Science 32 (3), 335–348.

McMillen, D. P., Redfearn, C., 2010. Estimation and hypothesis testing for nonparametric hedonic house price functions. Journal of Regional Science 50 (3), 712–733.

Medda, F., 2012. Land value capture finance for transport accessibility: a review. Journal of Transport Geography 25, 154–161.

Mendenhall, W., Wackerly, D. D., Scheaffer, R. L., 1989. Mathematical statistics with applications (Fourth ed.). PWS-Kent, Ch. Nonparametric statistics, pp. 674–679.

Milakis, D., Vlastos, T., Barbopoulos, N., 2008. Relationships between urban form and travel behaviour in Athens, Greece. a comparison with Western European and North American results. European Journal of Transport and Infrastructure Research 8 (3), 201–215.

Millard-Ball, A., Murray, G., ter Schure, J., January 2006. Car-sharing as a parking management strategy. In: Proceedings of the 85th Annual Meeting of the Transportation Research Board. Washington D.C.

Millard-Ball, A., Murray, G., ter Schure, J., Fox, C., Burkhardt, J., 2005. Car-sharing: Where and how it succeeds. Tech. rep., Transit Cooperative Research Program (TCRP) Report.

Miyamoto, K., Vichiensan, V., Shimomura, N., Paez, A., 2004. Discrete choice model with structuralized spatial effects for location analysis. Transportation Research Record: Journal of Transportation Research Board 1898, 183–190.

Moniruzzaman, M., Páez, A., 2012. Accessibility to transit, by transit, and mode share: application of a logistic model with spatial filters. Journal of Transport Geography 24, 198–205.

Morrison, M., 2000. Aggregation biases in stated preference studies. Australian Economic Papers 39 (2), 215–230.

Müller, K., Axhausen, K. W., 2010. Population synthesis for microsimulation: State of the art. ETH Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau (IVT).

Municipality of Athens, 2008. Development business plan for the Municipality of Athens 2007-2010. Part A: Strategic plan.

Munkres, J., 1957. Algorithms for the assignment and transportation problems. Journal of the Society for Industrial & Applied Mathematics 5 (1), 32–38.

Munoz-Raskin, R., 2010. Walking accessibility to bus rapid transit: Does it affect property values? the case of Bogota Colombia. Transport Policy 17 (2), 72–84.

Murphy, J. J., Allen, P. G., Stevens, T. H., Weatherhead, D., 2005. A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics 30 (3), 313–325.

Nagel, K., 2013. MATSim User's Guide. Technical University of Berlin (TUB).

Nassauer, J. L., 2012. Landscape as medium and method for synthesis in urban ecological design. Landscape and Urban Planning 106 (3), 221–229.

Nelson, J. P., 1980. Airports and property values: A survey of recent evidence. Journal of Transport Economics and Policy 14 (1), 37–52.

Nelson, J. P., 1982. Highway noise and property values: A survey of recent evidence. Journal of Transport Economics and Policy 16 (2), 117–138.

Nelson, J. P., 2004. Meta-analysis of airport noise and hedonic property values: Problems and prospects. Journal of Transport Economics and Policy 38 (1), 1–27.

Nelson, J. P., 2008. Hedonic Property Value Studies of Transportation Noise: Aircraft and Road Traffic. Springer, Department of Ecnomics, Pennsylvania State University.

Nerenberg, V., Bernard, M. J., Collins, N. E., 1999. Evaluation results of San Francisco Bay Area station-car demonstration. Transportation Research Record: Journal of the Transportation Research Board 1666, 110–117.

Nham, B., Siangliule, K., Yeung, S., 2008. Predicting of transport from iphone accelerometer data. Tech. rep., Stanford University.

Nicolai, T., 2012. Integrating MATSim to UrbanSim for the sustaincity studies brussels and zurich. Tech. rep., Technische Universität Berlin.

Nicolai, T., Nagel, K., 2011. MATSim4UrbanSim – guide on UrbanSim usage of the travel model plug-in. Tech. rep., Transport Planning and Transport Telematics (VSP), Transport Planning and Transport Telematic (VSP), Technical University of Berlin (TUB).

Nicolai T., Wang L., N. K., Waddell, P., 2011. Coupling an urban simulation model with a travel model - a first sensitivity test. In: Computers in Urban Planning and Urban Management. Lake Louise, Canada.

Nobre, A., Pacheco, M., Jorge, R., Lopes, M. F. P., Gato, L. M. C., 2009. Geo-spatial multi-criteria analysis for wave energy conversion system deployment. Renewable Energy 34 (1), 97–111.

Novotny, J., 2007. On the measurement of regional inequality: does spatial dimension of income inequality matter? Annals of Regional Science 41 (3), 563–580.

Nuissl, H., Haase, D., Lanzendorf, M., Wittmer, H., 2009. Environmental impact assessment of urban land use transitions – A context-sensitive approach. Land Use Policy 26 (2), 414–424.

OASTH, 2013. Organization of Urban Transportation of Thessaloniki. http://oasth.gr/organization/general_eng.php, [Online; accessed 25-March-2013].

Odec, J., Rekdal, J., Hamre, T. N., 2003. The socio-economic benefits of moving from cordon toll to congestion pricing: The case of Oslo. In: Proceedings of the 92nd Annual Meeting Transportation Research Board.

Ohta, H., Fuji, S., Nishimura, Y., Kozuka, M., January 2009. Psychological analysis of acceptance of pro-environmental use of automobile: cases for carsharing and eco-car. In: Proceedings of the 88th Annual Meeting of the Transportation Research Board. Washington D.C.

Oketch, T., Carrick, M., 2005. Calibration and validation of a micro-simulation model in network analysis. In: Proceedings of the 84th TRB Annual Meeting, Washington, DC.

OPUS, January 2011. The Open Platform for Urban Simulation and UrbanSim. Users Guide and Reference Manual. The UrbanSim Project University of California Berkeley, and University of Washington.

Ouma, Y. O., Josaphat, S. S., Tateishi, R., 2008. Multiscale remote sensing data segmentation and post-segmentation change detection based on logical modeling: Theoretical exposition and experimental results for forestland cover change analysis. Remote Sensing of Environment, Computer and Geoscience 34 (7), 715–737.

Owojori, A., Xie, H., March 2005. Landsat image-based LULC changes of San Antonio, Texas using advanced atmospheric correction and object-oriented image analysis approaches. In: Proceedings of the 5th International Symposium on Remote Sensing of Urban Areas.

Ozbay, K., Bartin, B., Yanmaz-Tuzel, O., Berechman, J., 2007. Alternative methods for estimating full marginal costs of highway transportation. Transportation Research Part A: Policy and Practice 41 (8), 768–786.

Pacionne, M., 2003. Urban environmental quality and human wellbeing – a social geographical perspective. Landscape and Urban Planning 65 (1), 19–30.

Paez, A., 2006. Exploring contextual variations in land use and transport analysis using a probit model with geographical weights. Journal of Transport Geography 14 (3), 167–176.

Pan, H., Zhang, M., 2008. Rail transit impacts on land use: Evidence from Shanghai China. Transportation Research Record: Journal of the Transportation Research Board 2048, 16–25.

Park, S. W., Bahng, D. W., Park, Y. W., 2010. Price run-up in housing markets, access to bank lending and house prices in Korea. Journal of Real Estate and Financial Economics 40 (3), 332–367.

Parker, D., Steven, M. Marco, A., Janssen, M., Hoffmann, J., Deadman, P., 2003. Multi-agent systems for the simulation of land-use and land-cover change: A review. Annals of the Association of American Geographers 93 (2), 314–317.

Parsons Brinckerhoff Quade and Douglas Inc, February 2001. The effect of rail transit on property values. Tech. rep., NEORail II, Cleveland, Ohio.

Pattanaik, P. K., 1968. Risk, impersonality, and the social welfare function. Journal of Political Economy 76 (6), 1152–1169.

Patterson, Z., Bierlaire, M., 2010. Development of prototype UrbanSim models. Environment and Planning B: Planning & Design 37 (2), 344–366.

Patterson, Z., Hurtubia, R., August 2008. A prototype integrated transportation land-use model for the Lausanne region. Tech. rep., École Polytechnique Fédérale de Lausanne.

Patterson, Z., Kryvobokov, M., Marchal, F., Bierlaire, M., 2010. Disaggregate models with aggregate data. Journal of Transport and Land Use 3 (2), 5–37.

Pennington, G., Topham, N., Ward, R., 1990. Aircraft noise and residential property values adjacent to Manchester international airport. Journal of Transport Economics and Policy.

Pérez, M., Rey, E., 2013. A multi-criteria approach to compare urban renewal scenarios for an existing neighbourhood. Case study in Lausanne Switzerland. Building and Environment 65, 58–70.

Picard, N., Antoniou, C., 2011. Econometric guidance for developing econometric guidance for developing urbansim models. first lessons from the SustainCity project. In: Proceedings of the 51st European Congress of the Regional Science Association International. Barcelona, Spain.

Pinkse, J., Slade, M. E., 1998. Contracting in space: An application of spatial statistics to discrete-choice models. Journal of Econometrics 85 (1), 125–154.

Power, A., 2012. Social inequality, disadvantaged neighbourhoods and transport deprivation: an assessment of the historical influence of housing policies. Journal of Transport Geography 21, 39–48.

Pritchard, D. R., 2008. Synthesizing agents and relationships for land use/transportation modelling. Ph.D. thesis, University of Toronto.

Putman, S. H., 1983. Integrated Urban Models: Policy Analysis of Transportation and Land Use. Pion Press, London.

R Development Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL http://www.R-project.org/

Rahman, A., Harding, A., Tanton, R., Liu, S., 2010. Methodological issues in spatial microsimulation modelling for small area estimation. International Journal of Microsimulation 3 (2), 3–22.

Rahman, A., Harding, A., Tanton, R., Liu, S., 2013. Simulating the characteristics of populations at the small area level: New validation techniques for a spatial microsimulation model in Australia. Computational Statistics & Data Analysis 57 (1), 149–165.

Ramjerdi, F., 2006. Equity measures and their performances in transport. Transportation Research Record: Journal of the Transportation Research Board 1983, 67–74.

Ravulaparthy, S., Dalal, P., Chen, Y., Goulias, K. G., 2012. Exploratory analysis of spatial hierarchical clustering in Los Angeles County, California. Transportation Research Record: Journal of the Transportation Research Board 2307, 132–140.

Research, C. S. E., Consulting, 2002. Der schweizer immobilienmarkt - fakten und trends. Tech. rep., Credit Suisse.

ReVelle, C., Swain, R., 1970. Central facilities location. Geographical Analysis 2 (1), 30–34.

Richardson, A. J., 2003. Simulation study of estimation of individual specific values of time using adaptive stated-preference study. Transportation Research Record: Journal of the Transportation Research Board 1804, 13–20.

RICS Policy Unit, October 2002. Land value and public transport. Tech. rep., RICS.

Rodier, C., 2009. Review of the international modelling literature: transit, land use, and auto pricing strategies to reduce vehicle miles traveled and greenhouse gas emissions. Transportation Research Record: Journal of the Transportation Research Board 2123, 1–12.

Rodier, C., Shaheen, S., January 2004. Carsharing and carfree housing: predicted travel, emission, and economic benefits. a case study of the Sacramento, California region. In: Proceedings of the 83th Annual Meeting of the Transportation Research Board. Washington D.C.

Rodier, C., Shaheen, S., 2008. Easyconnect: Low-speed modes linked to transit planning project. California PATH research report. research report ucb-its-prr-2008-17. Tech. rep., Institute of Transportation Studies, University of California, Berkeley.

Rodriguez, D. A., Targa, F., 2004. Value of accessibility to Bogota's bus rapid transit system. Transport Reviews 24 (5), 587–610.

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. Journal of Political Economy 82 (1), 34–55.

Rosing, K. E., ReVelle, C. S., Rosing-Vogelaar, H., 1979. The p-median and its linear programming relaxation: An approach to large problems. The Journal of Operational Research Society 30 (9), 815–823.

Ross, S., 1997. Simulation. Statistical Modeling and Decision Science. Academic Press.

Roukouni, A., Basbas, S., Kokkalis, A., 2012. Impacts of a metro station to the land use and transport system: the Thessaloniki metro case. In: Procedia - Social and Behavioral Sciences. Vol. 48. pp. 1155 – 1163.

Salvini, P., Miller, E., 2005. Ilute: An operational prototype of a comprehensive microsimulation model of urban systems. Networks and Spatial Economics 5, 217–234.

Santos, B., Antunes, A., Miller, E. J., 2008. Integrating equity objectives in a road network design model. Transportation Research Record: Journal of the Transportation Research Board 2089, 35–42.

Savelyev, A. K., Janoxicz, K., Thatcher, J., Xu, S., Mulligann, C., Luo, W., 2011. Volunteered geographic services: Developing a linked data driven location-based service. In: Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Semantics and Ontologies. pp. 25–31.

Schaefer, M., 2011. 'Standortmosaik Zurich' or the ecology of access. Anthos, 50–53.

Schafer, J., 2010. Analysis of incomplete multivariate data. CRC Press.

Scheiner, J., Holz-Rau, C., 2007. Travel mode choice: affected by objective or subjective determinants? Transportation 34 (4), 487–511.

Schetke, S., Haase, D., Kötter, T., 2012. Towards sustainable settlement growth: A new multi-criteria assessment for implementing environmental targets into strategic urban planning. Environmental Impact Assessment Review 32, 195–210.

Scholz, R. W., Tietje, O., 2002. Embedded case study methods: Integrating quanitative and qualitative knowledge. Sage, Thousand Oaks, California.

Sener, I. N., Pendyala, R. M., Bhat, C. R., 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. Journal of Transport Geography 19 (2), 294–303.

Ševčíková, H., E., R. A., Waddell, P., 2007. Assessing uncertainty in urban simulations using bayesian melding. Transportation Research Part B: Methodological 41 (6), 652–669.

Ševčíková, H., E., R. A., Waddell, P., 2011. Uncertain benefits: Application of bayesian melding to the Alaskan way viaduct in Seattle. Transportation Research Part A: Policy and Practice 45 (6), 540–553.

Shaheen, S., Cohen, A., 2007. Growth in worldwide carsharing: an international comparison. Transportation Research Record: Journal of the Transportation Research Board 1992, 81–89.

Shaheen, S., Cohen, A., Martin, E., 2010a. Carsharing parking policy: review of North American practices and San Francisco Bay area case study. Transportation Research Record: Journal of the Transportation Research Board 2187, 146–156.

Shaheen, S., Guzman, S., Zhanq, H., 2010b. Bikesharing in Europe, the Americas and Asia: Past, present and future. Transportation Research Record: Journal of the Transportation Research Board 2143, 159–167.

Shaheen, S., Hua, Z., Elliot, M., Guzman, S., 2011. China's Hangzhou public bicycle. Transportation Research Record: Journal of the Transportation Research Board 2247, 33–41.

Shaheen, S., Martin, E., January 2007. Assessing early market potential for carsharing in China: a case study of Beijing. In: Proceedings of the 86th Annual Meeting of the Transportation Research Board. Washington D.C.

Shaheen, S., Rodier, C., Murray, G., Cohen, G., Martin, E., 2010c. Carsharing and public parking policies: Assessing benefits, costs, and best practicies in North America. Tech. Rep. Report CA-MTI-10-2612, Mineta Transportation Institute.

Shaheen, S., Schwartz, A., Wipyewski, K., 2004. Policy considerations for carsharing and station cars. Transportation Research Record: Journal of the Transportation Research Board 1887, 128–136.

Siora, E., 2011. Land use evolution through classification of remotely sensed imagery and econometric models. Master's thesis, School of Rural and Surveying Engineering, National Technical University of Athens.

Siverman, B. W., 1986. Density estimation for statistics and data analysis. Tech. rep., School of Mathematics University of Bath, UK. Statistics and Applied Probability, London: Chapman and Hall.

Small, K. A., Steimetz, S., 2007. Spatial hedonics and the willingnes to pay for residential amenities. Journal of Regional Science 52 (4), 635–647.

Smith, A. H., Batty, M., Crooks, A., Milton, R., 2009. Mapping for the masses accessing web 2.0 through crowdsourcing. Social Science Computer Review 27 (4), 524–538.

Smith, T., Nelischer, M., Perkins, N., 1997. Quality of an urban community: a framework for understanding the relationship between quality and physical form. Landscape and Urban Planning 39 (2), 229–241.

So, H. M., Tse, R., Gansen, S., 1966. Estimating the influence of transport on house prices: Evidence from Hong Kong. Journal of Property Valuation and Investment 15 (1), 40–47.

Spiliopoulou, C., Antoniou, C., 2012. Analysis of illegal parking behaviour in Greece. In: Transport Research Arena. Athens, Greece.

Srinivasan, K. K., Ramadurai, G., Muthuram, V., Srinivasan, S., 2007. Determinants of changes in mobility and travel patterns in developing countries: A case study of chennai. Transportation Research Record: Journal of the Transportation Research Board 2038, 42–52.

Steimetz, S., 2010. Spatial multipliers in hedonic analysis: a comment on “spatial hedonic models of airport noise, proximity, and housing prices”. Journal of Regional Science 50 (5), 995–998.

Steinitz, C., 2012. A framework for geodesign: Changing geography by design. Journal of Landscape Architecture 7 (2), 87.

Surprenant-Legault, J., Patterson, Z., El-Geneidy, A. M., 2013. Commuting trade-offs and distance reduction in two-worker households. Transportation Research Part A: Policy and Practice 51, 12–28.

Sustaincity, October 2009. Micro-simulation for the prospective of sustainable cities in Europe. Seventh Framework Program FP7, Annex.

Talen, E., 1998. Visualizing fairness: Equity maps for planners. Journal of the American Planning Association 64 (1), 22–38.

Tanton, R., Edwards, K., 2013. Spatial microsimulation: a reference guide for users. Vol. 6. Springer.

Tanton, R., Vidyattama, Y., Nepal, B., McNamara, J., 2011. Small area estimation using a reweighting algorithm. Journal of the Royal Statistical Society: Series A (Statistics in Society) 174 (4), 931–951.

Taylor, D. B., Miller, D., Iseki, H., Fink, C., 2008. Nature and/or nurture? analyzing the determinants of transit ridership in urbanized areas. Transportation Research Part A: Policy and Practice 43 (1), 60–77.

Teitz, M. B., Bart, P., 1968. Heuristic methods for estimating the generalized vertex median of a weighted graph. Operations Research 16 (5), 955–961.

Termorshuizen, J. W., Opdam, P., van den Brink, A., 2007. Incorporating ecological sustainability into landscape planning. Landscape and Urban Planning 79 (3), 347–384.

Thaler, R. H., Sunstein, C. R., 2009. Nudge, Improving Decisions About Health, Wealth, and Happiness. Yale University Press.

Thalmann, J., 2012. Ökosystemleistungen im siedlungsraum: Analyse des potentials der lebensraumqualität für vier verdichtungsszenarien in schlieren. Master’s thesis, GEO511 - Masterarbeit an der mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich (in German).

Theebe, M. A. J., 2004. Planes, trains, and automobiles: The impact of traffic noise on house prices. The Journal of Real Estate Finance and Economics 28 (2-3), 209–234.

Tiefelsdorf, M., Boots, B., 1995. The exact distribution of Moran’s I. Environment and Planning A 27 (6), 985–999.

Tiefelsdorf, M., Griffith, D. A., 2007. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. Environment and Planning A 39 (5), 1193–1221.

Tiefelsdorf, M., Griffith, D. A., Boots, B., 1999. A variance-stabilizing coding scheme for spatial link matrices. Environment and Planning A 31, 165–180.

Train, K., 2009. Discrete Choice Methods with Simulation (second ed.). Cambridge University Press, New York.

Transportation Research Board, 1999. A handbook for measuring customer satisfaction and service quality. Report 47, TRCP.

Transportation Research Board, 2004. Transit capacity and quality of service manual. Report 100, TRCP.

Trulia, 2013. Trulia's rent vs. buy report: Full methodology. Tech. rep., www.trulia.com.

Tu, K.-J., Lin, L. T., 2008. Evaluative structure of perceived residential environmental quality in high-density and mixed-use urban settings: An exploratory study on Taipei city. Landscape and Urban Planning 87 (3), 157–171.

Turcu, C., 2012. Local experiences of urban sustainability: Researching housing market renewal interventions in three English neighbourhoods. Progress in Planning 78 (3), 101–150.

Tyrinopoulos, Y., Antoniou, C., July 2008. Public transit user satisfaction: Variability and policy implications. Transport Policy 15 (4), 260–272.

Tyrinopoulos, Y., Antoniou, C., 2013. Factors affecting modal choice in urban mobility. European Transport Research Review 5 (1), 27–39.

UN-Habitat, 2013. Urban planning for city leaders. http://www.unhabitat.org/pmss/listItemDetails.aspx?publicationID=3385, Nairobi.

Van Acker, V., Mokhtarian, P. L., Witlox, F., 2010. Car ownership explained by the structural relationships between lifestyles, residential location and underlying residential and travel attitudes. Submitted to Transport Policy.

Van Eggermond, M., Lehner, M., Earth, A., 2011. Modeling hedonic prices in Singapore applying spatial hedonic regression. In: FCL Conference. Singapore.

Van Kamp, I., Leidelmeijer, K., Marman, G., de Hollander, 2003. Urban environmental quality and human well-being: Towards a conceptual framework and demarcation of concepts; a literature study. Landscape and Urban Planning 65 (1), 5–18.

Vandenbulcke, G., Steenberghen, T., Thomas, I., 2007. Accessibility indicators to places and transports. Tech. rep., Politique Scientifique Fédérale et SPF Mobilité et Transports, Brussels.

Vandenbulcke, G., Steenberghen, T., Thomas, I., 2009. Mapping accessibility in Belgium: A tool for land-use and transport planning? Journal of Transport Geography 17 (1), 39–53.

Vautin, D. A., Walker, J., January 2011. Transportation impacts of information provision and data collection via smartphones. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board. Washington D.C.

Viegas, J. M., 2001. Making urban road pricing acceptable and effective: searching for quality and equity in urban mobility. Transport Policy 8 (4), 289–294.

Voas, D., Williamson, P., 1998. Testing the acceptability of random number generators. Tech. rep., Working Paper 1998/2). Liverpool: Population Microdata Unit, Department of Geography, University of Liverpool. (Available from: http://pcwww. liv. ac. uk/microdata).

Voas, D., Williamson, P., 2000. An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata. International Journal of Population Geography 6 (5), 349–366.

Voas, D., Williamson, P., 2001. Evaluating goodness-of-fit measures for synthetic microdata. Geographical and Environmental Modelling 5 (2), 177–200.

Von Haefen, R., 2003. Incorporating observed choice into the construction of welfare measures from random utility models. Journal of Environmental Economics and Management 45 (2), 145–165.

Waddell, P., 2002. UrbanSim: Modelling urban development for land use, transportation and environmental planning. Journal of the American Planning Association 68 (3), 297–314.

Waddell, P., 2010. Modeling residential location in urbansim. In: Pagliara, F., Preston, J., Simmonds, D. (Eds.), Residential Location Choice, Models and Applications. Springer Berlin Heidelberg, pp. 165–180.

Waddell, P., 2011. Integrated land use and transportation planning modelling: Addressing challenges in research and practice. Transport Reviews 31 (2), 209–229.

Waddell, P., Ulfarsson, G., 2003. Dynamic simulation of real estate development and land prices within and integrated land use and transportation model system. In: Proceedings of the 82nd Annual Meeting of the Transportation Research Board.

Waddell, P., Ulfarsson, G., 2004. Introduction to urban simulation: design and development of operational models. Handbook in Transport 5, 203–236.

Waddell, P., Ulfarsson, G., Franklin, J., Lobb, J., 2007. Incorporating land use in metropolitan transportation planning. Transportation Research Part A: Policy and Practice 41 (5), 382–410.

Walker, J. L., 2001. Extended discrete choice models: Integrated framework, flexible error structures and latent variables. Ph.D. thesis, MIT.

Walsh, M., 1990. Global trends in motor vehicle use and emissions. Annual Review of Energy 15 (1), 217–243.

Walz, A., Lardelli, C., Behrendt, M., Grêt-Regamey, A., Lundström, C., Kytzia, S., Bebi, P., 2007. Participatory scenario analysis for integrated regional modeling. Landscape and Urban Planning 81 (1), 114–131.

Wang, X., Kockelman, K. M., 2006. Tracking land cover change in a mixed logit model: Recognizing temporal and spatial effects. In: Proceedings of the 85th Annual Meeting of the Transportation Research Board. Washington D.C.

Wang, X., Kockelman, K. M., 2008. Bayesian inference for ordered response data with dynamic spatial ordered probit model. Journal of Regional Science 49 (5), 877–913.

Wang, X., Kockelman, K. M., 2009a. Application of the dynamic spatial ordered probit model. Transportation Research Record: Journal of the Transportation Research Board 2136, 45–56.

Wang, X., Kockelman, K. M., 2009b. Application of the dynamic spatial ordered probit model: Patterns of land development change in Austin, Texas. Papers in Regional Science 88 (2), 345–365.

Wang, X., Kockelman, K. M., Lemp, J. D., 2012. The dynamic spatial multinomial probit model: analysis of land use change using parcel-level data. Journal of Transport Geography 24, 77–88.

Wang, Y., Kockelman, K. M., Wang, X., 2011. Anticipating land use change using geographically weighted regression models for discrete response. Transportation Research Record: Journal of the Transportation Research Board 2245, 111–123.

Wang, Y., Kockelman, K. M., Wang, X., 2013. Understanding spatial filtering for analysis of land use-transport data. Journal of Transport Geography 31 (1), 123–131.

Washington, S. P., Karlaftis, M. G., Mannering, F. L., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and hall/CRC, London.

Wegener, M., 1982. A multilevel economic-demographic model for the Dortmund region. Sistemi Urbani 3, 371–401.

Wegener, M., 2004. Overview of land-use transport models. In: Hensher, D., Button, K. (Eds.), Transport Geography and Spatial Systems. Vol. 5. Pergamon/Elsevier, Kidlington, UK, pp. 127–146.

Wegener, M., March 2011. From macro to micro-how much micro is too much? Transport Reviews 31 (2), 161–177.

Weinberger, R. R., 2000. Light rail proximity: Benefit or detriment in the case of Santa Clara Country, California? Transportation Research Record: Journal of the Transportation Research Board 1747, 104–113.

Weinberger, R. R., 2001. Commercial rents and transportation improvements: The case of Santa Clara country's light rail. Tech. rep., Lincoln Institute of Land Policy, Cambridge, MA.

Wheaton, W., 1977. A bid rent approach to housing demand. Journal of Urban Economics 4 (2), 200–217.

Wickham, H., 2014. ggplot2: An implementation of the Grammar of Graphics.
URL http://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf

Wiek, A. amd Binder, C., Scholz, R. W., 2006. Functions of scenarios in transition processes. Futures 38 (7), 740–766.

Wilhem, S., Matos, M. G., 2013. spatialprobit: Spatial Probit Models.
URL http://cran.r-project.org/web/packages/spatialprobit/spatialprobit.pdf

Williamson, P., 2013. An evaluation of two synthetic small-area microdata simulation methodologies: Synthetic reconstruction and combinatorial optimisation. In: Spatial Microsimulation: A Reference Guide for Users. Springer, pp. 19–47.

Williamson, P., Birkin, M., Rees, P. H., 1998. The estimation of population microdata by using data from small area statistics and samples of anonymised records. Environment and Planning A 30 (5), 785–816.

Wilson, A. G., 1998. Land-use/transport interaction models: Past and future. Journal of Transport Economics and Policy 32 (1), 3–26.

Wissen, U., von Wirth, T., Kunze, A., Neuenschwander, N., 2012. How does the Limmattal present itself in the year 2030? four regional scenarios. Tech. rep., Sustainable Urban Patterns.

Won Kim, C., Phipps, T. T., Anselin, L., 2003. Measuring the benefits of air quality improvement: a spatial hedonic approach. Journal of Environmental Economics and Management 45 (1), 24–29.

Wood, S. N., 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. The Journal of American Statistical Association 99 (467), 673–686.

Wood, S. N., 2014. mgcv: GAMs with GCV/AIC/REML smoothness estimation and GAMMs by PQL.

Wu, J., 2010. Urban sustainability: an inevitable goal of landscape research. Landscape Ecology 25 (1), 1–4.

Xifilidou, A., Karanikolas, N., Spatalas, S., 2012. The effect of central metro stations on real estate values. Journal of Land use, Mobility and Environment 2, 185–193.

Yang, D., Xue, G., Fang, X., Tang, J., 2012. Crowdsourcing to smartphones: Incentive mechanism. design for mobile phone sensing. In: Proceedings of the 18th annual international conference on Mobile computing and networking. ACM, pp. 173–184.

Zaidi, A., Harding, A., Williamson, P., 2009. New frontiers in microsimulation modelling. Ashgate Vienna.

Zhang, Q., Wang, J., Peng, X., Gong, P., Shi, P., 2002. Urban build-up land change detection with road density and spectral information from multitemporal landsat TM data. International Journal of Remote Sensing 23 (15), 3057–3078.

Zhao, F. Park, N., 2004. Estimation of AADT using geographically weighted regression models. Transportation Research Record: Journal of the Transportation Research Board 1876, 99–107.

Zhao, Y., Kockelman, K., Karlström, A., January 2008. Welfare calculations in discrete choice settings: The role of error term corelation. In: Proceedings of the 87th Annual Meeting of the Transportation Research Board. Washington D.C.

Zheng, Y., Zhang, L., Xie, X., Ma, W.-Y., 2009. Mining interesting locations and travel sequences from gps trajectories. In: Proceedings of the 18th international conference on World Wild Web. Madrid, Spain.

Zhou, B., Kockelman, K. K., 2008. Neighborhood impacts on land use change: A multinomial logit model of spatial relationships. Annals of Regional Science 42 (2), 321–340.

Zhou, B., Kockelman, K. K., 2011. Land use change through microsimulation of market dynamics: An agent-based model of land development and locator bidding in Austin, Texas. Transportation Research Record: Journal of the Transportation Research Board 2255, 125–136.

Zhou, B., Kockelman, K. M., Gao, R., 2011. Opportunities for and impacts of carsharing: a survey of the Austin, Texas market. International Journal of Sustainable Transport 5 (3), 135–152.

Ziliak, S. T., McCloskey, D. N., Deirdre, N., 2008. The cult of statistical significance: How the standard error costs us jobs, justice, and lives. University of Michigan Press.

Zöllig, C., Axhausen, K. W., September 2011. A conceptual, agent-based model of land development for UrbanSim. In: 51th ERSA Conference. Barcelona.

Zook, M., Graham, M., Shelton, T., Gorman, S., 2010. Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. World Medical & Health Policy 2 (2), 7–33.