Essays on urban and spatial economics

Rosa Sanchis-Guarner

30th March 2012

A thesis submitted to the Department of Geography of the London School of Economics for the degree of Doctor of Philosophy, London

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

London, 30th March 2012

Rosa Sanchis-Guarner

Statement of conjoint work

Chapters 2, 3 and 4 of this thesis are based on research I undertook while working as a Research Assistant at the Spatial Economics Research Centre. The aim of this work was to produce two policy reports for the UK Department of Transport, which were published in April 2009 and October 2010. This work was produced jointly with Dr. Steve Gibbons, Dr. Teemu Lyytikäinen and Professor Henry Overman.

I built on this research to produce the three chapters on the "Economic Impacts of Transport Policy". I developed a separate methodology chapter (chapter 2) in which I added a theoretical section and a literature review section which are not present in previous work. I also substantially extended the description of the construction of the accessibility measures and produced a larger set of descriptive statistics.

Chapter 3 is partly based on a joint paper derived from the policy reports mentioned above, to which I added extra regression results not produced in the paper.

Chapter 4 shares the methodological framework and some data sources with chapters 2 and 3 but it is my own work.

This statement is to confirm I contributed a minimum of 50% to chapters 2 and 3 as agreed to by the undersigned.

Rosa Sanchis-Guarner *Teemu Lyytikäinen* *Steve Gibbons* *Henry G. Overman*

Contents

Acknowledgements

Writing a thesis can be an arduous and an extremely enjoyable experience at the same time. Along my PhD I have been helped and encouraged by many people. These pages provide me with the opportunity of acknowledging them.

I am greatly indebted to my supervisor Henry Overman for his continuous support along these years. He has provided invaluable guidance while giving me enough flexibility to make and learn from my own mistakes. He has witnessed my evolution as a researcher and contributed to it enormously, helping me to move forward when most needed. Henry has taught me to think critically about economics, to challenge trivial explanations and to put an emphasis on being rigorous in empirical research. I really appreciate the amount of time he has devoted to help me and the trust he has put on my ability to conduct research.

I am also very thankful to the other two persons with whom I have worked more closely during these last years. Steve Gibbons has helped me in many occasions, responding to innumerable questions and emails. The clarity with which he understands and explains complicated issues is a source of inspiration. I have learnt hugely working under his supervision. Teemu Lyytikäinen contributed enormously towards this thesis by providing a vast amount of support, both professionally and personally. We spent many hours working together at the VML Data Lab and we had countless discussions over tea at the CEP Social Space. Working with such clever and modest people has been incredibly enjoyable and I hope we share many projects in the future.

Matilde Mas and Pilar Chorén, from *Instituto Valenciano de Investigaciones Económicas* (IVIE), Francisco Goerlich, from the University of Valencia, and Félix Alonso Luengo, from the Spanish Housing Department, kindly helped with the data in chapter 1. The research in chapters 2 to 4 would not have been possible without the assistance of the staff of the *Virtual Microdata Laboratory* at the ONS and the *Secure Data Service* at the UK Data Archive. I am specially thankful to Richard Welpton for his invaluable help with the data sources and service. I would also want to acknowledge the *Fundación Ramón Areces*, the Bank of Spain and the LSE Department of Geography for financial support.

My colleagues from the Economic Geography cluster at the LSE Department of

Geography, from the Spatial Economics Research Centre and from the LSE Centre for Economic Performance also provided useful insights and suggestions for my research. I am specially thankful to Olmo Silva, Christian Hilber, Ralf Martin, Giulia Faggio and Elias Einiö.

I have been lucky enough to share these years with a group of bright fellow PhD students who provided another invaluable source of support and inspiration. I am indebted to Carlo Menon, Max Nathan, Felix Weinhardt, Max Cali, Ben Faber and Neil Lee for their help and encouragement at the different stages of my PhD. I also shared countless conversations about our PhDs, and many other things, with Dimitri Szerman, Giuseppe Berlingieri, Joachim Groeger, Ashwini Natraj and Patrick Blanchenay, both inside and outside the School.

Nacho, Yago y María has sido un apoyo constante durante estos años. Dejarlos en Barcelona fue uno de los momentos más duros de mi vida pero ellos me han demostrado que, a pesar de la distancia, el "equipo" está ahí para lo que necesite. Mis "tordas" y la colla de "Entrepins" han hecho que no olvide de dónde vengo. Desde mi llegada a Londres en 2006 he conocido a un montón de personas fantásticas que han hecho estos los años más interesantes de mi vida. Mención especial merecen Joseba y Jana, que me ayudaron mucho cuando Arnau estaba lejos.

No pueden faltar unas palabras de gratitud para mis padres, Manuel y Ana, y para mis hermanos, Tania y Manuel. Sé que que esté lejos también ha sido duro para ellos, y estoy muy agradecida por su apoyo constante. Hablar por "Skype" con mi sobrino Arnau es siempre uno de los mejores momentos de mi semana.

Les últimes paraules són per Arnau, la persona sense qui la meua feina i la meua vida no serien possibles. Qui anava a dir que aquell xic prim amb gorra de marinero que vaig conèixer a l'estiu del 2000 és convertiria en el company de la meua vida. No hi ha paraules per expressar quant m'has ajudat durant tots aquestos anys. Aquesta tesi és el final d'una època i alhora el principi d'una altra que espere caminem l'un al costat de l'altre.

Abstract

This thesis is composed of four chapters. The first one investigates the impact of immigration on housing markets. The rest study the effects of transport policy on economic outcomes.

Chapter 1 provides causal estimates of the effects of an increase of foreign-born population on house prices. I use data for the Spanish provinces between 2001 and 2010. In order to infer causality I construct an instrument based on past location patterns by immigrant nationality. I find positive effects of the increase in the share of foreign-born population on both rental and purchase prices. The estimated elasticities are 0.6% for rental prices and 2% for purchase prices. I also investigate the relationship between immigration and native location (native displacement) and I find that immigrants attract natives to the same regions they locate. When I re-estimate the effects using solely the variation on population growth which is due to exogenous location of foreign-born, I find that estimates are around 30-40% smaller than if we ignored the relationship between immigration and native location decisions.

Chapters 2 to 4 investigate the effects of road improvements on aggregate and individual economic outcomes, using data for Great Britain during the period 1998-2008. Chapter 2 develops the methodology to estimate the economic impacts of transport improvements. We summarise the existing evidence and the theoretical channels through which transport policy can impact firm, worker and aggregate economic outcomes. To capture the effect of road improvements, we construct a measure of accessibility to employment through the road network. For this purpose, we collect novel data on 31 major road improvement projects and combine this information with the trunk road network in Great Britain in 2008. This information is used to calculate optimal travel times between locations at each point in time, which are used in the computation of the accessibility measures.

The last two chapters discuss the empirical results, for ward and firm outcomes (chapter 3) and for individual labour market outcomes (chapter 4). I find positive effects of accessibility on ward employment and number of plants, a limited effect on plant employment and no effect on productivity. Accessibility from workplace has substantial impacts on individual wages and total hours worked, while accessibility from home only seems to have an effect on reducing the travel time to work.

Introduction

I Overview

I.1 Spatial economics as a discipline

Since the publication of Krugman's works at the beginning of the 90s (Krugman, 1991a,b), there has been a renewed interest to include geography within mainstream economic analysis. The research described in this thesis provides solid empirical evidence to answer a number of open questions in urban and spatial economics.

Economics is the social science that studies the production, distribution and consumption of goods and services. It analyses the distribution of resources among economic agents (firms and workers). It also examines the way inputs (labour, capital and intermediate goods) are combined and transformed into final goods and services, which are distributed and consumed. Spatial economics is specifically concerned with the allocation of resources over space and the location of economic activity. This covers location theory, spatial competition and regional and urban economics (Duranton, 2008). It is connected to other areas of economics like international trade, real estate economics and local public economics. It embraces what geographers call "economic geography" and economists call "geographical economics" (following the definitions of Martin, 1999).

As discussed in Duranton (2008), the importance of the spatial dimension depends on the type of economic question we are investigating. Some questions are intrinsically "spatial", i.e. the spatial dimension plays a central role. For example, why are there cities? Why do they grow? Where do firms locate? For other questions we are trying to unveil what is the relative importance of space. For instance, do technological spillovers depend on geographical proximity? Do countries that are closer trade more? Finally, for other economic questions the importance of space is much smaller. For example, the determination of interbank interests rates is quite unlikely to be affected by space.

Economic agents (firms and workers) are distributed across space and they interact spatially. The frequency and "quality" of these interaction depend on their location and on how close they are to other economic agents. These interactions can affect economic outcomes and thus impact where the economic agents locate. For example, the performance of a worker in the labour market depends on his characteristics (age, gender, qualifications, ability) but it can also depend on how close he is located relative to jobs. If his home and job are spatially separated, he has to spend part of his working day traveling to work. The length of this commute may affect his labour/leisure optimal choice, his labour market status or even his productivity (van Ommeren & Gutiérrez-i-Puigarnau, 2011). If workers live closer to firms, these have better access to potential employees and might find it easier to adjust their employment when facing a productivity shock (Overman & Puga, 2010). Other firms and workers might find it profitable to relocate into these employment cores (also called agglomerations) in order to gain from proximity. This example illustrates how the interactions between workers and firms depend on their location and how, if interactions have meaningful effects on economic outcomes, the concentration of economic agents in the space can endogenously determine the location decisions of agents.

Spatial economics encompasses urban economics and economic geography. According to the definition of Quigley (2008), urban economics deals with the spatial arrangements of households, firms, and capital in metropolitan areas. It also studies the externalities which arise from the proximity of households and from land uses, and the public policy issues which arise from the interplay of these economic forces. In short, urban economics studies cities and why they exist and grow (Glaeser, 2008). On the other hand, the various definitions of economic geography (Combes et al., 2008b; Brakman et al., 2009) stress the unequal distribution of economic activity across space. This branch of spatial economics concerned with the explanation of these inequalities. It searches to understand the nature, extent, causes and consequences of these disparities.

Two branches of the spatial economics literature have received a great deal of attention in the last years, both from a theoretical and an empirical point of view. The "new" economic geography (or NEG) (Krugman & Venables, 1995; Fujita et al., 2001) studies the uneven distribution of economic activity across space using a general equilibrium framework and rigorous microeconomic foundations. Even if many of its mechanisms had been largely discussed in the "proper" economic geography literature[1], Krugman and followers put the role of space at the centre of the mainstream economics discussion.

The main contribution of NEG is to give micro-foundations to the emergence of economic agglomeration and spatial inequality (Venables, 2008). The core building blocks of new economic geography models are product differentiation (modeled

[1]See Overman (2004); Duranton & Rodríguez-Pose (2005) and Duranton & Storper (2006) for an extensive discussion on the dialogue between economists and geographers.

through a love of variety assumption), increasing returns to scale and transport costs, which together create pecuniary externalities in agents' location choices. When combined with either factor mobility or intermediate inputs, these three building blocks give rise to forces of cumulative causation and agglomeration (Redding, 2010).

At the same time, a large body of theoretical and empirical literature has been devoted to investigate the existence, magnitude and causes of agglomeration (see Puga, 2010, for a review) and the concentration of economic activity in cities (Glaeser, 2008). Differently to NEG, this literature explains the existence of agglomeration due to the productivity advantages of being located in cities. Agglomeration economies concern the positive externalities that arise from economic concentration and that might have positive effects on firm and worker productivity. The spatial clustering of industries in specific places gives rise to increased interactions between economic agents which can be capitalised into higher productivity and wages. The mechanisms behind these interactions have attracted considerable attention.

The most common classification of agglomeration economies is that of Marshall (1890), which identifies three mechanisms: input sharing (linkages between input suppliers and final producers), labour market pooling (think local labour market interactions) and knowledge spillovers. Duranton & Puga (2004) provide an alternative classification of the mechanisms that give rise to local increasing returns: sharing, matching and learning[2]. Some papers have tried to empirically disentangle the different mechanisms through which agglomeration economies are operating (Rosenthal & Strange, 2001; Jofré-Monseny et al., 2011). Empirically, the different mechanisms are however hard to distinguish (Overman & Puga, 2010) and agglomeration economies are difficult to identify (Combes et al., 2011). Substantial attention has been given to the fact that wages are larger in cities (see for example Combes et al., 2010).

The research carried out in this thesis relates to these core issues of spatial economics. The issues under investigation, which are explained in detail below, are connected to urban labour economics, transportation, agglomeration economies, spatial mobility, real estate and regional economics. This thesis aims to contribute to these literatures by answering original research questions using rigorous empirical methods. The rest of the introduction is organised as follows. Section I.2 summarises the research methodology followed in this thesis.Section II provides a brief summary of the chapters. Finally, section III explains the main contributions of my research.

[2]Although many other forces can lead to agglomeration, like natural advantages, consumption amenities or historical accidents. See Strange (2008) for a review.

I.2 Research methodology

This thesis is divided in two parts. The first part, chapter 1, studies the impact of immigration on housing markets, using data for Spain between 2001 and 2010. The second part, chapters 2 to 4, studies the impact of transport policy on aggregate and individual economic outcomes, focusing on road construction in Great Britain between 1998 and 2008. This second part comprises a methodological chapter (2) and two chapters in which the empirical results are discussed (3 and 4). A common methodological approach is applied thorough this thesis, which is the standard methodology in applied economic studies and, in particular, in what has been called "geographical" economics (as discussed by Overman, 2004).

A comprehensive review of the literature helps to identify gaps in the existing empirical evidence and to formulate relevant research questions which have not been satisfactorily addressed in previous studies. The analysis of the relevant related papers allows us to define a set of theoretical predictions on the expected effects of the interest variables on outcomes. The research questions are therefore guided by existing theories and empirical evidence. They are formulated in order to be clear and specific. In this sense, an effort is made to synthesize complex phenomena into simple research questions. We prioritize the rigorous answer of a simple question over the investigation of wider issues.

Given that our empirical approach is based on the estimation of reduced-form empirical specifications, the statement of clear theoretical predictions is important for the formalisation of the empirical equations and for the interpretation of the size and sign of the estimated coefficients. Specifically, section 1.2.2 of chapter 1 discusses the channels through which changes in the population size of regions caused by immigration inflows might affect housing demand and prices. Section 2.2 of chapter 2 reviews the different channels through which transport policy, in particular road construction, affects worker and firm economic outcomes.

The empirical methodology is developed in order to test the hypothesis derived from the theoretical framework. The expected empirical relationships between the outcome variable and the explanatory variable in which we focus our analysis (regressor of interest) are formalised into empirical specifications which are tested in the data. In the thesis, the empirical specifications are discussed in detail. The outcome variable and regressor of interests are clearly identified and the different elements of the equations and how they affect the outcome variable are carefully explained. The variables are accurately defined in order to capture the economic relationships which we are interested in. To quantitatively assess the potential effects we perform regression analysis and apply several econometric techniques. We rigourously estimate coefficients that summarise the nature of the economic relationship under study.

To carry out the empirical exercises several datasets are used, both for Spain and Great Britain. The geographical level of analysis was chosen to provide an appropriate setup for the testing of the research hypothesis. An effort is made to employ the best available data and spatial units for the research questions investigated. For example, in the first part of the thesis, to be able to assess the effect of immigration on house prices, I collect data both on housing purchase and rental prices, because both prices are relevant indicators of housing markets. With the aim of appropriately capturing the effects of immigration, to calculate the total number of immigrants I use register data on the population of foreign-born in each spatial unit, instead of using a weighted total derived from a sample. The geographical unit used is the province, which, as discussed in the text, is the best unit available to approximate for labour and housing markets. Alternatively, in the case of the assessment of the economic effects of transport policy, to measure transport policy we use road construction. In particular, we capture the effects of road construction by using a measure of accessibility to employment from a given location. With respect to the spatial unit, in this part of the thesis we perform the analysis using very small geographical units (wards) because our identification strategy is partly based on the use of small spatial scale.

The aim of the regression analysis is the estimation of causal effects, so particular attention is devoted to the correct identification of the parameters of interest. This involves the estimation of unbiased parameters which can be interpreted as causal effects. Omitted variable bias, reverse causation, simultaneity bias, spurious correlation, unobserved effects, measurement error or poor proxies can challenge the validity of our estimates.

Identification is important for two reasons. First, even if interesting, simple associations or partial correlations can be uninformative or misleading when drawing policy implications from the empirical results. If we want to be able to derive policy recommendations from our empirical results it is crucial to estimate the effects in the correct direction of causality[3]. Secondly, sometimes the theoretical predictions on the expected impacts can be ambiguous and the direction of the net effects is unclear. Hence, obtaining robust empirical evidence becomes of increasing importance in order to learn about the different theoretical channels which might be driving the results. Causal estimates are thus more informative than partial correlations when linking the results to the theoretical framework.

Even though randomised or natural experiments would be the ideal setting to estimate causal effects, as noted by Duranton (2008), it is difficult to find these ex-

[3]For the economic geography field, Combes (2011) argues that addressing endogeneity appropriately is basic to be able to draw policy implications.

periments within the spatial economics field. Consequently, we rely on other econometric techniques in order to claim causality of our estimates. The identification strategy we follow can be summarised in the following points[4]:

1. We use longitudinal datasets (panel data), both at the aggregate level (panel of regions) and at the individual level (panel of workers and firms). This has three advantages. First, it allows us to control for unobservable time-invariant characteristics of the observation units which might be correlated with the regressor of interest and the outcome variable inducing bias in our estimates (fixed-effects estimation). We can furthermore control for common yearly shocks by using time dummies. Secondly, it allows us to explore the timing of the effects (for example lagged effects) as we observe the individuals at several points in time. Finally, in the case of aggregate observations (provinces in chapter 1 and wards in chapters 3 and 4), since the number of spatial units is fixed, using panel data allows us to increase the number of observations used in the regressions which improves the precision of the estimates and the validity of the econometric tests.

2. In order to reduce omitted variable bias, besides using fixed-effects estimation and time dummies, we include several control variables in the estimated empirical specifications which might be correlated with the interest and the outcome variables at the same time. Attention is paid to the issue of "bad controls" (Angrist & Pischke, 2009).

3. When necessary, we make use of instrumental variables techniques. Instrumental variables estimation, if the instruments are valid, allows us to infer causality when there are reasons to believe our estimates could be inconsistent due to endogeneity, reverse causality or measurement error. The instruments that we use and their validity (relevance and exogeneity) are discussed in detail in the main text.

4. In chapters 2 to 4 we use micro-datasets (individual datasets). This has again three advantages. In the first place, even for aggregate outcomes, we construct the variables aggregating up from the individual observations. This allows us to control the way the variables are defined and constructed, which would not be possible if we were using aggregate datasets. Secondly, it allows us to exploit the small geographical scale and to focus on specific observations in order to tackle endogeneity issues (more on this in the relevant chapters). Finally, micro-datasets include a large number of observations and, as discussed above, this has advantages in terms of precision and hypothesis testing.

[4]More details are given in the chapters.

The last steps of the empirical methodology are the interpretation of the coefficients and the testing of the robustness of the results. From the set of different regression results, we choose a preferred specification and we discuss the size and sign of the causal estimates. These are compared to previous findings in the literature and discussed in relation to the theoretical predictions. Finally, we test the robustness of the main results to the use of different data, specifications, inclusion and exclusion of controls, etc. These allows us to improve the external validity of our findings.

II Summary of the chapters

II.1 Chapter 1: Spatial impacts of immigration: Evidence from the Spanish housing market

Chapter 1 investigates the effect of large immigration inflows on the growth of house prices. I study the effect of changes in foreign-born population on the growth of purchase and rental prices. I use the case of Spain during the period 2001 to 2010 to test my hypothesis in the data. The study of the impact of immigration on economic outcomes has produced a large body of empirical literature in the recent years, especially the study of its labour market effects. However, evidence on the effects of immigration on prices is scarcer. The Spanish case has attracted substantial attention of researchers due to the large immigration inflows that the country experienced since the beginning on the 2000s. At the same time, there has been a large increase on house prices, specially purchase prices. These almost doubled in the boom years (2001 to 2008), and slowly decreased since the beginning of the housing crisis.

Immigrants can influence house prices via their effects on housing demand and costs. Neglecting the effects on costs (by providing labour in the construction sector), for a given housing supply and population in the province, an immigration inflow increases demand for dwellings pushing prices up. In the long-run, both population (natives) and housing supply adjust, so the total "net" effect on prices depends further on how supply and natives respond to immigration inflows.

In this chapter I estimate a similar specification to that of Saiz (2007), using first-differences fixed-effect estimation and instrumental variables. My observational units are the 50 Spanish provinces during the period 2001-2010. Provinces are the smallest geographical unit for which data is consistently available and I argue in the text that they are a good approximation to housing and labour markets.

Initially, I regress the lagged annual inflow of foreign-born over the initial population (the "immigration ratio") of a province on the annual growth of house prices. The estimated specification includes controls, fixed effects and trends in order to reduce omitted variables bias. To deal with the endogeneity of immigration inflows due to unobservables which drive both immigration and price changes, I use a instrumental variable strategy. I construct an instrument based on past location patterns by immigrant nationality. The instrument is carefully constructed. In the most demanding instrumental variables (IV) results, I find causal elasticities of prices with respect to immigration of 2.7 for purchase prices and of 1.05 for rental prices. These estimates are comparable to previous findings.

I then test the effect that immigration inflows have on the mobility of natives. Natives might "crowd-out" from the provinces in which immigrants are locating due to competition for amenities and services or discrimination, but they might also move in if immigrants have desirable characteristics for natives (for example if they like diversity or if immigrants provide desirable services for natives). I find that for each 100 immigrants locating in a province, 43 natives moved to the same location. The attraction effect is causally estimated and very robust across specifications.

I argue that overlooking this fact misleads the interpretation of the effect of immigration on prices, as the estimated coefficients using standard approaches like that of Saiz (2007) capture the combined demand effects of immigrants plus natives. I propose a methodology to isolate the impact of immigrant demand from the demand from natives. I regress the total population growth in a province on the growth in house prices and I use exogenous variation on the location of immigrants as an instrument for changes in population. This way, I only use the variation in population growth which stems from immigration and not from increased native inflows. I find coefficients which are around one third smaller than with the standard approach: the elasticities are 1.9 for purchase prices and 0.63 for rental prices. This is consistent with the previous elasticities (2.7 and 1.05) being inflated due the fact that the standard methodology ignores the impact of immigration on natives location. I interpret the elasticities obtained with my methodology as the "net" effects of immigration on prices. Quantitatively, my estimates predict that around one third of the total average annual growth of purchase prices and around one quarter of the total average annual growth of rental prices can be attributed to the effect of immigration on housing demand. I finally test the robustness of these results and investigate the role played by housing supply.

II.2 Chapters 2 to 4: Economic Impacts of Transport Policy

Chapters 2 to 4 assess the impact of transport policy on economic outcomes, using data for Great Britain during the period 1998 to 2008.

Chapter 2: Methodology

Chapter 2 motivates the research, reviews the relevant literature and develops the methodology to estimate the impacts of road construction on aggregate and individual economic outcomes.

Transport policy is an important area of the economic policy of countries. Infrastructure projects require large amounts of investment and their location can be quite controversial. Therefore, the correct evaluation and assessment of transport policy is very important for policy evaluation as a whole, and yet, little robust empirical evidence exists. The aim of these chapters is to assess the impact that road construction has on firms and workers. We focus on road construction due to the importance that motor transport has both in the transportation of goods and in the movement of workers (commuting).

In this chapter I first review and summarise the channels through which transport policy as a whole and specifically road construction affect workers and firms. On the one hand, the theoretical predictions on the net effects of transportation improvements on firms and aggregate outcomes are ambiguous. Better transport infrastructure brings places and people closer together, by reducing transport and commuting costs. This has two effects on the actual size of markets. Firstly, for a given location of firms and workers, effective density increases, as it becomes easier to reach other locations using the improved transportation network. Secondly, new infrastructure increases the attractiveness of locations, which may boost spatial concentration if firms and workers relocate. These effects may reinforce each other and create positive agglomeration spillovers. Besides, improved access to markets also strengthens competition, thus forcing the exit of the less productive firms and thus increasing aggregate productivity. Finally, firms use transport services as a production input, so changes in the supply and relative prices of transport affect the input mix used by the firms and their demand of other inputs, for example labour.

On the other hand, transport policy affects labour markets through different channels. First of all, reduced transport costs bring employers and workers closer together. Transport improvements change commuting costs and hence affect the size of labour markets. For unemployed or inactive workers transport improvements modify search costs and reservation wages, and in this way, they can help reduce frictional unemployment and increase the employment probabilities of those who are jobless. According to the so called "spatial mismatch hypothesis", accessibility to jobs is an important determinant of labour market participation. Secondly, improvements in the transport network can increase the scope of agglomeration economies, as for a given physical distance employers and employees are nearer to each other. Increased competition for jobs and workers (tighter labour markets) can foster productivity, and additionally we could see more and better job matches taking place in these enlarged labour markets. Finally, transport investments could be capitalised into non-labour prices and affect congestion, which in turn affect location workers and therefore residential sorting.

After reviewing the empirical evidence on some of the effects summarised above, I describe the policy context of the paper. I then explain in detail the methodology used to estimate the effects of road construction on workers and wards (in chapter 3) and on individual labour market outcomes (in chapter 4).

To capture the effect of road construction on economic outcomes we construct an index of accessibility to employment using the road network from each location at each year between 1998-2008. This measure is similar to the market-access measures which have been used in the literature to assess the effect of market-size and agglomeration on productivity and wages. This index captures the amount of employment which is reachable using the road network and changes because places become closer together due to road improvements. For a given location it is a weighted sum of the economic size (employment) of the surrounding locations, inversely weighted by the cost of reaching them using the road network. The "costs" are approximated by the optimal driving travel time between the locations. The spatial unit we use is the ward, of which there are more than 10,000 in Great Britain. To calculate employment at the ward level we use the Business Structure Database (BSD), provided by the Office for National Statistics (ONS). This micro-dataset is a register of all the alive establishments in the UK and it allows us to calculate the employment at any spatial scale for all years between 1998 and 2008.

The first step on the construction of the accessibility is the identification of the road improvements and the calculation of travel times. For this, we collected a novel dataset on road construction undertaken in Britain between 1998 and 2007. We combine this data with the 2008 road network to reconstruct the networks "as they were" in the years prior to 2008. We then apply ArcGIS® network tools to compute least-cost (minimum journey time) routes between any pairwise ward combinations. We use these travel times as the costs in the calculation of the accessibility indices. In chapter 2, we also provide abundant description and summary statistics of the travel times and the accessibility measures.

We estimate the effects of accessibility on aggregate and individual economic outcomes using regression analysis. When regressing accessibility to employment on economic outcomes, for example ward number of plants, there are three endogeneity issues we need to tackle. We exploit micro-data and small spatial scale in order to address the sources of bias. Firstly, cross-sectional estimates of the effect of accessibility on economic outcomes could be biased if the model does not capture underlying time-invariant factors (such as place specific productive advantages) that affect both effective density and economic outcomes. We use a fixed-effects estimation method to address this problem. When we use fixed-effects we exploit the within ward/individual variation over time. However, in the fixed effects framework, changes in accessibility can arise because of road improvements and because employment relocates. The second endogeneity issue is that accessibility changes due to relocation of employment may be partly driven by the outcome variable studied or be correlated with the same unobserved shocks. To address this source of bias, we construct an instrument which uses accessibility changes stemming only from the transport improvements. In practice, we construct an alternative accessibility index for which we fix employment at the beginning of the period and we only change the travel times between wards, which vary annually due to road construction. When deciding where to make transport investments, the government could be targeting specific areas because these have specific productivity or wage trends. In order to reduce the possible bias caused by the endogeneity of the placement of the transport investments, we focus on observations which are located within 10, 20 or 30 kilometres of road schemes. This way, we compare individuals and locations which are close to the improvements and we exploit the fact that the impact of the improvements varies considerably even within the distance band. It is quite unlikely that the improvements are aimed at specific individuals or wards within those narrowly defined distance bands, specially after controlling for different growth trends around the schemes.

Chapter 3: Aggregate and firm outcomes

In this chapter we present the empirical results on the effect of accessibility on aggregate and firm outcomes. We test the effect of accessibility on ward employment, number of plants and aggregate productivity. We also investigate its effects on individual outcomes: firm employment, total factor productivity, labour productivity, gross output and average wages.

For the aggregate outcomes, we estimate a linear specification which relates the outcome with accessibility. We use the fixed-effects IV strategy outlined above, and focus on wards within 20 kilometres of road schemes. We use the BSD to calculate ward employment and number of establishments for years 1998 to 2008. We find positive effects of accessibility on total employment and number of plants, and also positive effects on some sectors (specifically construction and producer services). The elasticities are stronger and more robust for the results on the number of plants.

To obtain the ward productivity results, we use an additional data source, the Annual Respondents Database (ARD), also provided by the ONS. The ARD dataset provides balanced-sheet information on a sample of small firms and on the universe of big firms (over 150 employees). We combine this information with additional data on the capital stock of firms. We implement a two-step strategy in which we first estimate yearly ward level productivity shifters and then regress them on accessibility in the second step. We find no effects of accessibility on ward productivity.

We also investigate the effect of accessibility on individual firms outcomes. Overall, when we examine the effect of all major road transport improvements with firm level data focusing on firms and plants that remain in situ before and after the opening of new road links, we find insignificant effects on the employment and productivity of firms. The fact that plant level employment is mostly unaffected by accessibility suggests that the positive ward level employment effect is mainly attributable to increased entry or decreased exit.

Chapter 4: Labour market outcomes

Finally, in chapter 4 we investigate the effects of accessibility on individual labour market outcomes. We study the impacts on individual wages and hours worked (both basic and total), on the probability of being employed and on travel time to workplace. To investigate these outcomes we use two additional datasets.

We use the Annual Survey on Hours and Earnings (ASHE) to test the effect of accessibility on weekly wages, weekly hours worked and hourly earnings, analysing both for basic and total (which includes overtime). Due to data restrictions, in our main regressions we use a panel of employees surveyed between years 2002 and 2008. This data has two main advantages. The first one is that the dataset is a panel, so we are able to control for time-invariant individual unobservables. The second one is that we have precise information of the location of the job and the home of the worker. We use this information to tackle the problem of endogenous spatial sorting due to changes in accessibility.

We regress individual outcomes on accessibility from both work and home, controlling by optimal commuting time, in order to investigate the different effects that accessibility from home or from workplace might have. We use the same distance-band and instrumental variables strategy as in the previous chapter. We use individual fixed-effects and instrumental variables, exploiting the changes in accessibility from work and home for each individual across time. We find positive effects of both accessibility from work and home on wages and hours worked.

However, if workers are sorting spatially in order to take advantage of changes in accessibility, we would not be able to identify the separate effects on labour market outcomes which stem from (endogenous) sorting and those which are due to changes in accessibility for a given location (externalities or spatial competition). Sorting could be an outcome of accessibility or could be due to other unobservable reasons correlated with the labour market outcomes. To overcome this issue we use individual-home-work fixed-effects and exploit the variation over time for individuals in a given location pair. When we do this, only accessibility from work seems to have an effect, and it does so on basic and total wages and on total hours worked.

These findings suggest that, at least partly, the effect of accessibility from home is driven by spatial residential sorting. They also suggest that workers are affected by changes in accessibility from workplace. The outcomes that adjust are those which are "flexible", i.e. wages and overtime. A possible explanation for these results is the existence of some sort of agglomeration externalities due to increased spatial competition in the workplace which is capitalised into wages, and due to higher earnings, workers find it worthwhile to work longer hours. These results are very robust to different specifications and the inclusion and exclusion of control variables.

In this chapter, we also study the effect of accessibility on other labour market outcomes, e.g. employment status and travel time to workplace. To carry out this research we use an additional dataset, the Labour Force Survey (LFS). This survey is a household-based quarterly survey which provides abundant information on some labour market outcomes (specifically employment status). The main drawbacks of this data is that it provides limited information on the location of the job and that, on an annual basis, it is not a panel of individuals.

We use repeated cross sections on individuals for the years 1998 to 2008 and define "pseudo fixed-effects" and estimate a pseudo-panel. In practice, we allocate a different fixed-effect to individuals with a combination of specific set of characteristics (in our case gender age group and ethnicity). We exploit the variation of different individuals around the mean of a "representative" individual, defined by the set of characteristics. We include additional controls and time dummies to reduce omitted variable bias. Once again, we use the instrumental variables and distance bands to tackle endogeneity issues.

With the LFS data, we can only estimate the effect of accessibility from home, as we do not have information on where the individuals are working or looking for jobs. We find no significant effects of accessibility on the probability of being employed, and the coefficients become quite imprecise when we become particularly rigorous with the endogeneity issues. We find some evidence on accessibility from home on reducing travel time to workplace, which points in the direction of accessibility affecting labour market outcomes through reducing commuting costs.

III Contributions

This thesis contributes to the empirical spatial economics literature on three dimensions:

III.1 Theoretical

The research carried out contributes to the theoretical debates on the effect of immigration on housing markets and on the effects of transport policy on economic outcomes.

Chapter 1 thoroughly discusses the different theoretical channels through which an increase in foreign-born population might affect house prices. An increase in foreign-born population in a given location could affect prices via (increased) demand and via (reduced construction) costs. If we focus on the effects which work via changes in housing demand, using Saiz (2007) terminology, we can distinguish between short-run and long-run effects. In the short-run, total population and housing stock of a regions are fixed. In the long-run housing supply (through construction) or housing consumption (through changes in housing density) might change. In addition, native population might relocate spatially due to the immigration inflow into the region. Therefore, the "net" effect of immigration depends on how housing supply and native population react to changes in foreign-born population. I provide a careful discussion of these channels which adds to the existing discussion in the literature and I design the empirical strategy based on the insights emerging from this discussion.

Chapters 2 to 4 investigate the effect of transport policy on economic outcomes. We assess the impact of road construction, measured using an accessibility to employment index, on aggregate and individual (workers and firms) outcomes. We study numerous outcomes: employment, number of plants, productivity, wages, hours worked, etc. Given the large number of outcomes we study, there are numerous predictions of the effects of transport policy on them and the different channels are intertwined. Chapter 2 includes a comprehensive discussion of these channels and relates them to the empirical evidence. The different channels are connected to different strands of the urban economics literature (agglomeration economies, spatial mismatch hypothesis, spatial competition). This discussion contributes to the existing literature by not solely discussing partial-equilibrium outcomes but by setting out the interlinked mechanisms thought which transport policy affects firm, worker and aggregate outcomes.

III.2 Methodological

Both parts of the thesis also make substantial methodological contributions. Chapter 1 develops a methodology to isolate the "net" effect of immigration on aggregate house prices. Standard approaches to estimate these effects fail to take into account the effect that immigration inflows have on native population mobility. In fact, my estimates predict that an inflow of 10 foreign-born into a province induces approximately 4 natives to locate into the same province. Therefore, the estimated reduced-form coefficient of a regression of immigration inflows on price growth would be capturing both the demand effect stemming from immigrants plus the induced demand effect by natives attracted to the same locations in which foreign-born are setting in. In this chapter, I propose a methodology to isolate the "net" demand effect of immigration on prices from the total demand effect. In practice, I estimate the effect that immigration has on house prices "via its effects on population changes". This allows us to gain insights both on the short and long run effects of immigration inflows on province house prices underlined by Saiz (2007).

Chapters 2 to 4 investigate the impact that transport policy might have on economic outcomes. We focus on road construction, making two major methodological contributions. The first one is the construction of an accessibility index which captures the impact that road construction has on the amount of employment which is reachable from a given location. We construct this index by combining the economic size of locations (employment) and the optimal travel times between them. These latter are calculated using an original dataset on detailed road schemes undertaken in Great Britain during the years 1998 to 2008 and combining this information with data on British major roads networks. This way, we are able to construct major roads networks in Great Britain for every year in our period of analysis, which is not readily available. To these data we apply ArcGIS® network tools to calculate how long it takes to cross any link in the networks and to compute least-cost (minimum journey time) routes between any pairwise ward combination. To our knowledge, this is the first time that optimal travel times by road between wards are calculated for Great Britain for every year in the period 1998-2008. In fact, our data allows us to potentially calculate optimal travel times between any pair of locations, no matter the spatial scale. Moreover, it is the first time an accessibility index of this nature is constructed at such a small spatial scale on a yearly basis for such a long period of time (11 years).

The second methodological contribution is the development of an identification strategy which allows us to tackle the different endogeneity issues involved in estimation of the effect of accessibility on economic outcomes. This method allows us to focus on the changes in accessibility stemming from road construction, so we can separate the effect of transport policy from the effect of spatial relocation of employment in the variation of accessibility. This strategy is explained in detail in the text.

III.3 Empirical

The research carried out in this thesis also makes several empirical contributions.

First of all, the empirical strategy is carefully developed to be able to imply causality in the estimates. For this purpose, a substantial number of specifications are estimated and a series of robustness checks are performed in order to check the validity of the results. Special attention is given to the correct identification of the coefficients. Thus, we are certain that the estimated effects are quite robust and, when possible, are comparable with previous estimates of the literature.

Secondly, the research questions are picked in order to fill relevant gaps in the literature. The existing evidence is critically reviewed in order to identify the areas in which robust evidence is needed. Specifically, chapter 1 answers a question for which empirical evidence already exists in the literature (what are the effects of immigration on house prices?) but does so in a more complete and informative way. I study the effects on increases in foreign-born population on both purchase and rental prices, and in the case of rental prices, this is new for the Spanish case. As explained above, I also develop a methodology to be able to differentiate between total and immigration-induced demand effects, which contributes to the empirical literature of spatial effects of immigration in general. Moreover, I estimate the causal impact of immigration on natives location for the Spanish case, which to my knowledge had not been done before.

In the case of chapters 2 to 4, they also make substantial empirical contributions. The estimates of the effect of accessibility on productivity and wages add to existing evidence of the effect of agglomeration and market access on these outcomes (see Mion & Naticchioni, 2009; Hering & Poncet, 2010; Martin et al., 2011, for some examples), although they focus on one specific channel which affects the level of agglomeration, e.g. road construction. There exists no previous evidence on the effect of road construction on individual economic outcomes (firm and workers). Moreover, we study how much road construction affects accessibility, which is also novel. Finally, my findings also shed light on some of the predictions of the spatial mismatch literature (see Inhanfeldt, 2006, for a review) and of the thin labour markets theory (Manning, 2003).

Chapter 1

Spatial impacts of immigration: Evidence from the Spanish housing market

1.1 Introduction

The study of the impacts of immigration has been a very active area of research in the last 20 years. Large immigration inflows have substantial effects on the spatial distribution of population within a country. Location choices of the foreign-born directly changes the composition and size of the population residing in a given area. Immigration inflows also influence the location decisions of natives, indirectly changing the population size of the different locations. An inflow of population affects the labour force of an area, and therefore impacts not only average wages and employment rates, but also their distribution. Changes in the labour market conditions will, as a result, affect other economic aspects such as productivity, skills composition, and ultimately growth and welfare. Not only immigrants affect the production factors, they also consume amenities and housing services in the places they locate and this way they influence the spatial equilibrium. As a consequence, the study of effects of immigration on the housing markets becomes central from a urban economics point of view.

Since seminal papers such as Borjas (1995), most of the theoretical and empirical contributions on the study of the impact of immigration in receiving regions have originated from the analysis of their labour market effects[1]. This strand of the immigration literature has focused on the analysis of the impact on natives wages and on the (individual) distribution of wages. The existing evidence is mixed: gener-

[1] Hanson (2008), Dustmann et al. (2008b), Longhi et al. (2009) or Pekkala-Kerr & Kerr (2011) provide recent reviews of the literature.

ally negative effects on the short run and positive on the long run have been found (Card, 2007). In any case the estimated effects have been very small and the debate on the labour impact of immigration is far from being over (Ottaviano & Peri, 2006; Dustmann et al., 2008a; Ottaviano & Peri, 2008a; Borjas, 2009)[2].

The spatial dimension of immigration has however not been sufficiently accounted for in most of the labour economics literature (Cushing & Poot, 2003). Increased spatial competition on the consumption of goods, amenities and housing services may push prices up; the sign and size of the impacts also depend on the response of the supply of goods and on any induced relocation of natives following the foreign-born inflows. The net effect on prices would then be the result of changes in the demand side (positive effects through increased demand) and of changes in the supply side (through low-skilled immigrants pushing down wages and costs). A small number papers have provided evidence on the effect of immigration on (consumption) goods prices (Lach, 2007; Cortés, 2008; Frattini, 2008; Zachariadis, 2011). They have mostly found negative effects of an increase of low-skilled immigration on (generally immigrant-labour intensive) goods. The same supply-demand mechanisms discussed above would operate in the analysis of the effects of immigration on house prices. Previous evidence for the US (for example Saiz, 2003, 2007; Ottaviano & Peri, 2011) has generally found positive causal impacts of immigration on both rents and prices[3].

In order to provide new evidence on the existence of the aforementioned effects, this paper studies the impact of the large immigration inflows on house prices using Spanish data. Between 2001 and 2010 both the number of foreign-born residing in Spain and house prices significantly increased, providing a suitable setup to gain further insights on the impact of immigrations on prices. Motivated by the substantial size of the immigration inflows, the number of empirical works analysing the impact of immigration in Spain on various economic outcomes has increased in recent years. Most of the papers have focused on the labour market impacts (Bentolila et al., 2008; Carrasco et al., 2008; Amuedo-Dorantes & de la Rica, 2008a,b; Gonzalez & Ortega, 2010), but a number of papers have studied other aspects like the effect of immigration on output mix (Requena et al., 2009), trade (Peri & Requena, 2010), productivity (Kangasniemi et al., 2009), or even crime (Alonso-Borrego et al., 2011).

A handful of recent works have also provided some evidence of the impact of immigration on house prices in Spain. Talavull de la Paz (2003) explores their different determinants using a sample of Spanish cities during the period 1989 to 1999. She

[2]The study of immigration impacts has also analysed on other aspects, for example regional convergence (Ozgen et al., 2009), innovation (Gandal et al., 2004), (Gauthier-Loiselle & Hunt, 2009), productivity (Peri, 2009) or crime (Buonanno et al., 2011).

[3]Other studies are Greulich et al. (2005), which analyses welfare effects (real rents) or Stillman & Maré (2008), which provides estimates for New Zealand.

investigates the role of population and economic activity specialisation as explanatory variables of cities prices differentials. She finds that population is strongly significant in explaining house price levels while economic structure does not appear to have any significant effect on house prices. Sosvilla-Rivero (2008) analyses the effect of immigration for the regions during the period 1995-2007 and assesses the over-valuation of the house prices with respect to economic fundamentals. He finds that almost half of the over-valuation can be attributed to immigration flows, which he interprets as a positive relationship between immigration and prices. Gonzalez & Ortega (2009) find important positive causal effects of immigration inflows on both house prices and dwelling construction. They focus on the period of "housing boom" in Spain (1998-2008) and use a similar methodology to Saiz (2007). Their paper is the closest to the present one, but their period of analysis is different and they do not study the effects on housing rents. Finally, García-Montalvo (2010) explores the role of land regulation and immigration on Spanish municipalities during the period 2001 to 2005 but, conversely to the other studies, he finds no effect of immigration inflows using a long-differences instrumental variables estimation.

The scope of the present paper is related to the aforementioned. To study the effects of immigration on house prices I exploit a panel of Spanish provinces (NUTS3) for the period 2001-2010. Fixed-effects and instrumental variables estimators are used in order to infer causality between the immigration inflow and the evolution of average house prices. I provide new evidence of the impact of immigration on the average purchase prices and novel evidence of the effect on rental prices.

One of the main contributions of this paper is the analysis of the effects of immigration on Spanish local rental prices, which, to my knowledge, has never been done before. I consider rents to be an important indicator of the demand for housing (from immigrants). Data analysis based on the National Immigration Survey 2007 shows that the housing tenancy choice of most immigrants (70%) is renting, not owning. Therefore, a substantial part of the impact of immigration on purchase prices could be indirect. Moreover, housing is not only a consumption good, but could also be considered an investment asset. In this sense, foreign investment in Spanish real estate has been substantial in the last years (Rodríguez & Bustillo, 2008). These two reasons make the analysis of rental prices an important complement to the analysis of purchase prices when interested on the impact of immigration on housing services.

The period of analysis covers a subperiod of high boom (2001-2008) and bust of the housing markets (2008-2010), which provides sufficient variation to adopt a very demanding empirical strategy. I find estimates of the effects of immigration on house prices comparable to previous findings (Saiz, 2007). The relationship between immigration location and native location is also explored, using the empirical test suggested by Peri & Sparber (2011). I estimate a significant and positive causal relationship between natives and immigrants location choices, which is very robust across specifications. Motivated by this finding, I provide novel estimates which use the variation in population attributable to exogenous immigration location choices. If natives are co-locating in the same locations as immigrants, because these provide some desirable services or because natives like diversity, standard estimates might be capturing both the demand increase from immigrants but also the induced demand increase from relocated natives. I argue that the estimates obtained using my technique are non-biased estimates of the net effect of population changes due to immigration. These turn out to be around 30 to 40% lower than those obtained with the standard methodology used for example in Saiz (2007) or Gonzalez & Ortega (2009). The estimated elasticities using this methodology are around 0.6% for rental prices and between 1.3 and 2% for purchase prices. These results are robust across specifications, to different data sources and to the use of different definitions of the instrument.

Finally, I explore the role played by housing supply in mitigating the house prices growth attributed to increased demand from immigrants. I find no significant effect of the growth of housing stock on prices, conditional on changes in demand induced by immigration. This could be due to financial constraints that immigrants face when accessing housing ownership or due to substantial numbers of non-occupied houses[4].

The rest of the paper is organised as follows. Section 1.2 describes the empirical strategy: the empirical specification is explained in 1.2.1, some issues related to the estimation of causal effects of immigration on house prices are discussed in 1.2.2, the identification strategy is explained in 1.2.3 and finally the data sources and some descriptive statistics are provided in 1.2.4 and 1.2.5. Section 1.3 discusses the results and the robustness tests. Finally, section 1.4 contains the conclusions and the discussion of the limitations of the analysis.

1.2 Empirical methodology

1.2.1 Empirical specification

In order to estimate the causal effect of changes in foreign-born population on the growth of house prices (purchase and rental), I use a linear empirical specification

[4]In fact, the estimates of the Spanish Buildings Census 2011 for stock of non-occupied dwellings (which includes non sold dwellings) is around 20%.

similar to Saiz (2007). It takes the form:

$$\Delta\log(r_{i,t}) = \beta\frac{FBinflow_{i,t-1}}{population_{i,t-1}} + \lambda_t + \gamma_r + \phi' Z_i + \delta'\Delta X_{i,t-2} + \varepsilon_{i,t} \tag{1.1}$$

The geographical unit of observation are the 50 Spanish provinces i, which are grouped into 17 regions r.[5] t denotes time periods (years). $\Delta\log(r_{i,t})$ is the change of the natural logarithm of housing purchase or rental prices in province i during year t, the immigration ratio during $t-1$ is $FBinflow_{i,t-1}/population_{i,t-1}$, λ_t are time fixed-effects, γ_r are regional fixed-effects, Z_i is a matrix of province time-invariant attributes and $\Delta X_{i,t-2}$ is a matrix of province time-varying controls. Finally, $\varepsilon_{i,t}$ is a random shock.

The independent variable of interest is the immigration ratio: it is defined as the inflow of immigrants into province i during a given period divided by the population in the province at the end of the previous period (original province's population). The inflow of immigrants during $t-1$ is calculated as the change in the foreign-born population between January $t-1$ and January t. Population in $t-1$ denotes the stock of total residents (natives and foreign-born) at the end of period[6] $t-2$. Using an immigration ratio instead of gross inflows as the measure of "immigration" has three advantages. For a given housing stock, the changes in demand which affect house prices depend on the number of immigrants moving into the province and on the demand from existing residents. Bigger regions, in terms of population, would be able to absorb larger numbers of immigrants, and could have different house price growth dynamics than less populated regions. Standardising the immigration inflow by the original population size of the province allows us to take into account the "relative" size of the immigration inflow, which better captures the effect of immigration on housing demand. By using the ratio we also eliminate any unobservables that might equally affect both the numerator (immigration inflow) and the denominator (original province's population). Finally, it allows us to interpret the coefficient β as an elasticity: a 1% increase in the ratio has a β% effect on the change in prices.

The immigration ratio is lagged one period with respect to the growth rate of prices. For example, I expect a change in the immigration ratio during 2001 to have an effect in prices the following year, i.e. between 2001 and 2002. The span of time between the arrival of immigrants into a region and the reaction of housing services prices is undetermined. Just after arrival into a province, immigrants start consuming housing services immediately, as they need accommodation. It is likely that the

[5] Provinces correspond to the European NUTS3 and regions to NUTS2. I exclude the African territories for their historical particularities and the lack of reliable data.

[6] The source of the population data dates the population numbers (total, foreign-born and natives) on the 1^{st} of January.

initial housing tenure of just-arrived immigrants is renting, specially if they are economic immigrants and they move into a location looking for a job. Therefore, the effect of immigration inflows on rental prices could be contemporaneous or lagged. The effect on housing purchase prices is more likely to be lagged. Unless immigrants settle into a region because of residential reasons (which is the case for a small fraction of the immigrant arrivals since 2001), immigrants probably take some time to accumulate enough wealth to buy a property. The main specification uses lagged immigration ratio with respect to the changes in prices, but I also investigate the contemporaneous relationship as a robustness test.

The first-differences setting eliminates any unobservable province characteristics which might be correlated with the level of house prices and the level of foreign-born population in the province. Time fixed-effects λ_t control for common shocks affecting the growth of prices of all provinces in Spain in a given year (for example, a tax deduction on mortgage payments, a subsidy to renting or a better financial climate). There could still exist some unobservable factors at the region or at the province level which are correlated with the changes in purchase/rental prices and changes in foreign-born stocks (numerator of the immigration ratio) and which are biasing the estimation of β. To reduce this bias, I add region fixed-effects γ_r (and regional trends $\gamma_r * t$ in some specifications). These fixed-effects control for time-invariant regional characteristics which might affect the prices growth and the immigration ratio and which are not common to the whole country. Alternatively I include province fixed-effects (γ_i). These control for unobservables at the province level which are correlated with changes in prices and in the immigration ratio and correspond to a first-differences fixed-effects estimation.

Vector Z_i contains time-invariant province attributes. They control for the fact that provinces with different levels of the time-invariant characteristics might have different growth trends in the levels house prices and in the stocks of foreign-born population. Given that region fixed-effects (γ_r) are also included, the province attributes control for differential growth trends of the provinces around their common regional trend. The matrix includes geographical characteristics (coast dummy, length of the coastline, surface of the national parks), weather (average temperature and average rain precipitation in January) and beginning of the period levels of several amenities (number of restaurants and bars in 2000, number of retails shops in 2000, number of doctors in 2000 and a comparative index of the importance of the tourism sector in 2000). When I use province fixed-effects, I control for all time-invariant attributes, so Z_i drops.

The vector $\Delta X_{i,t-2}$ contains time-varying characteristics (in changes). Even if our aim were to reduce the omitted variable bias, this would not be the case if the variables included in $X_{i,t-2}$ were "bad" controls, in other words, variables that could well be outcomes variables in equation (1.1) (Angrist & Pischke, 2009). To try to mitigate the effect of bad controls, I use a lag with respected to the immigration ratio, so the variables are measured one year before the immigrants locate in the province. Hence, I use the changes in the variables during $t-2$, one period before the inflows ($t-1$) and two periods before the change in prices (t). I control for the growth in gross domestic output (GDP) and the changes in the unemployment rate. Richer provinces which are growing faster and employing more people could be attracting more immigrants and thus could also have higher growth in house purchase and rental prices. I also control for changes on the number of credit establishments and on the share of saving banks because they could have affected the availability of credit, which might have pushed purchase house prices up by influencing housing tenure decisions (Cuñat & Garicano, 2010).

If we believed there is high time dependence on both the immigration ratio and on the growth of purchase/rental prices[7], lagging the controls one period with respect to the immigration ratio would not be enough to overcome the problem of bad controls. In this case, we would not be sure that our control, for example GDP growth, was not directly determined by prospects of future changes in prices and immigration. Given that equation (1.1) controls for regional/province fixed-effects and time fixed-effects, the time-varying controls would only be eliminating the bias induced by annual changes in the province characteristics which are not captured by these fixed-effects and which are affecting the change in prices and the change in immigration ratio at the same time. In other words, a annual shock in GDP in province which is not common to all province in Spain and which is different from the average growth in the period. These changes are likely to be small, so the reduction in the bias caused by the introduction of time-varying controls is likely to be small (which is the case, as explained in section 1.3). The empirical results are very robust to the exclusion of $\Delta X_{i,t-2}$, and the estimated β coefficient is very similar with and without time-varying controls. The main results are obtained including time-varying controls, but the qualitative conclusions would remain unaltered if we excluded them.

1.2.2 How do immigrants affect house prices?

Correct identification of coefficient β would yield insights into the causal effect of immigration on house prices. If we consider changes in foreign-born population as a source of population growth (Card, 2007), large immigrants inflows would be expected to have a positive impact on the evolution of housing purchase and rental

[7]The correlation between the immigration ratio and the lagged immigration ratio is 0.60, between changes/lagged changes in purchase prices is 0.76 and changes/lagged changes in rental prices 0.50.

prices. The economic intuition behind this is a simple demand-supply result. For a given level of population in the region, after a large immigration inflow, increased competition in housing markets forces both newly arrived immigrants and stayers to bid higher to buy or rent a property. Given supply, a positive immigration inflow into a region could be translated into an increase in demand of housing services, thus pushing up prices and rents.

This is the intuition behind the model developed in Saiz (2007). Increases in foreign-born population in a given location increases total population and then pushes demand and prices in the short run. In the long run (net effects) we need to take into account the effect of the changes on housing supply (construction), on housing consumption (density) and on the mobility of natives or previous residents (displacement). I discuss the implications of considering these aspects on the interpretation of coefficient β below.

Table A.1 shows that construction of new private dwellings between 2001 and 2010 was very high, of almost 5,000,000 houses. Given this large number, we would expect increased supply to, at least partially, mitigate the rise in prices caused by the increase in demand. In this paper I try to account for the effect of housing supply on house prices. Including supply changes in the estimation of equation (1.1) as an extra control is very problematic, because even if lagged, housing construction is very likely to be a "bad" control for the reasons explained before. There are difficulties to find a perfect instrument for housing construction with the currently available Spanish data, hence conclusions drawn from these results must be taken with caution.

In addition, we could argue for a limited effect of construction via a high number of non-occupied dwellings. According to Census data, in 2001 15% of the housing stock was empty. Although no official number exists, in 2010 the government recognised the existence of around between 700,000 empty houses, while non-official statistics quantify this number between 1.7 and 3 millions in 2008 (at the peak of the boom) and at least 1.5 millions in 2010[8] (after the crisis started). This would suggest that not all the new construction was occupied by new tenants or owners (maybe because of credit restrictions or expectation on future growth in prices), so increased supply did not help to alleviate price growth.

Immigration can also have an effect on prices through input costs, by pushing down (construction) costs through lower wages (Gonzalez & Ortega, 2009; Zachariadis, 2010). According to the Wage Structure Survey (National Institute of Statistics), in 2007 wages paid on the construction sector were around 10% lower than the average wages, and the average paid wages to non natives were much lower than those paid to natives (20% lower for EU citizens, 55% for Latin American, 65% for East

[8] SEOPAN and *Anuario Estadístico del Mercado Inmobiliario Español* (RR de Acuña & Asociados).

European and 47% for the rest of the world, including North African immigrants). On the other hand, immigrants employment is concentrated in the construction sector.

Table A.2 illustrates this fact. It displays the percentage of workers in the construction sector over total employment, for the total labour force (immigrants plus natives), for the native labour force and for the immigrant labour force. The last row shows the proportion of immigrants over total labour force. In 2008, at the peak of the housing boom, over 20% of the immigrants that work do so in the construction sector, compared to the 12% of the total. In 2010, at the end of the period of analysis (when foreign-born workers still represent more than 13% of the labour force), immigrants are still over-concentrated in the construction sector. If we compare these proportions to the numbers at the beginning of the period, we notice that while in 1998 immigrants and natives were equally concentrated in the construction sector, by the end of if immigrants disproportionably work in construction. Putting the lower wages and the concentration of immigrants in low-paid jobs together, it can also be argued that immigrants could negatively affect the growth of house prices through pushing down the labour cost in the housing construction sector. However, the lack of good quality data to test the cost hypothesis prevents us from drawing strong conclusions.

Housing density remained relatively stable during the period. Table A.1 also shows the ratio of population over total (private) housing stock[9] during the period 2001 to 2010. Even if we cannot draw definitive conclusions, mainly due to the lack of reliable data on housing vacancies, these numbers suggest that, if anything, intensive construction kept the ratio of houses/population relatively stable (or even increased it) over the period of analysis.

The net (long-run) effect of immigration inflows on any local economic outcomes depends on what the literature has called "native displacement". This issue gained renewed interest after the publication of Borjas (2003). This paper criticized regional immigration studies of the labour market impacts of foreign-born inflows, claiming that the United States (US) works as a single labour market and that the existence of displacement hampers the estimation of regional effects. Any estimated regional effect of an inflow of immigrants would be the net results of changes in labour supply which results from the inflows plus any changes from natives relocation. The existence of native displacement has been used as an explanation for the lack of robust estimates of the impact of immigration on wages across US labour markets.

The relocation of population across regions within a country would hinder the identification of area-level effects, as the effects would dissipate throughout the

[9] Private housing in Spain represents around 90% of total stock.

country[10]. This same issue would apply to the estimation of the effect of immigration on house prices. For a given housing stock, immigration inflows increase prices through increases in housing demand in the short run. Consequently, total changes in housing demand in the long run depend on how and if the natives relocate spatially after or at the same time as the immigrants arrive.

In a recent article, Peri & Sparber (2011) review the existing evidence of native displacement in the US and, using simulated data, they test the relevance of the tests which have been previously performed in the literature. They conclude that, based on the existing tests, there is no robust evidence in favour of the existence of native displacement in the USA. Due to the traditionally low native mobility in Spain (Decressin & Fatas, 1995), we would expect natives to react little to immigration inflows. Yet, Spain had never experienced immigration inflows of the magnitude of those of the last years, so this setting provides an interesting scenario to test internal mobility[11].

Peri & Sparber (2011) suggest to test the native displacement hypothesis using a variation of the test proposed by Card (2007)[12]. We can use a "native ratio" in the left-hand-side of an specification similar to (1.1) and estimate:

$$\frac{natives\_inflow_{i,t}}{population_{i,t}} = \alpha \frac{FBinflow_{i,t}}{population_{i,t}} + \lambda_t + \gamma_r + \phi' Z_i + \delta' \Delta X_{i,t-1} + \varepsilon_{i,t} \tag{1.2}$$

where the variables in the right-hand-side denote the same elements as in (1.1). The sign and size of α would inform us about the relationship between immigration inflows and native relocation. If the estimated α is negative this would indicate that natives are leaving the regions where the immigrants locate: displacement would be complete if $\alpha = -1$ or less than proportional if $-1 < \alpha < 0$.

It is commonly assumed that immigrants would displace natives from the regions they migrate into. Native population might move out from areas where immigrants move in for several reasons; for example competition in the labour market, competition in the consumption of amenities or public goods or even segregation. It could also be the case that immigration inflows have no effect on natives location ($\alpha = 0$), for example if immigrants specialise in different tasks than natives (Ottaviano & Peri, 2008b), so they would not compete for the same jobs, or if they

[10] Numerous papers have investigated the relationship between immigration and natives mobility, for example Card & DiNardo (2000), Card (2001), Hatton & Tani (2005), Borjas (2006), Card (2007), Cortés (2008) and Mocetti & Porello (2010).

[11] In their report, Fernández-Huertas et al. (2009) provide some non-causal evidence on the relationship between immigration and native location. They find positive correlations although they claim that the size is negligible to have any considerable impact on the estimation of local effects of immigration. Given my results, in section 1.3.2 I argue differently.

[12] Card's specification uses population growth as the left-hand-side variable, which includes both natives and immigrants.

consume different goods (Mazzolari & Neumark, 2011). In this case we could expect immigrants not to affect natives location decisions. Immigrants could even attract natives to the location they settle in. This could happen if immigrants are attractive to natives because they provide cheaper labour-intense goods (as suggested by Cortés, 2008) or because they generate positive externalities on natives wages or rents (Ottaviano & Peri, 2006). Another plausible explanation for the existence of "attraction" is that there exist unobservable time-varying characteristic or amenity of a region which attract both natives and immigrants and that we are unable to capture in our econometric model.

In his study of the effect of immigration on American rents, Saiz (2007) claims that if native outflows completely off-set immigration inflows, we would expect the increase in housing demand by immigrants to be completely balanced out by a decrease of housing demand from natives. The total effect, and therefore the parameter β in equation (1.1), would be zero. If natives leave the area in greater numbers than immigrants enter, β would be negative because it would mean total housing demand (for a given supply) is decreasing. He suggests that finding a positive local effect of immigration in rents allows us to reject the complete native displacement in the labour market[13].

If our aim is to draw conclusions on the (net) effect of immigration on prices, it is therefore essential to estimate the effect of immigration inflows on native location decisions. If no causal relationship exists between immigration location and native re-location, then we can be quite certain that we are estimating the effect of (increased demand from) immigration on prices. But if a sizeable causal relationship exists, we need to be more cautious about the interpretation of our results. In this paper I estimate the displacement hypothesis with Spanish data and I propose a methodology to investigate the long-run effect of immigration on prices, specifically, the effect of immigration on prices through its effect on total population changes. This is discussed in detail in section 1.3.2.

Finally, other minor issues on the estimation of β should be considered, which are discussed in the appendix (section A.1).

1.2.3 Identification strategy

As detailed in section 1.2.1 above, the first step to achieve correct identification of the effects of immigration on prices is to include region and province fixed-effects. These control for time-invariant unobservables at the region or province level correlated at the same time with the immigration inflows and the growth in

[13] As discussed in section A.1, the total effect also depends on the relative income/access to credit of natives and immigrants.

prices. The fixed-effect estimator exploits the variation in price changes and immigration inflows within provinces across time around the average changes during the period 2001-2010 (net of common national shocks as we are including time dummies too). We need a substantial amount of variation to be able to identify the β parameter precisely. Because our period of analysis covers both a period of high growth (2001-2007) and of economic crisis (2008-2010), there is a fair amount of variation in the data to be able to identify the parameter of interests even after applying first-differences and including year and province fixed-effects in most specifications.

Nevertheless, even after including province fixed-effects, consistent estimation of β still requires the regressor of interest to be uncorrelated with the time-varying part of the error (local time-varying shocks affecting price growth and immigrant location at the same time). If this is not the case, we would still be finding inconsistent estimates of the coefficient we are interested in. There is no prior on the direction of the bias. The estimated β would be upward biased if immigrants are going to provinces with positive shocks or better economic prospects, while it would be downward biased if, for some reason, immigrants locate in province in which prices are growing slower[14].

In order to infer causality on the relationship between immigration and house prices growth, I estimate equation (1.1) using an instrumental variables approach. I construct the instrument adopting the "shift-share" methodology, which has extensively been used before, for example by Card (2001), Ottaviano & Peri (2006) or Peri (2009). It exploits the fact that immigrants tend to disproportionately locate in areas where immigrants from the same nationality/ethnicity have located before, to take advantage of social and economic established networks. I use historical location patterns (1991 for most specifications) to predict current location patterns, and I use these predicted inflows as an instrument for the actual inflows.

The regressor of interest is the immigration ratio: immigration inflow during $t-1$ divided by total population (foreign-born plus natives) at the end of $t-2$. The population data is dated on the 1^{st} of January. Denoting foreign-born population as $FBstock$ we can express the immigration ratio for province i as:

$$\frac{FBinflow_{i,t-1}}{population_{i,t-1}} = \frac{FBinflow_{i,t-1}}{FBstock_{i,t-1} + natives_{i,t-1}} = \frac{FBstock_{i,t} - FBstock_{i,t-1}}{FBstock_{i,t-1} + natives_{i,t-1}} \tag{1.3}$$

I construct the instrument following Peri (2009). I predict the stock of foreign-born $FBstock_{i,t-1}$ using previous location patterns by nationality, and I use this prediction to calculate the immigration inflow of the numerator (calculated as change in

[14] In a similar exercise to the present one, Gonzalez & Ortega (2009) claim that, given that they are controlling for economic prospects of the regions, it is expected than immigrants go to provinces where prices grow slower, so we expect the parameter to be downward biased.

the stock) and I also use it in denominator as part of total population. Other papers on the impact of immigration on housing (Saiz, 2007; Gonzalez & Ortega, 2009) use a similar instrument but the denominator uses actual population data, not including the predicted foreign-born in the denominator, which is incorrect.

I denote provinces with r (i is the specific province for which we are calculating the share and R is the 50 provinces in Spain), time periods with t, nationalities or ethnic groups with n (N being the total number of nationalities) and the years I use to calculate the historical location patterns as *base*. A list of the nationalities used (119 groups) appears in table A.4. To impute the immigrant population in each province by nationality of origin, I first calculate, for each province and each nationality, the share of immigrants (over the total number in Spain) that were located in that region in the base year. The base year is the reference year of "past" location patterns, which is normally some years before the start of our period of analysis. The share is defined:

$$share^n_{i,base} = \frac{FBstock^n_{i,base}}{\sum_r^R FBstock^n_{r,base}} = \frac{FBstock^n_{i,base}}{FBstock^n_{Spain,base}} \tag{1.4}$$

This share is the proportion of immigrants located in a particular province i over the total immigrants from the same nationality located elsewhere in Spain, at some past period of time *base*.

The imputed foreign-born stock of a specific nationality n in province i at time t, $imp\_FBstock^n_{i,t}$, is calculated allocating current total national stocks based on the historical share:

$$imp\_FBstock^n_{i,t} = share^n_{i,base} * \left(\sum_{r \neq i}^R FBstock^n_{r,t}\right) \tag{1.5}$$

The current national stock of nationality n, $\sum_{r \neq i}^R FBstock^n_{r,t}$, is calculated summing the stock of foreign-born of that nationality in all provinces in Spain except i. I exclude i to avoid using the stock I am trying to instrument for in the construction of the prediction of foreign-born.

To calculate the imputed total (all nationalities) foreign-born stock in province i at time t, we sum across nationalities:

$$imp\_FBstock_{i,t} = \sum_n^N \left(imp\_FBstock^n_{i,t}\right) \tag{1.6}$$

I use the change of the imputed total foreign-born population to calculate the imputed total inflow of immigrants (recall that population data is dated 1^{st} January). This imputed value is divided by the imputed population (imputed foreign-born plus natives) in province i at the beginning of the period $t-1$ in order to obtain the first instrument for the immigration ratio as defined in expression (1.3). The instrument is constructed as follows:

$$IV1\_ratio_{i,t-1} = \frac{(imp\_FBstock_{i,t} - imp\_FBstock_{i,t-1})}{imp\_FBstock_{i,t-1} + natives_{i,t-1}} = \frac{imp\_FBinflow_{i,t-1}}{imp\_population_{i,t-1}} \quad (1.7)$$

For this instrument to be valid it has to be sufficiently correlated with the immigration ratio but uncorrelated with the local shocks that affect house prices variations, conditional on the controls and fixed-effects. The relevance of the instrument can be assessed by the value of the F-statistics of the instrument in the first stage of the 2-stage-least-squares (2SLS) regressions, and additionally by using under-identification and weak identification tests. The exogeneity of the instrument depends on several conditions. Given the way the predicted foreign-born stock (1.5) is constructed we need that[15]:

1. The unobserved factors determining the location of immigrants in one province with respect to another in the base year (1991) is uncorrelated with the relative economic prospects of the two provinces during the period of analysis (2001-2010). In other words, immigrants in 1991 did not come in the prospects of future growth during the 2001-2010 decade.
2. The only channel through which foreign-born geographical distribution in the base year (1991) affects current changes in house prices is through its influence on shaping the current immigrants location patterns (exclusion restriction).
3. The total (national) flow of immigrants in a given year (second term in the interaction) has to be exogenous to specific unobservable province local shocks.

The choice of the base year determines the validity of conditions (1) and (2) but also the strength of the instrument. If the base year is very close to t, the instrument would be strong but its exogeneity can be jeopardised. If the base year is very far from t, it is more likely that the instrument is exogenous, but it may not be strong enough. For the main results, the IV is computed using data for the 1991 Census (foreign-born by country of nationality). In 1991 there was a sufficient stock of foreign-born in each province from each nationality to assure that our instrument is strong. Conditions 1 and 2 require that that location choices in base year are not driven by factors correlated to current changes in house prices (Saiz, 2007). These conditions are quite likely to be valid given that between 1991 and 2001 there was an important economic crisis (1992-1993) followed by economic recovery and growth (from 1997). We can assume immigrants were not able to predict future shocks (not captured in the province nor in the time fixed-effects) ten years before our period of

[15] Adapted from Cortés (2008).

analysis starts[16].

The validity of condition 3 depends on the way the second term of the interaction is constructed. First, to avoid using the inflow for that we want to instrument in our prediction (just scaled by $share^n_{i,base}$), the term $\sum_{j\neq i}^{J} FBstock^n_{j,t}$ is defined as the total inflow of immigrants from nationality n coming to Spain at time t minus the inflow of immigrants from nationality n coming to province i at time t. But we still require this term to be orthogonal to current local shocks. This assumption may be violated if location in provinces other than i is correlated with unobservable economic conditions of province i at a given point in time t. This is probable, specially if our spatial units are small and the economic conditions that attract immigrants are spatially correlated. For example, the economic condition in "economically big" provinces (like Madrid or Barcelona) could influence the total number of immigrants deciding to come to Spain, even if they end up locating somewhere else (based on their ethnic networks). To solve this issue a similar strategy to Saiz (2007) and Ortega & Peri (2009) is adopted. I compute the predicted total stock by country of origin from the results of a gravity model which depends only on push factors. Details of this procedure are given in the appendix (section A.2.1). Using the predictions from equations (A.2) and (A.3), I redefine the instrument as:

$$IV2\_ratio_{i,t-1} = \frac{imp\_pred\_FBinflow_{i,t-1}}{imp\_pred\_FBstock_{i,t-1} + natives_{i,t-1}} \tag{1.8}$$

However, there could still exist a final issue with the construction of (1.8) which might make the instrument invalid. Total population stock, which appear in the denominator, is the results of the sum of the foreign-born (imputed prediction) plus the natives. As discussed in section 1.2.2, the number of total natives residing in a given province might depend on the number of foreign-born in the same location. For this reason, I use a similar shift-share strategy to compute a prediction for the location of natives $imp\_natives_{i,t-1}$, based on past location patterns. Details are given in the appendix (section A.2.2).

Substituting the actual native stock by its prediction in equation (1.8), I finally define the main instrument as:

$$IVmain\_ratio_{i,t-1} = \frac{imp\_pred\_FBinflow_{i,t-1}}{imp\_pred\_FBstock_{i,t-1} + imp\_natives_{i,t-1}} \tag{1.9}$$

I use $IVmain\_ratio_{i,t-1}$ in the main instrumental variables estimation results and different variations of it (changing the base year share, the national inflow used and

[16] As a robustness check 1930, 1940 and 1985 location patterns is be used below. I also use 2001 location patterns as base year (first year of the period of analysis) to compute the instrument, which makes it a very strong instrument but unlikely to be exogenous.

the prediction of the native stocks) in the robustness checks.

1.2.4 Data sources and variable construction

The spatial unit of analysis is the province (NUTS3). I exclude Ceuta and Melilla because of their particular history and lack of data. I have population and economic data for the 50 provinces during period 1998-2010. However, I focus on the period 2001-2010 for several reasons. First, Fernández-Huertas et al. (2009) and Bertoli et al. (2011) recommend the use of population data coming from the population registers (*Padrón*) from 2001 because its reliability improves after that year. In section 1.3.5, I test the robustness of the results to the inclusion of years prior to 2001. Secondly, it is after 2001 that the stock of foreign-born starts increasing significantly. It could be the case that most entries started in 2001 or that the stocks started to be correctly measured after that year. To mitigate measurement error I then focus on 2001-2010 for the main analysis. Thirdly, the rental prices data is only available from 2001 so focusing on this time period allows us to compare the rental and purchase prices results over the same time period. Finally, using the housing boom and bust allows adoption of a demanding estimation strategy as there is more variance in the house price growth data.

The available data sources for population and foreign-born numbers in Spain are the Census (each 10 years) and the population municipality registers (annual). The number of residents in a municipality is registered by the city councils in an administrative register called the Municipal Register (*Padrón*). The annual records of the municipal register, dated on the 1^{st} January of each year, is obtained from its updates. In 1996 a modification of register regulations was undertaken. A continuous and computerised management system for municipal registers was introduced, based on the coordination of all of the municipal registers by the National Statistical Institute. This new system yields more accurate and up-to-date information. The first population data available with this new system is dated 1^{st} January 1996.

One advantage of this register is that it gives very precise information on the population figures, on a yearly basis, for very small geographical units (up to census sections). Another important advantage is that it collects the total number of foreign-born residents even if they are illegal immigrants. However, it has two disadvantages. For confidentiality issues, data availability on the characteristics of the population is limited (only age, gender and nationality). In addition, the immigration figures may be over-estimated because immigrants have to actively cancel their register when they move out of the country (if they move within the country their new register cancels out the old one). Alternatively, I could have used data from the Home Office (*Ministerio del Interior*), which collects figures of residence permits and legal immigrants on an annual basis. However, using the register data has the additional advantage that it is consistent with the total population data and that it collects all foreign-born that reside in a given province, not only the legal ones.

The house price data comes from Uriel-Jiménez et al. (2009), published by the Valencian Institute of Economic Research (henceforth IVIE) jointly with the BBVA Foundation (FBBVA). The database covers the period 1990-2007 and the IVIE prices are calculated using the original data from the Spanish Housing Department (*Ministerio de Vivienda*). The Housing Department official data provides the average price per square meter on dwellings purchases in the private sector. It is provided every quarter for all the provinces. The original price data is calculated by the Housing Department from the data provided by the Professional Association of Valuation Societies (*ATASA - Asociación Profesional de Sociedades de Valoración*). It is a weighted average of the valuation price and the number of valuations[17]. The IVIE dataset of house prices is constructed by weighting the official prices provided by the Housing Department to take into account the location of the dwelling and when it was built. As the IVIE data is only available until 2007, the dataset was expanded until 2010 by applying the provincial price growth rates from the Housing Department official data series.

Data on rental prices comes from the Housing Department and the National Institute of Statistics (*INE*). I combine data from National Observatory of Rented Properties (*Observatorio Estatal de la Vivienda en Alquiler*) and the consumer price indices (CPI provinces - rents component) to calculate the average rent price per square meter of the each province, from 2001 to 2010.

As time varying controls I use the number of credit establishments in a given province and the share of saving banks (to control for credit availability), the growth of GDP and the growth of the unemployment rate. Data on the number of banks comes from the La Caixa Spanish Economic Yearbook (*La Caixa Anuario Económico de España*), which collects data at the municipality and the province level for several socioeconomic indicators. Data on the growth of GDP comes from the Regional Economic Accounts of the National Institute of Economics. The province unemployment rate was calculated using the IVIE data on human capital (*Estimación de las Series de Capital Humano 1964-2010*) and it is defined as the ratio of unemployed over working-age population.

I also use time-invariant province characteristics in the specifications without province fixed-effects. These include: geographical characteristics (a dummy if the province is located on the coast, the length of the coastline and the surface if the national parks, from the National Geographical Institute); weather conditions (average rainfall and average temperature in January, from the National Agency of Meteor-

[17] There also exists data on transaction prices, but only after 2004.

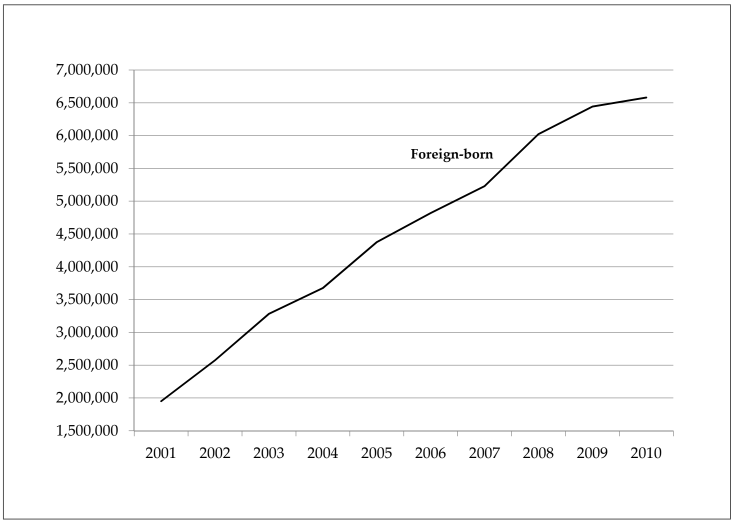
ology) and initial province attributes in 2000 (number of retails shops, number of restaurants and bars, relative weight of the tourism sector, from La Caixa Spanish Economic Yearbook; and number of doctors from the National Institute of Statistics).

1.2.5 Descriptive statistics

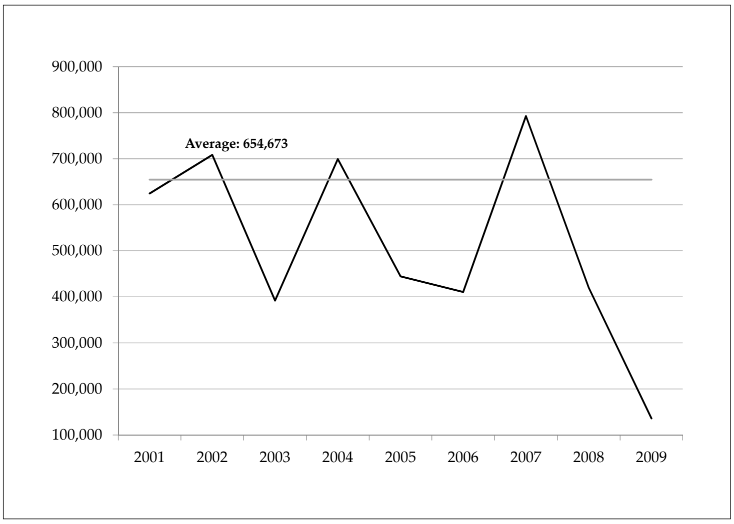
Table A.7 contains summary statistics of all the variables for the 50 provinces over the 9 year period (2002/2010 for the prices and 2001-2009 for the population variables), i.e, for the all the 450 observations of the panel pooled. The mean total change in log (annual growth) for rental prices is 3.3%, while for purchase prices is between 6.30 and 7.2%, depending on the source. Average provincial population growth if 12.5%, while the immigration ratio is 10.5%. The table also displays the summary statistics for the province time-invariant attributes and the time-varying controls. The final rows present summary statistics of the variables related to the supply of housing, which are used in section 1.3.6.

Tables A.8 and A.10 show the mean and the standard deviation only for the house prices growth and the population variables (the ratios). Table A.8 displays the summary statistics for each province over the 9-year time period. The last rows of both tables display the total (across provinces and years) mean and standard deviation, which correspond to the values in table A.7. Table A.9 display the sum of the squares from the ANOVA decomposition to illustrate the sources of variation of the variables across provinces. The analysis of these tables allows us to look at both the spatial (provinces) and the temporal (years) dimensions of the summary statistics. From these two tables we can infer that most of the variation in the growth of prices is within provinces, especially for purchase prices. This fact allow us to use a demanding empirical strategy (inclusion of region and province trends) and still identify the parameters. For the population variables there is nevertheless more variation between provinces than within them, except for the immigration ratio, where the within and between variation across provinces is fairly similar. Table A.10 shows the statistics for each year across the 50 provinces. We observe that between year growth of rents changes very little between 2002 and 2008, and it decreases when the housing crises starts. For purchase prices these changes are even bigger, as the growth rate becomes negative from 2009 onwards.

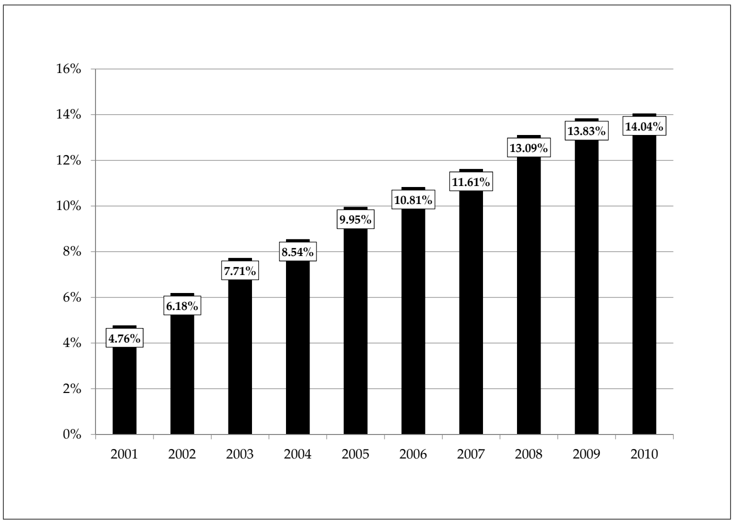
Figure 1.1 shows the Spanish average time evolution of the stocks (top panel) and inflows (middle panel) of foreign-born and the share of foreign-born over population (bottom panel), between years 2001 and 2010. The share of foreign-born over total population rose from 4.8% to 14% and the number of foreign-born increased 237% (from 1,950,452 the 1^{st} January 2001 to 6,579,121 the 1^{st} January 2010, according to the Population register). In every year of the period, the inflows of foreign-born



(a) Foreign-born stocks



(b) Foreign-born changes



(c) Percentage of foreign-born over population

Figure 1.1: Immigration stocks and inflows 2001-2010

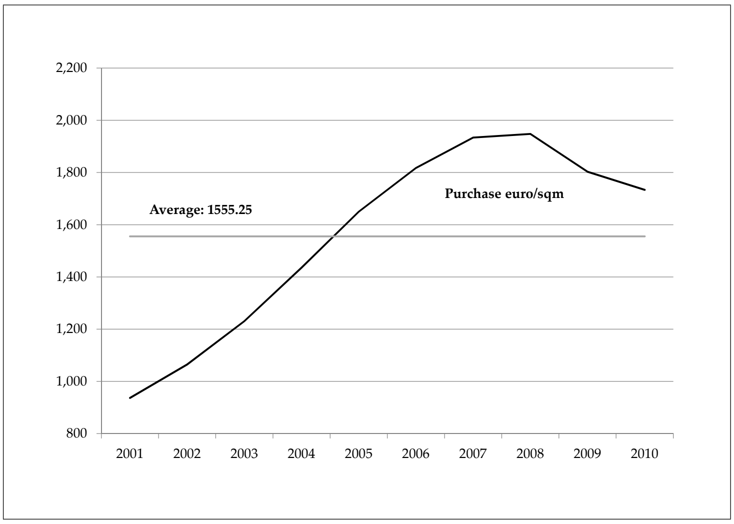
were over 100,000 persons, and the average for the period is over 650,000. The three spikes in the inflows in figure 1.1(b) correspond to three events described in Bertoli & Fernández-Huertas (2011): the 2000 law which allowed access to municipality public services when registered, the 2004 illegal immigration amnesty and the accession of Romania and Bulgaria to the EU in 2007. We can also observe these peaks in column 5 of table A.10 .

Figures A.1 and A.2 in the appendix display several maps which show the spatial distribution of the stocks of foreign-born, the changes in the stock between 2001 and 2010, the share of foreign-born at the beginning and the end of the period and the total growth of foreign-born population. The different colours represent the 5 quantiles of the values of the mapped variable. The provinces on the coast and Madrid are the ones which have higher levels of immigrants and have received most of the inflows. In 2001 the highest shares of immigrants were also concentrated on the coastal provinces and Madrid, but in 2010 many inner provinces have high shares of immigrants. This is confirmed in map A.2.3, in which we can observe that the regions with fewer immigrants in 2001 (map A.1.1) have been among the ones which have experienced the highest growth rate in the amount of foreign-born population between 2001 and 2010.

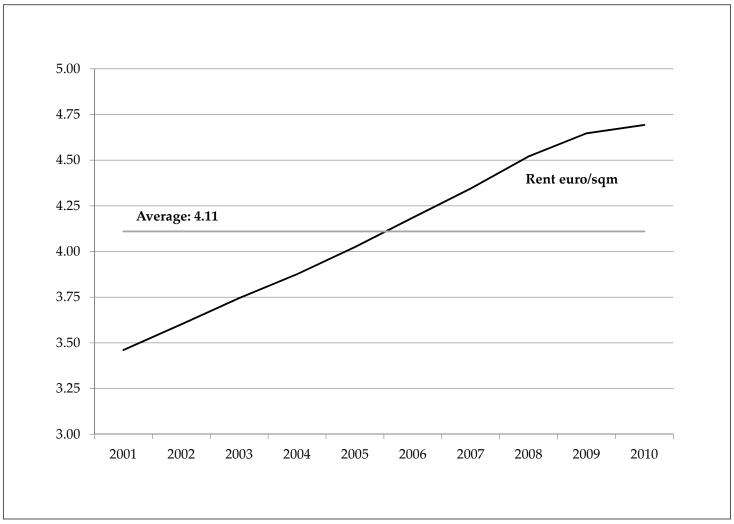
The top two panels of figure 1.2 shows the evolution of the average price in Spain for the period of analysis. During the "housing boom" years (2001-2008) housing purchase prices rose between 109 and 115%, followed by the construction sector crises which decreased average prices around 12%, with strong regional disparities both in the escalation and in the collapse of the prices. Rental prices also increased importantly during this period, around one point above the general CPI index during 2001-2010. During the whole period it raised 35%[18].

Construction of new dwellings also increased greatly during these years; between 2001 and 2010 5,312,245 new dwellings were constructed. The bottom panel of 1.2 shows the evolution of the total and private housing stocks and table A.1 also displays the total stock of dwellings in Spain during 2001-2010. Figure A.3 shows the spatial distribution of the growth of prices and housing stock between 2001 and 2010 (long-differences). Purchase prices increased greatly in all provinces. Some inner provinces (close to economic centers like Madrid, Barcelona, Sevilla or Valencia) have experience the highest growth rates in purchase prices, probably due to the fact that prices were lower in those provinces in 2001. These seem to be also the locations in which construction has been concentrated, as we can see in the bottom

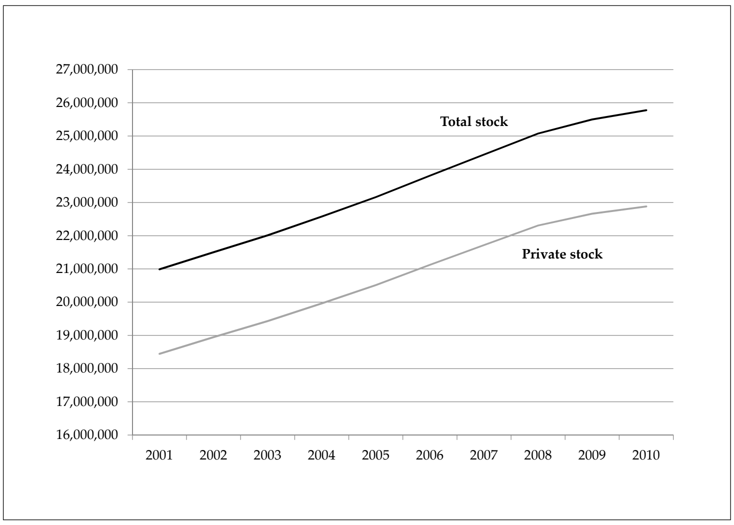
[18]Rental prices are based on the whole stock of properties available for renting (the already rented and the just rented), and are tightly connected to national CPI prices, so the scope for growth is smaller than in the case of purchase prices. On the other hand, the changes on house prices depend solely on new properties sold. Therefore, one can expect the increase on purchase prices to be much more volatile than that of rents.



(a) Growth of purchase prices



(b) Growth of rental prices



(c) Growth of dwelling stocks

Figure 1.2: house prices growth and dwellings construction 2001-2010

panel of the figure.

Most of the growth in prices and construction stopped in 2008 with the global economic crises and between 2008 and 2010 prices have decreased and construction of new dwellings has virtually stopped, but their levels are still above the average values of the end of the 90s.

1.3 Empirical results

1.3.1 Effects of immigration on house prices

1.3.1.1 Fixed-effect estimates

Tables 1.1 and 1.2 present the results of the estimation of equation (1.1), for rental prices and for purchase prices respectively. These results are obtained using data on annual changes on prices during the period 2002-2010 and data on the immigration ratio lagged one period (2001-2009). The number of observations is 450 (50 provinces times 9 years). In all specifications the standard errors are clustered at the province level, to allow for arbitrary correlation of the idiosyncratic shocks for a given province across time, and are robust to heteroskedasticity. All specifications include time dummies to control for national shocks. Different columns show results for different specification which diverge in the dummies and trends which are included and in the included controls (time invariant, time-varying or both). Specifications range from more to less demanding in terms of data variation: OLS results (column 1 to 3) to first-differences fixed-effects model (column 8 to 10).

The first three columns of tables 1.1 and 1.2 show the results obtained by OLS, adding time-invariant controls and time-varying controls in columns 2 and 3. The estimations by OLS of columns 2 and 3 include extra time-invariant controls (dummy if province is an island and dummy if the region has a single province), which drop when introducing the regional dummies. Columns 4 to 7 include regional dummies: column 4 only includes regional dummies, column 5 adds time-invariant province attributes, which control for different growth trends of provinces with different levels of weather conditions and amenities, and columns 6 and 7 add time-varying controls and regional trends. Columns 8 to 10 include province dummies and correspond to the first-differences fixed-effects model. Column 8 only includes the dummies, and 9 and 10 add the time-varying controls and the regional trends[19]. The specification in column 10 is very demanding, because it is identifying the parameters off the variation of provinces growth in prices and immigration rations around the 2002-2010 means, conditional on common national shocks and a linear regional

[19]Province trends absorb too much variation and do not allow to identify the parameters.

Change in the log of rental prices in t	OLS			Region dummies				Province dummies		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of immigrants in t-1 over population end t-2	0.346**	0.324**	0.280*	0.303**	0.418***	0.379***	0.347**	0.297**	0.284**	0.213
	[0.132]	[0.152]	[0.152]	[0.133]	[0.134]	[0.130]	[0.149]	[0.125]	[0.120]	[0.154]
Dummy if region only has one province		-0.001	-0.001							
		[0.002]	[0.002]							
Dummy if province is an island		-0.014**	-0.016**							
		[0.006]	[0.006]							
Coast dummy		-0.003	-0.003		-0.003	-0.003	-0.003			
		[0.004]	[0.004]		[0.004]	[0.004]	[0.004]			
Length of coastline		0.000	0.000		0.000	0.000	0.000			
		[0.000]	[0.000]		[0.000]	[0.000]	[0.000]			
Log of the surface of natural parks		-0.000	-0.001		-0.001	-0.001	-0.001			
		[0.001]	[0.001]		[0.001]	[0.001]	[0.001]			
Log of hours of average temperature (January)		0.002	0.002		0.005	0.006	0.006			
		[0.004]	[0.004]		[0.006]	[0.006]	[0.006]			
Log of hours of rain precipitation (January)		-0.000	0.001		0.003	0.004	0.004			
		[0.002]	[0.002]		[0.004]	[0.005]	[0.005]			
Log of number of retails shops in 2000		0.003	0.003		0.004	0.004	0.004			
		[0.003]	[0.003]		[0.003]	[0.003]	[0.003]			
Log of number of restaurants and bars in 2000		0.002	0.002		0.003**	0.002*	0.002*			
		[0.001]	[0.001]		[0.001]	[0.001]	[0.001]			
Importance of tourism sector - comparative index 2000		0.000	0.000		-0.000	-0.000	-0.000			
		[0.000]	[0.000]		[0.000]	[0.000]	[0.000]			
Log of the number of doctors - 2000		-0.001	-0.002		-0.002**	-0.002***	-0.002***			
		[0.001]	[0.001]		[0.001]	[0.001]	[0.001]			
Change of log of GDP in t-2			0.010			0.004	0.004		-0.004	-0.006
			[0.023]			[0.025]	[0.026]		[0.026]	[0.028]
Change of log of unemployment rate in t-2			0.043			0.050	0.048		0.050*	0.047
			[0.030]			[0.031]	[0.033]		[0.028]	[0.031]
Change of log of number of credit establishments in t-2			0.096**			0.082*	0.082*		0.081*	0.081**
			[0.040]			[0.045]	[0.043]		[0.041]	[0.040]
Change of percentage of savings banks in t-2			-0.038			-0.057	-0.013		-0.124*	-0.091
			[0.083]			[0.079]	[0.084]		[0.067]	[0.071]
Time invariant controls	No	Yes	Yes	No	Yes	Yes	Yes	No	No	No
Time-varying controls	No	No	Yes	No	No	Yes	Yes	No	Yes	Yes
Region trends	No	No	No	No	No	No	Yes	No	No	Yes
Observations	450	450	450	450	450	450	450	450	450	450
Adjusted R^2	0.36	0.40	0.41	0.39	0.44	0.45	0.45	0.46	0.47	0.48

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies.

Table 1.1: OLS and FE results rental prices

trend.

For all specifications displayed in table 1.1 the estimated effect of immigration on rental prices is positive and statistically different from zero, except in the most demanding specification (in column 10). Focusing on the specifications with regional dummies, the introduction of province time-invariant attributes increases the coefficient from 0.303 to 0.418. This change would indicate that the (joint) correlation between the included time-invariant province characteristics and the immigration ratio is negative. In any case, in this specification regional dummies are included, so the province time invariant characteristics control for differential growth trends in the province relative to the regional trend. Therefore, the interpretation of the change in the size of the coefficient becomes difficult. Adding time-varying controls changes the coefficient slightly, which indicates that the inclusion or exclusion of the time-varying controls does not affect our qualitative results because most of the variation of these regressors is collected by the region and the time dummies and by the province time-invariant characteristics. Adding the regional trends in column 7 decreases the coefficient a little bit more. Columns 8 to 10 repeat the specifications of columns 4, 6 and 7 adding province dummies, i.e. the first-differences fixed-effects model (with and without time-varying controls and with regional trends). The coefficient decreases with respect to that of the specifications with regional dummies. This would suggest there exist unobservable province characteristics (like amenities) which attract immigrants and are positively correlated with rental price growth. All models have reasonably high explanatory power, and the adjusted R^2 are between 0.36 and 0.48.

The estimated elasticities of the changes in the immigration ratio on log changes of rental prices presented in table 1.1 range from 0.28% to 0.42%. These numbers would imply that an increase of the share of foreign-born on the original population of a province of 10% would cause an increase on the rental prices between 2.8% and 4.8% the following year. These numbers are much smaller than previous estimates found by Saiz (2007), which are around 8-10%. A possible explanation is the legal environment in Spain, as compared to the US case. In Spain the standard legal tenancy agreement for privately let properties establishes (by default it is of 5-year length) that the annual increase on the rental price would of the same amount as the change in the national general consumer price index (CPI)[20]. Therefore, most of the variation of changes in rental prices, given that we control for national CPIs by including time dummies and for provincial trends by including province dummies, would come from variation on provincial CPI indices with respect to the national CPI and from growth in rental prices stemming from newly signed tenancy agreements. Because of this, we can expect the impact of immigration on rental prices to

[20] *Ley de Arrendamientos Urbanos 29/1994, del 24 de Noviembre de 1994.*

Change in the log of purchase prices in t	OLS			Region dummies				Province dummies		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of immigrants in t-1 over population end t-2	0.588**	0.697**	0.787**	0.720**	0.882**	0.980***	0.512*	1.128***	1.184***	0.581
	[0.285]	[0.310]	[0.312]	[0.320]	[0.338]	[0.323]	[0.270]	[0.420]	[0.415]	[0.350]
Dummy if region only has one province		-0.010**	-0.010**							
		[0.005]	[0.004]							
Dummy if province is an island		-0.060***	-0.057***							
		[0.015]	[0.015]							
Coast dummy		-0.012*	-0.011		-0.001	-0.001	-0.000			
		[0.007]	[0.007]		[0.006]	[0.006]	[0.006]			
Length of coastline		0.000**	0.000**		0.000	0.000	0.000			
		[0.000]	[0.000]		[0.000]	[0.000]	[0.000]			
Log of the surface of natural parks		0.003*	0.003*		-0.000	-0.000	-0.001			
		[0.002]	[0.002]		[0.001]	[0.001]	[0.001]			
Log of hours of average temperature (January)		0.019**	0.019**		0.018**	0.016**	0.018**			
		[0.008]	[0.008]		[0.007]	[0.007]	[0.007]			
Log of hours of rain precipitation (January)		-0.008*	-0.009*		0.007	0.005	0.005			
		[0.004]	[0.004]		[0.005]	[0.005]	[0.005]			
Log of number of retails shops in 2000		0.001	0.002		0.004	0.004	0.001			
		[0.005]	[0.005]		[0.004]	[0.004]	[0.003]			
Log of number of restaurants and bars in 2000		0.000	0.000		0.001	0.002	0.002			
		[0.002]	[0.002]		[0.002]	[0.002]	[0.002]			
Importance of tourism sector - comparative index 2000		-0.000	-0.000		-0.000***	-0.000***	-0.000***			
		[0.000]	[0.000]		[0.000]	[0.000]	[0.000]			
Log of the number of doctors - 2000		-0.004	-0.003		-0.002	-0.001	-0.001			
		[0.002]	[0.002]		[0.002]	[0.002]	[0.002]			
Change of log of GDP in t-2			-0.030			-0.044	-0.092		-0.066	-0.119*
			[0.078]			[0.080]	[0.068]		[0.078]	[0.065]
Change of log of unemployment rate in t-2			-0.136*			-0.112	0.016		-0.107	0.016
			[0.080]			[0.078]	[0.066]		[0.077]	[0.065]
Change of log of number of credit establishments in t-2			-0.129			-0.158	-0.053		-0.183	-0.080
			[0.101]			[0.122]	[0.112]		[0.144]	[0.135]
Change of percentage of savings banks in t-2			0.177			0.096	-0.163		0.127	-0.176
			[0.185]			[0.174]	[0.151]		[0.175]	[0.152]
Time invariant controls	No	Yes	Yes	No	Yes	Yes	Yes	No	No	No
Time-varying controls	No	No	Yes	No	No	Yes	Yes	No	Yes	Yes
Region trends	No	No	No	No	No	No	Yes	No	No	Yes
Observations	450	450	450	450	450	450	450	450	450	450
Adjusted R^2	0.82	0.83	0.83	0.84	0.84	0.84	0.86	0.85	0.85	0.87

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies.

Table 1.2: OLS and FE results purchase prices

be limited, due to the existing legal limits to its growth[21].

Table 1.2 displays the results for the effect of immigration on housing purchase prices. The estimates are bigger than in the case of rental prices, as expected, and range between 0.6 to 1.18. All models have high explanatory power; the adjusted R^2 are between 0.82 and 0.85. The coefficients are always positive, except in the most demanding specification. OLS estimates suggest that provinces which are islands, where the coast is longer, the weather hotter and drier, have more natural parks and lower unemployment growth have higher increases in purchase prices. From column 4 we include regional dummies. As before, the province time-invariant attributes are negatively correlated with the prices and the immigration ratios. Adding time-varying controls in column 6 increases the coefficient even further. Finally, the estimates in columns 8 and 9 predict an elasticity of the purchase prices with respect to the immigration ratio of slightly over 1. A 10% increase in the ratio would imply an increase in the purchase price of around 12% on the following year. As for the case of the effects on rental prices, these estimates are also below Saiz (2007) estimates. This could be due to the different spatial and temporal scale of the two empirical exercises, given that Saiz (2007) uses decennial data and different spatial units (metropolitan areas) while I use annual data and bigger geographical units (provinces). The more than proportional increase on prices could be caused by a less than proportional response of the supply (due to non-sufficient construction or non-sufficient access to housing) or by induced increased demand by natives if they are moving to the same places as immigrants. These potential explanations would be examined in the following subsections. I first deal with the correct identification of the parameters of interest by using an instrumental variables estimation strategy.

1.3.1.2 Instrumental variables

In order to be able to infer causal effects from the estimates of coefficient β in equation (1.1), I implement the instrumental variables strategy explained above. Tables 1.3 and 1.4 present the results using instrument as defined in equation (1.9). The predicted stocks and inflows of foreign-born by nationality from the gravity model estimation of columns 1 in tables A.6 and A.5 is used to construct $imp\_pred\_FBstock_{i,t-1}$, and year 1991 is used as the base year for the predicted location patterns of both natives and foreign-born. The tables have the same structure as tables 1.1 and 1.2. As previously, time fixed-effects are included in all the specifications and the standard errors are clustered at the province level. The tables also display some test of the validity of the instruments. Details on them are given in section A.3.

[21] The smaller margin for adjustment of rents is also illustrated in figure 1.2. Rents grow slower

Change in the log of rental prices in t	OLS			Region dummies				Province dummies		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of immigrants in t-1 over population end t-2	0.665**	0.562	0.443	0.924**	0.974**	0.946**	1.047**	0.120	0.151	0.126
	[0.291]	[0.487]	[0.480]	[0.387]	[0.393]	[0.426]	[0.504]	[0.461]	[0.472]	[0.579]
Dummy if region only has one province		-0.001	-0.001							
		[0.003]	[0.003]							
Dummy if province is an island		-0.012*	-0.014**							
		[0.007]	[0.007]							
Coast dummy		-0.004	-0.004		-0.004	-0.004	-0.004			
		[0.005]	[0.005]		[0.004]	[0.004]	[0.004]			
Length of coastline		0.000	0.000		0.000	0.000	0.000			
		[0.000]	[0.000]		[0.000]	[0.000]	[0.000]			
Log of the surface of natural parks		-0.000	-0.000		-0.001	-0.001	-0.001			
		[0.001]	[0.001]		[0.001]	[0.001]	[0.001]			
Log of hours of average temperature (January)		0.002	0.001		0.003	0.004	0.003			
		[0.004]	[0.004]		[0.006]	[0.006]	[0.006]			
Log of mm of rain precipitation (January)		0.002	0.002		0.004	0.005	0.005			
		[0.003]	[0.002]		[0.004]	[0.005]	[0.005]			
Log of number of retails shops in 2000		0.004	0.003		0.008**	0.008*	0.009*			
		[0.003]	[0.003]		[0.004]	[0.004]	[0.005]			
Log of number of restaurants and bars in 2000		0.002	0.002		0.002	0.002	0.002			
		[0.001]	[0.001]		[0.001]	[0.001]	[0.001]			
Importance of tourism sector - comparative index 2000		-0.000	0.000		-0.000	-0.000	-0.000			
		[0.000]	[0.000]		[0.000]	[0.000]	[0.000]			
Log of the number of doctors - 2000		-0.001	-0.002		-0.002***	-0.003***	-0.003***			
		[0.001]	[0.001]		[0.001]	[0.001]	[0.001]			
Change of log of GDP in t-2			0.003			-0.019	-0.016		-0.000	-0.004
			[0.027]			[0.028]	[0.028]		[0.025]	[0.026]
Change of log of unemployment rate in t-2			0.038			0.044	0.039		0.050*	0.047
			[0.032]			[0.032]	[0.035]		[0.027]	[0.029]
Change of log of number of credit establishments in t-2			0.092**			0.064	0.051		0.082**	0.083*
			[0.041]			[0.050]	[0.054]		[0.042]	[0.043]
Change of percentage of savings banks in t-2			-0.034			-0.065	-0.019		-0.127*	-0.095
			[0.082]			[0.072]	[0.075]		[0.066]	[0.073]
Time invariant controls	No	Yes	Yes	No	Yes	Yes	Yes	No	No	No
Time-varying controls	No	No	Yes	No	No	Yes	Yes	No	Yes	Yes
Region trends	No	No	No	No	No	No	Yes	No	No	Yes
Observations	450	450	450	450	450	450	450	450	450	450
LM test stat under-identification (K-P)	8.89	9.60	9.61	6.02	12.66	12.51	12.62	7.42	7.56	7.37
P-value of under-identification LM statistic	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01
F-stat weak identification (K-P)	19.01	10.48	9.77	26.51	24.50	21.25	24.59	14.73	14.70	16.25

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies.

Table 1.3: Instrumental variables results rental prices

Table 1.3 presents the results for changes in rental prices. As expected, the standard errors increase when we use instrumental variables. The estimated β are larger than in table 1.1 for all specifications. This would suggest that immigrants are moving, conditional on the controls and the fixed effects, to provinces which are experiencing negative shocks in the growth of rental prices, and therefore the OLS and FE estimates are downward biased. Columns 4 to 7 include regional dummies. The elasticities are around 1%, very similar to those found by Saiz (2007) for the US context. Column 7 reports the results using region dummies and trends and time-invariant and time-varying province characteristics. We can be quite sure about the causal interpretation of these results, because the specification is already very demanding in terms of variation (which ensures the validity of the instrument) and the instrument is strong. The coefficients in the last three columns (8 to 10) become much smaller (the standard errors are very similar to those of the specification with regional dummies). As explained before, given that rents growth is limited to CPI annual changes for most rental agreements, the province dummies could be capturing most of the variation in the growth rate of rental prices, so once we controls for endogenous location of the immigrants by means of the instrument, the coefficient is largely reduced and becomes insignificant. This could also be because the instrument is weaker when introducing the province fixed-effects. The F-stat of included instruments is lower than in columns 4 to 7, and it is not above the Stock-Yogo critical value at 10% (although it is identified at the 5% level). Given that the variation in the instrument comes from national changes in the inflows of immigrants (captured by the time fixed-effects) and from the past location patterns (capture by province characteristics which are removed in the first-differences setting), it is reasonable that the strength of the instrument is decreased when adding province fixed-effects.

Table 1.4 shows the results for changes in purchase prices. The standard errors are more than doubled with respect to those of table 1.2. The results by OLS and only including region dummies (columns 1 to 4) are insignificant when we use an instrument, as compared to significant at the 5% level in table 1.2. When we control for time-invariant province characteristics in column (5), the coefficient becomes statistically different from zero at the 1% significance level and the size of the coefficient is much larger. Some of the province attributes are also highly significant (and they are jointly significant), which reinforces the need of controlling for provincial attributes in order to be able to correctly identify parameter β, even after the inclusion of region dummies. Introducing regional trends in column 7 diminishes the coefficient marginally. Finally, columns 8 to 10 show the results using the first-differences fixed-effects model with instrumental variables. The coefficients increase substantially, and are highly significant. They suggest an elasticity of purchase prices with

than prices in the boom years but they also slow down less after 2009.

Change in the log of purchase prices in t	OLS			REGION DUMMIES				PROVINCE DUMMIES		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of immigrants in t-1 over population end t-2	-0.322	0.390	0.525	0.803	2.131***	2.419***	2.166***	3.028***	3.035***	2.739**
	[0.766]	[0.944]	[1.052]	[0.505]	[0.700]	[0.812]	[0.757]	[1.010]	[0.997]	[1.233]
Dummy if region only has one province		-0.009**	-0.010**							
		[0.005]	[0.005]							
Dummy if province is an island		-0.063***	-0.059***							
		[0.015]	[0.015]							
Coast dummy		-0.010	-0.009		-0.004	-0.003	-0.003			
		[0.008]	[0.008]		[0.007]	[0.007]	[0.007]			
Length of coastline		0.000***	0.000***		-0.000	0.000	0.000			
		[0.000]	[0.000]		[0.000]	[0.000]	[0.000]			
Log of the surface of natural parks		0.003	0.003		0.001	0.001	0.001			
		[0.002]	[0.002]		[0.002]	[0.002]	[0.002]			
Log of hours of average temperature (January)		0.020**	0.020**		0.013	0.010	0.012			
		[0.008]	[0.008]		[0.008]	[0.009]	[0.008]			
Log of mm of rain precipitation (January)		-0.010	-0.010		0.010	0.007	0.008			
		[0.007]	[0.007]		[0.007]	[0.007]	[0.007]			
Log of number of retails shops in 2000		-0.000	0.001		0.013**	0.014**	0.013*			
		[0.006]	[0.006]		[0.006]	[0.007]	[0.006]			
Log of number of restaurants and bars in 2000		0.000	0.000		-0.001	0.000	0.000			
		[0.002]	[0.002]		[0.002]	[0.002]	[0.002]			
Importance of tourism sector - comparative index 2000		-0.000	-0.000		-0.000***	-0.000***	-0.000***			
		[0.000]	[0.000]		[0.000]	[0.000]	[0.000]			
Log of the number of doctors - 2000		-0.004*	-0.004*		-0.003	-0.003	-0.003			
		[0.002]	[0.002]		[0.002]	[0.002]	[0.002]			
Change of log of GDP in t-2			-0.018			-0.103	-0.141**		-0.119	-0.154**
			[0.080]			[0.081]	[0.067]		[0.081]	[0.069]
Change of log of unemployment rate in t-2			-0.128			-0.129*	-0.004		-0.107	0.018
			[0.087]			[0.078]	[0.063]		[0.077]	[0.068]
Change of log of number of credit establishments in t-2			-0.123			-0.204	-0.127		-0.198	-0.121
			[0.106]			[0.136]	[0.128]		[0.150]	[0.137]
Change of percentage of savings banks in t-2			0.172			0.074	-0.177		0.168	-0.096
			[0.179]			[0.169]	[0.135]		[0.167]	[0.156]
Time invariant controls	No	Yes	Yes	No	Yes	Yes	Yes	No	No	No
Time-varying controls	No	No	Yes	No	No	Yes	Yes	No	Yes	Yes
Region trends	No	No	No	No	No	No	Yes	No	No	Yes
Observations	450	450	450	450	450	450	450	450	450	450
LM test stat under-identification (K-P)	8.89	9.60	9.61	6.02	12.66	12.51	12.62	7.42	7.56	7.37
P-value of under-identification LM statistic	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01
F-stat weak identification (K-P)	19.01	10.48	9.77	26.51	24.50	21.25	24.59	14.73	14.70	16.25

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * p<0.05, ** p<0.01, *** p<0.001. All specifications include year dummies.

Table 1.4: Instrumental variables results purchase prices

respect to the immigration ratio of around 3% (for a 1% increase in the immigration ratio). The specification of column 10, which is highly demanding, yields a positive and significant at 5% level estimated elasticity around 2.7%. This number is very similar to that found by Saiz (2007)) and confirm the findings of Gonzalez & Ortega (2009), which with a similar methodology, find that a 1% increase in the immigration ratio increase house prices 3.2%, using data for the period 1998-2008.

Previous research has found estimates of positive sign and similar magnitude. Are the estimated elasticities size and sign what we would expect? If the supply of housing is responding sufficiently to increases in demand, we should not expect an increase in foreign-born population to have such a large effect on house prices growth. Indeed, as discussed in section 1.2.2, during the period of analysis the construction of new dwellings was very intense, so we could have anticipated a moderate effect of an increase in demand induced by the increase in foreign-born population. However, as I show in subsection 1.3.6 below, the increase in supply did not seem to have had an effect on moderating the increase in prices. This could be because, even if new dwellings were constructed, they were not constructed in the places where immigrants wanted or could afford to live, because immigrants had credit restriction to access the housing markets or because many dwellings were not occupied[22].

Furthermore, if immigrants and natives are co-locating in the same provinces, demand in these location could be increasing more than the resulted from foreign-born inflows. If we want to be able to properly interpret the size of β, we need to study the relationship between natives and immigrant location decisions.

1.3.2 Effects of immigration on natives location

Section 1.2.2 discusses the issues related to the interpretation of the coefficient β when we do not take into account natives mobility. Specifically, standard tests of the effect of immigration on local prices that estimate specifications similar to (1.1) are unable to separate effects due to the increase on local housing demand from foreign-born from the effect on prices which is due to (induced) changes in demand from native population relocation Saiz (as noted by 2007). Using Saiz (2007) words, it would be capturing the "long-run" total effect of immigration. The estimated β would be the net result of increased increased demand from immigrant inflows plus/minus any changes in demand due to inflows or outflows of natives. If immigrants actually attract natives, the estimated coefficient β would be the sum of both changes in demand. We would not be able to disentangle the effect created by increased demand from immigrants (inflows) from the changes in demand from

[22]In fact, some estimates suggest that around 15-20% of houses are unoccupied.

natives. β would be capturing the effect of "total changes in demand" from both natives and immigrants, which would be higher (upward biased) than the "net" effect of immigration inflows on prices (through changes in total population).

In this section I estimate a causal relationship between native location and immigration inflows[23]. Table 1.5 shows the results of the estimation of equation (1.2). Columns 1 to 5 show the results using regional dummies and province fixed-effects and columns 6 to 10 repeat the estimations using instrumental variables (instrument 1.9). The specification vary in the inclusion of controls and the trends. I use the same estimation strategy as in the estimation of (1.1). As the inflow of natives and immigrants are contemporaneous, the time-varying controls are lagged one period with respect to time t.

As before, when instrumenting the immigration ratio coefficients increase substantially. Our preferred estimates are those of columns 9 and 10, which use the instrument and the first-differences fixed-effects estimation. These estimates predict that a for each 100 immigrants locating in given province in a given year, around 43 natives located in the same province in the same year. Table A.11 shows the results using the "lagged" immigration ratio. The results are slightly weaker (i.e. not significant in the most demanding specifications). Additionally, table A.12 in the appendix shows the same results but using the immigration and native ratio for people aged 16-64, assuming that is the working-aged part of the household who takes the decision on relocating. The results are very similar: for each 100 immigrants aged 16-64 locating in given province in a given year, almost 37 natives aged 16-64 located in the same province in the same year, and around 18 in the following year.

These findings suggest that natives and immigrants are locating in the same provinces, mostly "contemporaneously". This could be because immigrants have some unobservable attributes which are desirable for natives, for example if natives like ethnic diversity or if immigrants are specializing in producing goods and services which are desirable for natives.

Fernández-Huertas et al. (2009) find a comparable result for a long-differences non-causal estimation from population growth regressed on the immigration ratio for the period 2001-2008. Their prediction is of 11 natives for each 100 immigrants. They argue that this number is sufficiently small to be negligible to have an impact on compensation or reinforcement of the impact of immigration inflows on the housing or the labour markets[24]. Nevertheless, my results contradict this assess-

[23] Other examples where the relationship between natives and immigration is explored are Stillman & Maré (2008) and Ortega & Verdugo (2011).

[24] The different results could be due to the fact that these authors do not use instrumental variables in their estimation and they use long differences between 2001 and 2008, so they only use 52 observations. In fact, when they perform the estimation at the municipality level, using over 8,000

Inflow of natives in t over population end t-1	Region dummies (1)	(2)	(3)	Province dummies (4)	(5)	Region dummies (6)	(7)	(8)	Province dummies (9)	(10)
Inflow of immigrants in t over population end t-1	0.363***	0.321***	0.330***	0.182***	0.155***	0.707***	0.652***	0.665***	0.454***	0.432***
	[0.085]	[0.076]	[0.093]	[0.039]	[0.040]	[0.218]	[0.199]	[0.226]	[0.112]	[0.128]
Coast dummy	0.000	0.000	0.000			-0.000	-0.000	-0.000		
	[0.002]	[0.002]	[0.002]			[0.002]	[0.002]	[0.002]		
Length of coastline	0.000***	0.000**	0.000**			0.000*	0.000*	0.000*		
	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]		
Log of the surface of natural parks	-0.000	-0.000	-0.000			0.000	0.000	0.000		
	[0.001]	[0.001]	[0.001]			[0.001]	[0.001]	[0.001]		
Log of hours of average temperature (January)	-0.005	-0.004	-0.004			-0.006	-0.005	-0.006		
	[0.006]	[0.005]	[0.005]			[0.005]	[0.005]	[0.005]		
Log of mm of rain precipitation (January)	0.001	0.001	0.001			0.001	0.002	0.002		
	[0.003]	[0.002]	[0.002]			[0.002]	[0.002]	[0.002]		
Log of number of retails shops in 2000	0.001	0.000	0.000			0.003	0.003	0.003		
	[0.002]	[0.002]	[0.002]			[0.003]	[0.003]	[0.003]		
Log of number of restaurants and bars in 2000	0.001	0.001	0.001			0.000	0.000	0.000		
	[0.001]	[0.001]	[0.001]			[0.001]	[0.001]	[0.001]		
Importance of tourism sector - comparative index 2000	-0.000	-0.000	-0.000			-0.000	-0.000*	-0.000*		
	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]		
Log of the number of doctors - 2000	0.001	0.001	0.001			0.000	0.000	0.000		
	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]		
Change of log of GDP in t-1		0.047***	0.046***	0.016**	0.014*		0.033***	0.036***	0.009	0.009
		[0.015]	[0.016]	[0.007]	[0.007]		[0.012]	[0.013]	[0.008]	[0.007]
Change of log of unemployment rate in t-1		-0.009	-0.009	-0.013**	-0.012*		-0.013	-0.013	-0.013*	-0.012*
		[0.009]	[0.008]	[0.006]	[0.006]		[0.009]	[0.009]	[0.007]	[0.007]
Change of log of number of credit establishments in t-1		0.033**	0.041**	-0.004	0.002		0.022*	0.026*	-0.007	-0.003
		[0.014]	[0.016]	[0.006]	[0.006]		[0.013]	[0.015]	[0.008]	[0.009]
Change of percentage of savings banks in t-1		0.008	0.017	-0.010	-0.008		0.003	0.015	-0.004	0.002
		[0.030]	[0.034]	[0.015]	[0.016]		[0.028]	[0.030]	[0.016]	[0.018]
Time invariant controls	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Time-varying controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Region trends	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Instrumented	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	450	450	450	450	450	450	450	450	450	450
Adjusted R^2	0.58	0.60	0.60	0.21	0.25					
LM test stat under-identification (K-P)						12.66	12.51	12.62	7.56	7.37
P-value of under-identification LM statistic						0.00	0.00	0.00	0.01	0.01
F-stat weak identification (K-P)						24.50	21.25	24.59	14.70	16.25

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies.

Table 1.5: Native displacement results

ment because the size of the "collocation" is substantially larger. Hence, it is quite likely that any impact of immigrants on the housing markets would be amplified by the arrival of natives. I investigate this possibility in the following subsection.

1.3.3 Effects of immigration on house prices revisited

After finding a sizeable "attraction" of immigrants on natives, a natural extension is then to test the effect of population on prices using "only" the variation in population caused by immigration inflows. Total population in a given province in a given point in time is the sum of the foreign-born and the natives living in that region. If the location of immigrants is simultaneously determined as the location of natives and it is positively related, the estimated coefficient β in equation (1.1) is actually capturing the effects of increased demand from immigrants plus the increased demand from relocated natives, i.e. the effect of total population changes. In this section I propose a methodology to isolate the effect of the changes that can be attributed to immigration. I use exogenous variation on the location of immigrants for a more direct test of the effect of immigration on house prices growth. By using this procedure, we would be able to separate the "net" the effect of foreign-born demand on prices from the total effect of immigration on prices (through changes in total province population).

In section 1.2.3 I described the shift-share strategy followed to construct a prediction for native location based on past location patterns (for stayer and for movers). I used this prediction in the denominator of the instrument for the immigration ratio (1.9). I can use this prediction to construct the equivalent for the native ratio:

$$IVnatives\_ratio_{i,t-1} = \frac{imp\_natives_{i,t} - imp\_natives_{i,t-1}}{imp\_pred\_FBstock_{i,t-1} + imp\_natives_{i,t-1}} \tag{1.10}$$

As before, I construct this instrument using 1991 as base year and the prediction from the gravity model of column 1 in table A.6 to construct $imp\_pred\_FBstock_{i,t-1}$.

Table 1.6 explores the effect of natives and immigrants as separate population groups on rental (top) and purchase (bottom) prices. They display the estimates of the instrumental variable regressions which include regional dummies and trends and the controls (table A.13 in the appendix replicates the estimates using province dummies instead of regional dummies). The last row of the tables indicates which is the instrument that we are using (1.9, 1.10 or both).

The first column instruments the immigration ratio with (1.9), which is equivalent to the results of tables 1.3 and 1.4. Column 2 estimates the effect of the native ratio and instruments it with (1.10). Column 3 uses both ratios as dependent vari-

observations, they find very similar estimates to mine.

ables, and uses both instruments[25]. Columns 4 to 6 use total population growth as the regressor of interest and instrument it either with both ratios (4), with the immigration ratio (5), with the native ratio (6).

Column 1 reproduces previous results of tables 1.3 and 1.4. In column 2 we use the instrument for natives (1.10) and find that natives do not have an effect on rental prices but do affect purchase prices. The instrument is very strong. Even if the instruments are strong, when accounting for changes in native ratios and immigration ratios "at the same time" in column 3, only the coefficients for the immigration ratio are significant and they are similar to those of column 1. However, models with two endogenous variables are difficult to interpret and to identify [26]. For this reason, in the remaining of the section I focus in one regressor of interest.

Change in the log of rental prices in t	REGION DUMMIES (1)	(2)	(3)	(4)	(5)	(6)
Inflow of immigrants in t-1 over population end t-2	1.047**		1.190*			
	[0.504]		[0.647]			
Inflow of natives in t-1 over population end t-2		0.214	-0.215			
		[0.167]	[0.266]			
Inflow of population in t-1 over population end t-2				0.210*	0.629**	0.158
				[0.121]	[0.295]	[0.120]
Change in the log of purchase prices in t	REGION DUMMIES (1)	(2)	(3)	(4)	(5)	(6)
Inflow of immigrants in t-1 over population end t-2	2.166***		2.202**			
	[0.757]		[1.032]			
Inflow of natives in t-1 over population end t-2		0.741***	-0.054			
		[0.255]	[0.525]			
Inflow of population in t-1 over population end t-2				0.628***	1.301***	0.544***
				[0.174]	[0.399]	[0.187]
Observations	450	450	450	450	450	450
LM test stat under-identification (K-P)	12.62	16.54	10.19	21.24	13.18	18.31
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)	24.59	116.52	7.29	78.20	32.29	123.45
P-value Hansen J statistic				0.10		
A-P F-test of excluded instruments (immigration)			15.61			
A-P F-test of excluded instruments (natives)			78.52			
Instrument(s)	IMM	NAT	BOTH	BOTH	IMM	NAT

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. N=450. All specifications include year dummies, region dummies and trends and controls.

Table 1.6: Instrumental variables results rental/purchase prices - different population groups

Column 4 use population growth as the main regressor and instruments it with

[25]For these specifications, last two rows of the tables display the Angrist-Pischke F test of excluded instruments. Angrist & Pischke (2009) (paged 217-18) introduced first-stage F statistics (the AP tests) for tests of weak identification when there is more than one endogenous regressor. In contrast to the Cragg-Donald and Kleibergen-Paap statistics, which test the identification of the equation as a whole, the AP first-stage F statistics are tests of whether one of the endogenous regressors is weakly identified. They are in both cases over 10 and statistically significant.

[26]According to Angrist & Pischke (2009), tackling two causal questions at the same can raise serious issues. Moreover, a second endogenous variable would fall into the category of "bad control". http://www.mostlyharmlesseconometrics.com/2010/02/multiple-endogenous-variables-what-now/.

both ratios. The instruments are exogenous (as indicated by the Hansen test) but not too strong (the K-P statistic is around 7.5). Total population change has an effect on both rental and purchase prices (only on purchase prices in the specification with province dummies A.13). The coefficient is smaller for rents than for purchase prices, and significant only at the 10% level. This estimate would inform us about the effect of total changes in population, stemming from both internal migrations, immigration from abroad and natural population growth, but does not tell us anything about the effect of immigration on prices. This is why in the remaining of the chapter I present the results for both the immigration ratio and the population ratio.

When we use population growth as the main regressors and instrument it with "only" one of the ratios, immigrants or natives, we are using "only" the variation on population growth that can be attributed to exogenous location choices of that group. This strategy allows us to isolate the effect of either natives or immigrants on prices "via" its effect on population changes. This way, we remove the bias introduced by simultaneous location of the other population group. For example, we can use the exogenous variation on the location of foreign-born (1.9) to instrument the total population changes in the provinces and try to isolate the sole effect of immigration on prices. In this case, parameter β would be the causal effect of the "growth in total population which is due to immigration inflows", because we would be using only the variation in population growth which stems from exogenous changes in the immigration ratio. In this setting, we expect the parameter β to be smaller than the one found in column 1, because it would be "only" capturing the effect of immigration through their effect on population changes.

The results confirm this intuition: the coefficients of column 5 are smaller than those of column 1 and suggest that these are upward biased because the also capture the effect of native demand. For natives only seem to have an effect on purchase prices, and in fact, they do not have any effect at all when we introduce province dummies in table A.13.

The strategy used to obtain the estimates of column 5 allows us to isolate the short-run effect of immigration on prices or the effect on prices which is due only to demand from immigrants (and not from "attracted" natives). Table 1.7 shows the estimates for the same specifications as tables 1.3 and 1.4 using population growth as the main regressor and instrumenting it with (1.9). Column 7 corresponds to column 5 of table A.13, and columns 8 to 10 show the estimates of the first-differences fixed-effects models. If we compare the estimates of our preferred specifications (column 7 for rents and column 10 for purchase prices) of table 1.7 and tables 1.3 and 1.4, we find that the estimates of the effect of immigration on price growth are around 40% lower for rental prices and 30% for purchase prices in table 1.7. Given that we found that for each 100 foreign-born, around 40 natives locate in the province in the same

Change in the log of rental prices in t	OLS			Region dummies				Province dummies		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of population in t-1 over population end t-2	0.429**	0.400	0.339	0.584**	0.570***	0.572**	0.629**	0.082	0.104	0.088
	[0.170]	[0.328]	[0.358]	[0.233]	[0.218]	[0.247]	[0.295]	[0.316]	[0.323]	[0.403]
Change in the log of purchase prices in t	OLS			Region dummies				Province dummies		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of population in t-1 over population end t-2	-0.208	0.278	0.402	0.508	1.248***	1.465***	1.301***	2.080***	2.088***	1.912**
	[0.505]	[0.647]	[0.769]	[0.317]	[0.388]	[0.443]	[0.399]	[0.604]	[0.602]	[0.841]
Observations	450	450	450	450	450	450	450	450	450	450
Time invariant controls	No	Yes	Yes	No	Yes	Yes	Yes	No	No	No
Time-varying controls	No	No	Yes	No	No	Yes	Yes	No	Yes	Yes
Region trends	No	No	No	No	No	No	Yes	No	No	Yes
LM test stat under-identification (K-P)	8.28	6.86	7.20	6.31	12.56	13.92	13.18	8.91	9.15	8.21
P-value of under-identification LM statistic	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)	24.96	8.85	7.93	24.00	30.12	30.12	32.29	27.14	26.75	31.31

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies.

Table 1.7: Instrumental variables results rental/purchase prices - population growth from immigration

year, we can attribute this bias to the induced demand from native co-locating with immigrants.

These findings seem to validate the adopted strategy in order to identify the "sole" effect of immigration on prices. They point towards the existence of an sizeable bias in previous estimates of the impact of immigration on prices (for example Sosvilla-Rivero, 2008; Gonzalez & Ortega, 2009; García-Montalvo, 2010, for the Spanish case), because they disregard the causal relationship between immigrants and native location, or at least suggest a misinterpretation of the coefficient. Coefficients obtained with the "standard" instrumental variables strategy would be capturing the "long-run" effect of immigration, i.e. the combined effect of immigrants and natives. My approach, to use the variation in population changes driven by immigration, allows us to estimate the "short-run" or net effect, in other words, the effect on prices via demand that stems only from immigrants. The comparison of the coefficients obtained with the two methodologies informs us about the effect of induced native demand. Moreover, as done below, it is also important to control for supply to be able to disentangle the net effect of immigration demand, because in the long-run supply can also adjust. In the following sections I interpret the size of the coefficients and then test the robustness of these results.

1.3.4 Interpretation of the size of the coefficients

Table 1.8 reproduces the main results of sections 1.3.1.2 and 1.3.3 to ease the comparison with the results from the robustness checks (presented in section 1.3.5 below) and to help with the interpretation of the size of the coefficients. It displays, for rental and purchase prices, the estimates for the models with regional dummies (columns 1 and 2) and with province dummies (columns 3 and 4); with and without inclusion of regional trends. The four first rows show the results on the growth of rental prices and the last four on the growth of purchase prices.

During the period of analysis, 2001 to 2010, purchase prices grew an average of 7.1% and rental prices grew an average of 3.3% (see table A.7 for more details). In total, during this period, purchase prices grew over 81% and rental prices around 35%. The average population growth during the period was 1.25% while the average immigration ratio was 1.05% (again, see table A.7 for more details). Between January 2001 and January 2010 population in Spain increased 14.4%, while the total change in foreign-born with respect to initial population was 11.3%.

In the most demanding significant results of table 1.8, we find an elasticity of housing purchase prices with respect to the population ratio (which is equivalent to population growth) of 1.9 and an elasticity of rental prices of 0.63. The elasticity with respect to the immigration ratio is 2.7 for purchase prices and 1.05 for rental prices. We can combine this elasticities with the numbers above to interpret the relative importance of immigration on total grow of house prices.

Thus, our findings suggest that the average annual growth in population caused an average annual growth in purchase prices of 2.34% and in rents of 0.8%. This is obtained multiplying the elasticities (1.9 and 0.63) by the average growth in population (1.25%). This is around one third of the total average annual growth of purchase prices and around one quarter of the total average annual growth of rental prices. These proportions are quite substantial. The relative importance of immigration on house price growth is even higher if we use the elasticities of prices with respect to the immigration ratio (e.g. "long-run" or total effects), as these are larger.

Change in the log of rental prices in t	REGION DUMMIES		PROVINCE DUMMIES	
	(1)	(2)	(3)	(4)
Inflow of immigrants in t-1 over population end t-2	0.946**	1.047**	0.151	0.126
	[0.426]	[0.504]	[0.472]	[0.579]
Inflow of population in t-1 over population end t-2	0.572**	0.629**	0.104	0.088
	[0.247]	[0.295]	[0.323]	[0.403]
Change in the log of purchase prices in t	REGION DUMMIES		PROVINCE DUMMIES	
	(1)	(2)	(3)	(4)
Inflow of immigrants in t-1 over population end t-2	2.131***	1.847***	3.028***	2.720**
	[0.700]	[0.652]	[1.010]	[1.264]
Inflow of population in t-1 over population end t-2	1.465***	1.301***	2.088***	1.912**
	[0.443]	[0.399]	[0.602]	[0.841]
Region trends	No	Yes	No	Yes

Clustered (province) standard errors in brackets. t=2002/2010/ All specifications include year dummies and controls. N=450. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table 1.8: Instrumental variables results rental/purchase prices - summary of main results

In what respects to the total growth in prices between 2001 and 2010, population growth cause an increase in purchase prices of 27.4% and an increase in rental prices of around 9%. These also correspond to around one third and one quarter of the total growth of purchase and rental prices between 2001 and 2010. Again, these numbers could seem too large as a single factor (immigration inflows) would be explaining a substantial proportion of the growth in the price of housing.

Given the magnitude of the immigration inflows and price increases experienced during the period of analysis described in section 1.2.5, these proportions could in fact be quite reasonable. Actually, approximately two thirds of the growth in purchase prices and three quarters of the growth of rental prices would be explained by other factors than immigration, like supply rigidity, speculative demand, empty dwellings, changes in the cost of construction (taxes, land prices, materials), etc. Therefore, the is still an important part of the growth of house prices which is not explained by immigration.

1.3.5 Robustness checks

In this section I present the robustness checks carried out in order to check the validity of my findings. The estimates are compared to those of table 1.8. Fernández-Huertas et al. (2009) argue that one should focus only on immigrants aged between 16-64 (economically active) because they are the ones taking decisions related to the housing and the labour markets. Gonzalez & Ortega (2009) also restrict their sample to working-age foreign-born. Table 1.9 shows the results for the comparable specifications of table 1.8 but restricting the foreign-born and the total population only to people aged 16-64. Both the results using the immigration ratio and the population growth are displayed. The coefficients are higher than for the baseline results, specially for purchase prices. But given the size of the standard errors, the point estimates of table 1.9 are included in the confidence interval of the estimates of table 1.8.

Change in the log of rental prices in t	Region dummies (1)	Region dummies (2)	Province dummies (3)	Province dummies (4)
Inflow of immigrants in t-1 over population end t-2	1.329**	1.533**	0.196	0.170
(working-age)	[0.610]	[0.747]	[0.614]	[0.784]
Inflow of population in t-1 over population end t-2	0.817**	0.930**	0.140	0.125
(working-age)	[0.354]	[0.433]	[0.437]	[0.575]
Change in the log of purchase prices in t	**Region dummies (1)**	**Region dummies (2)**	**Province dummies (3)**	**Province dummies (4)**
Inflow of immigrants in t-1 over population end t-2	2.969***	2.668***	3.921***	3.686**
(working-age)	[1.007]	[0.995]	[1.322]	[1.746]
Inflow of population in t-1 over population end t-2	2.091***	1.925***	2.815***	2.727**
(working-age)	[0.644]	[0.605]	[0.825]	[1.233]
Region trends	No	Yes	No	Yes

Clustered (province) standard errors in brackets. t=2002/2010. All specifications include year dummies and controls. N=450. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 1.9: Instrumental variables results rental/purchase prices - population aged 16-64

Table 1.10 shows the results for changes in rental and purchase prices using the immigration ratio and population growth during the same period as the growth in prices, as opposed to the lagged ratio used in the main results. The results for total and only working-age population are displayed. The effect of changes on rental prices remain positive and significant and of very similar size to those of tables 1.8. However, the results on purchase prices become very imprecise. This suggests that immigrants choose renting as the tenancy status when they initially settle in the country while the effect on purchase prices is lagged because they either take some time to be able to afford buying a house or the indirect effect on prices (via purchase by native in order to rent them to immigrants) is also delayed (a native buys a house and the following year rents it to an immigrant).

Table A.14 estimates the same relationship as in table 1.4 but using different data sources for the purchase prices. The top panels display the results using the data-

Change in the log of rental prices in t	Region dummies (1)	(2)	Province dummies (3)	(4)
Inflow of immigrants in t over population in end t-1	0.994*	1.119*	-0.042	0.012
	[0.531]	[0.586]	[0.425]	[0.416]
Inflow of population in t over population in end t-1	0.618*	0.691**	-0.029	0.008
	[0.324]	[0.352]	[0.298]	[0.290]
Inflow of immigrants in t over population in end t-1	1.416*	1.645*	-0.055	0.016
(working-age)	[0.747]	[0.860]	[0.557]	[0.565]
Inflow of population in t over population in end t-1	0.920*	1.052**	-0.042	0.012
(working-age)	[0.471]	[0.526]	[0.420]	[0.425]
Observations	400	400	400	400
Change in the log of purchase prices in t	Region dummies (1)	(2)	Province dummies (3)	(4)
Inflow of immigrants in t over population in end t-1	1.315	1.029	1.107	0.631
	[1.655]	[1.541]	[1.992]	[1.926]
Inflow of population in t over population in end t-1	0.796	0.618	0.761	0.441
	[0.985]	[0.913]	[1.344]	[1.331]
Inflow of immigrants in t over population in end t-1	1.848	1.506	1.435	0.854
(working-age)	[2.340]	[2.259]	[2.589]	[2.603]
Inflow of population in t over population in end t-1	1.136	0.914	1.026	0.628
(working-age)	[1.416]	[1.356]	[1.823]	[1.900]
Observations	450	450	450	450
Region trends	No	Yes	No	Yes

Clustered (province) standard errors in brackets. t=2002/2009 for rents, t=2001/2009 for purchase. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies and controls. Time-varying controls in t-1.

Table 1.10: Instrumental variables results rental/purchase prices - contemporaneous inflows

base from IVIE. The left top panel restricts the sample to 2007, i.e., I do not use the extended sample for 2008-2010 that I constructed by expanding the original data with the Department of Housing data province annual growth rates. Excluding years 2008 to 2010 removes the year of the housing "crises". This limits the variation of the data because it only uses the information on the "boom" years, in which the both immigration and prices were increasing greatly. This is possibly why when we control by regional trends the estimates become insignificant. The right top panel extends the sample to include observations from 1999.[27] The estimates are less significant than in the baseline results even if the standard errors are very similar to before. The estimates become even insignificant in the first-differences fixed-effect estimation. This would point towards the existence of an attenuation bias due to measurement error on the population data before 2001, as explained in section 1.2.4. The bottom panels show the results using the data provided by the Housing Department (average of the four quarters and only the 2^{nd} quarter). The coefficients are smaller and less precise than those of table 1.8 but reasonably similar.

Table A.19 adds a second instrument to be able to test the exogeneity of the in-

[27] Population and immigration inflows can be constructed from 1998, but data availability on some of the time-varying regressors forces us to start using data from 1999. In any case, results not including time-varying controls using data from 1998 are very similar.

strument by means of the Hansen J statistic. I constructed a second shift-share instrument where I calculated the predicted inflow to a given province in a given year of each nationality, (1.5), as the product of the inverse distance between the country centroid to Madrid plus the euclidian distance from province i to Madrid (share) and the national inflow is the inflow of this nationality to Italy[28] (shift).

I use inverse distance to Spain to compute the prediction, inspired by Ottaviano & Peri (2006), who use the distance from the closest gateway into the US in the construction of the instruments for immigration[29]. Additionally, during the period 2001-2008 the European countries experience intense migration inflows (Pekkala-Kerr & Kerr, 2011). I use the inflow from Italy because this country is not "too far" from Spain in terms of distance, culture and economic conditions, because it had high rates of immigration during these years (Buonanno et al., 2011) and because is one of the few countries in the "OECD International Migration Statistics" dataset for which we have fewer missing values. This instrument is not strong enough by itself (the F-stat of the first stage is around 4.7) but as it is based on different variation sources as our main instrument it is sufficiently good to be used as a second instrument to allow for the testing of the orthogonality conditions. The last row of table A.19 shows the p-values of the Hansen test which in all cases confirm the exogeneity of our instrument. The coefficients are very similar to those of table 1.8, even though the estimates for the effects on purchase prices are a bit smaller. Additional robustness checks are available in the appendix (section A.4).

1.3.6 The role of housing supply

In this section I explore the role played by the supply on potentially mitigating the increase in prices[30]. This analysis applies mostly to the effect on purchase prices, because dwellings have always to be "bought" before they go to the rental market. I provide the results for rental prices in the appendix for completeness but I only discuss the results for purchase prices in this section. First I include some extra time-invariant controls in the models to control for different price growth trends based on some attributes of the provinces related to supply. Additionally, the growth in the stock of (private) dwellings is added as an extra control. Given that this variable

[28]The data sources for the construction of this instrument are the "OECD International Migration Statistics" for data on the stock and inflows of foreign born by nationality during 2001-2008 and the CEPII gravity database for the distance from the country to Spain. The internal distance of Madrid is calculated as $(2/3) * \sqrt{(area/\pi)}$.

[29]I also used distance to the closest port of entry, being the ports the 5 airports which according to the Spanish Airports Regulator data on airport traffic in 2000. According to the Spanish National Statistical Institute 63% of the immigrants between 1998 and 2010 arrived in Spain by plane. The results are very similar, mainly due to the fact that the majority of entries are through the Madrid airport.

[30]Section A.5 in the appendix explores the relationship of immigration with the construction sector.

is quite likely to be endogenous, I finally construct an instrument for this variable.

Several attributes related to housing supply, and therefore to prices, are added to the basic regression model of equation (1.1). These are time-invariant and control for different trends on the growth of prices and foreign-born population. I include the share of developable land in the province, orographic characteristics of the terrain (height and ruggedness) and the percentage of rented and empty dwellings over total dwelling in the province in 2001.

I control for the share of developable land because in these provinces land would be potentially cheaper (for a given demand of land) so construction of new homes could be more intense. Total area and total developable area[31] were calculated using GIS and raster maps of land use year for 2000, provided by the *Corine* Land Cover data project (European Environment Agency). I use a relative index (Spain=100) of average terrain height and ruggedness of each province, which are obtained from Goerlich-Gisbert & Cantarino-Martí (2010). Orographic characteristics of the terrain have been commonly used as proxies for supply constraints (Hilber & Vermeulen, 2009; Saiz, 2010). Provinces in which renting is more common can have different trends in the growth of supply. The percentage of rented properties over total occupied properties is obtained from 2001 Census data from the Spanish National Statistical Institute (INE). I also include the proportion of empty homes over total homes. Prices in provinces in which the proportion of unoccupied dwellings is larger could be growing at a slower rate because the supply of homes is higher in these locations.

I also calculated the stock of (private) dwellings in the different years (from 1999 to 2010) combining data from the Housing Department. Data on the housing stock is only available from 2001. Using the entry and exit flows I calculated a rate of depreciation and I updated the stock of the dwellings for years prior to 2001 combining the depreciation rate and construction of dwellings data. I focus on private dwellings, but the results using total dwellings are very similar.

Table 1.11 shows the results on purchase and rental prices adding this extra supply controls. Columns 1 to 6 show the results for the immigration ratio and columns 7 to 12 for the population growth. All specifications instrument the immigration or the population growth with (1.9). Columns 1 and 7 show the results for the specification using regional dummies and the set of original province attributes. The equivalent coefficient for this specification in tables 1.4 and 1.7 is that of column 5, which includes dummies and time-invariant province attributes.

Columns 2 and 8 add the province time-invariant characteristics related to sup-

[31] The categories included in developable land are: Green urban areas, Non-irrigated arable land, Permanently irrigated land, Rice fields, Vineyards Fruit trees and berry plantations, Olive groves, Pastures, Annual crops associated with permanent crops, Complex cultivation patterns, Land principally occupied by agriculture, Agro-forestry areas, Broad-leaved forest, Coniferous forest, Mixed forest, Natural grasslands, Moors and heartland, Sclerophyllous vegetation and Burnt areas.

Change in the log of purchase prices in t	REGION DUMMIES				PROVINCE DUMMIES		REGION DUMMIES				PROVINCE DUMMIES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Inflow of immigrants in t-1 over population end t-2	2.131***	1.919**	2.416***	1.867*	3.228***	2.764**						
	[0.700]	[0.749]	[0.865]	[1.038]	[1.018]	[1.386]						
Inflow of population in t-1 over population end t-2							1.248***	1.249***	1.714***	1.331*	2.234***	1.927**
							[0.388]	[0.467]	[0.588]	[0.701]	[0.632]	[0.958]
Coast dummy	-0.004	-0.000	0.000	-0.000			-0.003	-0.001	-0.001	-0.001		
	[0.007]	[0.005]	[0.005]	[0.005]			[0.006]	[0.005]	[0.006]	[0.006]		
Length of coastline	-0.000	-0.000	0.000	0.000			-0.000	-0.000	0.000	0.000		
	[0.000]	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]	[0.000]		
Log of hours of average temperature (January)	0.013	0.014	0.014	0.015			0.021**	0.020*	0.024*	0.022*		
	[0.008]	[0.009]	[0.010]	[0.010]			[0.010]	[0.012]	[0.014]	[0.012]		
Log of mm of rain precipitation (January)	0.010	0.002	-0.005	-0.002			0.008	0.005	-0.003	-0.001		
	[0.007]	[0.005]	[0.006]	[0.006]			[0.007]	[0.008]	[0.007]	[0.006]		
Log of number of retails shops in 2000	0.013**	0.006	0.003	0.003			0.009*	0.006	0.002	0.002		
	[0.006]	[0.006]	[0.007]	[0.005]			[0.005]	[0.006]	[0.008]	[0.005]		
Log of number of restaurants and bars in 2000	-0.001	-0.000	0.001	0.001			-0.001	-0.001	-0.000	0.000		
	[0.002]	[0.003]	[0.003]	[0.002]			[0.002]	[0.003]	[0.003]	[0.002]		
Importance of tourism sector - comparative index 2000	-0.000***	-0.000***	-0.000***	-0.000***			-0.000***	-0.000*	-0.000	-0.000*		
	[0.000]	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]	[0.000]		
Log of the number of doctors - 2000	-0.003	-0.002	-0.001	-0.001			-0.004*	-0.004	-0.002	-0.002		
	[0.002]	[0.002]	[0.002]	[0.002]			[0.002]	[0.003]	[0.003]	[0.002]		
Log of the surface of natural parks - 2000	0.001	0.001	0.001	0.001			0.001	0.001	0.001	0.001		
	[0.002]	[0.002]	[0.002]	[0.002]			[0.002]	[0.003]	[0.003]	[0.002]		
Share of developable land in 2000 (Corine)		0.064***	0.082***	0.067**				0.041**	0.059***	0.048***		
		[0.018]	[0.023]	[0.027]				[0.019]	[0.018]	[0.018]		
Relative index of altitude		-0.000*	-0.000**	-0.000*				-0.000	-0.000	-0.000		
		[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]		
Relative index of ruggedness		0.000	0.000	0.000				0.000	0.000	0.000		
		[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]		
Percentage of rented properties in 2001		0.050	0.108	0.083				-0.012	0.044	0.035		
		[0.095]	[0.099]	[0.096]				[0.120]	[0.117]	[0.103]		
Percentage of empty homes in 2001		0.078	-0.000	-0.020				0.201	0.150	0.095		
		[0.158]	[0.156]	[0.156]				[0.184]	[0.173]	[0.182]		
Change of log of GDP in t-2			-0.097	-0.138**	-0.119	-0.153**			-0.143*	-0.174**	-0.138*	-0.172**
			[0.078]	[0.062]	[0.082]	[0.068]			[0.086]	[0.069]	[0.080]	[0.067]
Change of log of unemployment rate in t-2			-0.111	0.011	-0.107	0.017			-0.100	0.018	-0.079	0.041
			[0.081]	[0.063]	[0.079]	[0.067]			[0.081]	[0.061]	[0.078]	[0.067]
Change of log of number of credit establishments in t-2			-0.184	-0.112	-0.197	-0.122			-0.194	-0.129	-0.182	-0.116
			[0.142]	[0.128]	[0.152]	[0.138]			[0.141]	[0.130]	[0.144]	[0.133]
Change of percentage of savings banks in t-2			0.137	-0.115	0.162	-0.097			0.104	-0.153	0.171	-0.100
			[0.164]	[0.145]	[0.168]	[0.156]			[0.165]	[0.137]	[0.165]	[0.149]
Log of change of stock of private dwellings in t-2			-0.340	-0.138	-0.253	-0.036			-0.499	-0.263	-0.280	-0.030
			[0.428]	[0.375]	[0.568]	[0.429]			[0.473]	[0.450]	[0.569]	[0.426]
Time invariant controls	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Supply time invariant controls	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No
Time-varying controls	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Region trends	No	No	No	Yes	No	Yes	No	No	No	Yes	No	Yes
Observations	450	450	450	450	450	450	450	450	450	450	450	450
LM test stat under-identification (K-P)	12.66	9.04	9.34	9.38	8.06	7.80	12.56	9.90	10.62	10.00	9.63	8.76
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)	24.50	18.80	15.67	14.82	16.76	15.63	30.12	21.04	16.52	14.30	30.29	30.28

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * p<0.05, ** p<0.01, *** p<0.001. N=450

Table 1.11: Instrumental variables results purchase prices - adding supply controls

ply. The estimated coefficients are 1.9 for immigration and 1.25 for population. The inclusion of these additional province trends, which are jointly highly significant, reduces the coefficient significantly for immigration, but it still remains significant at 10% level. This suggests that the coefficient were upward biased and that these variables jointly control for factors which make prices grow faster. For the population growth, the coefficients of specifications 7 and 8 are mostly identical. The coefficient of the share of developable land is positive and significant, suggesting that in provinces which had more developable land with respect to the region mean had faster price growth. This could be because there was some speculative component and this provinces were becoming more attractive, and therefore prices were growing faster, because there was potential to built new real estate development. The index of altitude is also significant but negative, even if the coefficient is very small. Province which have higher altitude, with respect to the regional mean, had slower growth in the prices.

Columns 3 and 9 introduce the time-varying controls, including the supply of housing (change in the stock of private dwellings)[32]. The coefficient now decreases with respect to that of columns 2 and 8. The coefficient on the growth of housing stock is negative but insignificant. Columns 4 and 10 add the regional trends. The remaining columns include province fixed-effects and controls. The only difference of these results with respect to those of columns 3 and 4 of table 1.8 are the inclusion of the growth of housing stock. The coefficients are very similar, but bigger, which could be due to the fact that construction coefficient is negative, although insignificant.

However, using the growth of housing stock as a control variable in table 1.11 could be problematic if more dwellings are constructed because developers expect house prices to rise in the future, which is likely to happen in periods of high price increases[33]. In order to mitigate the potential endogeneity of this regressor, I construct an instrument for the stock of private housing in a given province. I use a similar instrument as Amior (2011) and Saiz (2010). I construct a predicted stock of housing combining the share of developable land in the provinces in 2000 (for the initial spatial distribution) and the changes in total annual national stock. I drop the share of developable land from the supply time-invariant attributes controls and use the two instruments (for immigration and for construction) in the models that estimate the effects of immigration on prices.

Results using this are presented in table 1.12. The tables display the results for the immigration ratio (columns 1 to 4) and for population growth (columns 4 to 4). The

[32] I tried other supply proxies like log of gross inflow or ratio of inflow over stock, with very similar results.

[33] Immigrant can also have a direct impact on dwelling construction, so the growth of housing stock is a "bad" control by definition. This issue is explored in section A.5 in the appendix.

Change in the log of purchase prices in t	REGION DUMMIES (1)	REGION DUMMIES (2)	PROVINCE DUMMIES (3)	PROVINCE DUMMIES (4)	REGION DUMMIES (5)	REGION DUMMIES (6)	PROVINCE DUMMIES (7)	PROVINCE DUMMIES (8)
Inflow of immigrants in t-1 over population end t-2	2.823***	2.457*	3.188***	2.849**				
	[1.046]	[1.279]	[1.000]	[1.380]				
Inflow of population in t-1 over population end t-2					1.801***	1.532**	2.197***	1.974**
					[0.591]	[0.721]	[0.627]	[0.948]
Coast dummy	-0.003	-0.003			-0.002	-0.002		
	[0.005]	[0.005]			[0.006]	[0.006]		
Length of coastline	-0.000	-0.000			-0.000	-0.000		
	[0.000]	[0.000]			[0.000]	[0.000]		
Log of hours of average temperature (January)	-0.006	-0.003			0.010	0.012		
	[0.016]	[0.016]			[0.015]	[0.013]		
Log of mm of rain precipitation (January)	0.002	0.003			0.003	0.003		
	[0.008]	[0.007]			[0.009]	[0.008]		
Log of number of retails shops in 2000	0.011	0.009			0.008	0.007		
	[0.009]	[0.008]			[0.008]	[0.006]		
Log of number of restaurants and bars in 2000	0.001	0.001			-0.001	-0.000		
	[0.003]	[0.003]			[0.003]	[0.003]		
Importance of tourism sector - comparative index 2000	-0.000***	-0.000***			-0.000	-0.000		
	[0.000]	[0.000]			[0.000]	[0.000]		
Log of the number of doctors - 2000	-0.002	-0.002			-0.004	-0.003		
	[0.003]	[0.002]			[0.003]	[0.003]		
Log of the surface of natural parks - 2000	0.001	0.001			0.001	0.001		
	[0.003]	[0.003]			[0.003]	[0.003]		
Relative index of altitude	-0.000*	-0.000*			-0.000	-0.000		
	[0.000]	[0.000]			[0.000]	[0.000]		
Relative index of ruggedness	0.000	0.000			0.000	0.000		
	[0.000]	[0.000]			[0.000]	[0.000]		
Percentage of rented properties in 2001	0.038	0.039			-0.019	-0.010		
	[0.128]	[0.118]			[0.142]	[0.127]		
Percentage of empty homes in 2001	0.085	0.057			0.217	0.164		
	[0.206]	[0.210]			[0.208]	[0.220]		
Change of log of GDP in t-2	-0.112	-0.146**	-0.119	-0.152**	-0.157*	-0.186***	-0.138*	-0.171**
	[0.083]	[0.067]	[0.082]	[0.069]	[0.089]	[0.072]	[0.080]	[0.068]
Change of log of unemployment rate in t-2	-0.126	-0.003	-0.107	0.016	-0.110	0.011	-0.079	0.040
	[0.081]	[0.064]	[0.078]	[0.067]	[0.081]	[0.061]	[0.078]	[0.067]
Change of log of number of credit establishments in t-2	-0.203	-0.132	-0.197	-0.123	-0.213	-0.147	-0.182	-0.116
	[0.145]	[0.134]	[0.152]	[0.139]	[0.142]	[0.133]	[0.145]	[0.133]
Change of percentage of savings banks in t-2	0.086	-0.160	0.163	-0.097	0.069	-0.190	0.173	-0.100
	[0.168]	[0.139]	[0.168]	[0.157]	[0.168]	[0.133]	[0.165]	[0.149]
Log of change of stock of private dwellings in t-2	-0.191	-0.149	-0.200	-0.153	-0.276	-0.199	-0.208	-0.124
	[0.550]	[0.462]	[0.684]	[0.552]	[0.588]	[0.502]	[0.688]	[0.541]
Time invariant controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region trends	No	Yes	No	Yes	No	Yes	No	Yes
Observations	450	450	450	450	450	450	450	450
LM test stat under-identification (K-P)	7.72	7.20	8.01	7.76	10.03	9.16	9.55	8.68
P-value of under-identification LM statistic	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)	5.71	4.81	7.88	7.44	8.32	6.86	14.26	14.22
A-P F-test of excluded instruments (immigration)	14.82	12.85	15.70	14.92	19.74	16.00	28.30	28.61
A-P F-test of excluded instruments (supply)	63.08	58.93	51.22	47.79	59.08	55.34	51.03	48.69

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * p<0.05, ** p<0.01, *** p<0.001. All specifications include year dummies, controls and region trends. N=450

Table 1.12: Instrumental variables results purchase prices - instrumenting supply

specifications include regional dummies and province dummies, with and without regional trends, and all include controls. The tables display the Angrist-Pischke and the Kleibergen-Paap F-test which support the strength of the instruments. The estimated coefficients of the effects of immigration are very similar to that of table 1.11, marginally bigger. The coefficients of the growth of dwellings stock is always negative but never significant. These results suggest that the increase in supply through construction of new dwellings did not have a causal impact on the mitigation of the growth of house prices. As discussed before, possible explanations to this could be the high number of non-occupied dwellings in Spain or the reduced financial capabilities of immigrants to buy a property, either because they earn less or have less job stability (Carrasco et al., 2008) or because they are subject to tighter credit constraints than natives (Díaz-Serrano & Raya, 2011).

1.4 Conclusions

This paper provides causal estimates of the effect of immigration on house prices (rental and purchase prices). Panel data for the Spanish provinces during the period 2001-2010 is used to obtain estimates on the elasticity of prices with respect to change in the immigration ratio (immigration inflows over population). I use fixed-effects and instrumental variables estimation techniques in order to be able to tackle the potential endogeneity of the foreign-born location patterns and the growth in prices. The equations estimated is based in Saiz (2007) and the estimates are very similar to those found by him in the US context. They also confirm results obtained in similar analysis (for example Gonzalez & Ortega, 2009). However, in this type of analysis the relationship between immigrants and native location patterns is overlooked. To be able to interpret the size of the estimated coefficients I then estimate the "displacement effect" and I find a strong causal positive relationship between immigrant and native location. This would suggest that estimates that do not take this fact into account overestimate the "net" effect of immigration on prices.

I then use population growth as the main regressor and instrument it with a prediction of the immigration inflows and stocks based on exogenous variation. This allows me to use only the variation in population growth which "stems from immigration". Approaches such as that of Saiz (2007) or Gonzalez & Ortega (2009) interpret coefficient β as the effect of immigration on prices, which is correct, but these authors are unable to separate the effects of immigration on prices via immigrant demand from the effects on prices via immigrants affecting native displacement and thus provinces total housing demand. This fact links back to Saiz (2007) discussion on the channels through which immigration affects house prices and the existence of "long-run" and "short-run" impacts.

The estimates obtained with my approach are between 30 and 40% smaller than those of the standard approach, i.e. those that use the immigration ratio as the main regressor. This would suggest that, if we are interpreting β as the demand effect of immigrants on prices, previous estimates were "too high" because they were capturing the combined demand effect of immigration plus native induced migration. This 30–40% difference in the coefficients would be due to induced demand by native attracted to the same regions in which the immigrants are locating. Quantitatively, my estimates predict that around one third of the total average annual growth of purchase prices and around one quarter of the total average annual growth of rental prices can be attributed to the effect of immigration on housing demand.

Finally, I this chapter I also explore the role played by the increase in supply on mitigating price growth cause by the increase in demand. I find no effect of supply on prices, even after instrumenting it. This could be cause because prices were growing based on expectations of future growth, but this statement is difficult to test with the currently available data.

The first limitation of my analysis is its external validity. The relationship between house prices and immigration is studied for a very particular period of house prices boom and bust and intensive immigration inflows. The panel is too short to be able to divide the data in sub-panels which would allows us to study the different relationships when prices are increasing or decreasing. A second limitation is the use of spatial unit: in order to have a sufficiently long panels and a reasonable set of control variables I use provinces as the geographical level of observation. Ideally I could have used municipalities or urban areas to be able to study effects which only occur at smaller spatial scales, but data is much more limited at this scale. A third limitation is the lack of richer data on foreign-born population characteristics (demographic and socioeconomic). If foreign-born that settle in Spain are less skilled than natives it is likely that they could have a different effect on prices because their purchasing power is lower. The Spanish Labour Force Survey provides data on a sample of workers by nationality and skills, but it only captures "legal" immigrants so it under-estimates the total number of foreign-born setting down in the country. I believe the population registers capture better the actual size of the population variables but the measure of skills in this database is very poor[34]. A fourth data driven limitation is the use of average purchase and rental prices which does not take into account neither occupancy or any characteristic of the house (quality). If prices are growing because "better" houses are being constructed in some regions but not in others we could be over estimating the effect of immigration. The same would happen in secondary homes are concentrated in some provinces and they are pushing

[34] Essentially, the skill level categories are very aggregated and they are not updated, only recorded when the register is created for the first time.

prices up.

Finally, I have estimated separate models for rental and purchase prices even if this two prices are very related to each other. If rental prices are relatively high and credit is cheap, individuals would prefer ownership towards renting (specially given that ownership ratios in Spain are among the highest in the world, over 80%). At the same time, if purchase prices are high or credit constraints increase, individuals could start renting more properties increasing demand. So demand for purchase and rents can change because population increases or because demand from one type of tenancy is switched to the other type. The current analysis does not capture this.

This set of limitations leave some questions open for future research. From 2004, when the Housing Department was instituted as a different government department separated from the Ministry of Public Works, the quality and quantity of housing market data available to researchers increased substantially. Agreements with the College of Property Registrars (*Colegio de Registradores de la Propiedad*) and with the Professional Association of Property Valuators (*Asociacin Profesional de Sociedades de Valoración-ATASA*) were set in order to make available richer statistics on housing construction, valuation and transaction prices. The Housing Department was re-integrated in the Ministry of Public Works in 2010, but if the current effort on collecting more and better statistics continues this would provide the necessary data to provide some interesting new insights on some of the under-investigated determinants of house price growth in the future.

Data disclaimer

The work of chapters 2, 3 and 4 is based on data from BSD, ARD, ASHE and LFS, produced by the *Office for National Statistics* (ONS) and supplied by the *Secure Data Service* (SDS) at the *UK Data Archive* and by the *Virtual Microdata Laboratory* (VML).

 The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

All the results have been granted final clearance by the staff of the VML and by the staff of the SDS.

Additionally, any interpretations or opinions expressed in this presentation are those of the authors and do not necessarily reflect the views of the *Department for Transport* who also provided data to the study.

Chapter 2

Economic Impacts of Transport Policy: Methodology

2.1 Introduction

Road network is a hugely important part of infrastructure in all countries. According to Transport Statistics Great Britain, between 1998 and 2008, more than 90% of passenger transport was done by road. As shown in table B.1, in 2001 over 80% of the commutes in the UK were done by motor vehicles (cars, vans, buses or motorcycles), and this percentage increases in city size[1]. Moreover, in these years, around 65% of goods transport was also done by roads[2].

New transport projects require large amounts of public (and private) investment and they generate controversy about their placement. Between 1998 and 2008, investment in road infrastructure increased in Great Britain more than 60% in nominal terms and around 40% in per capita terms. Investment in rail also increased substantially (see table B.2). Investment in roads accounted for over 40% of total infrastructure investment (roads, trains, ports and airports). As shown in table B.3, the trunk and motorway network was extended by around 300 kilometres between 1999 and 2009[3].

Understanding the relationship between transport improvements and economic outcomes is essential to the design of transport policy, and given the importance of road transportation for the movements of people and goods, the correct identification of the impact of road investments is also important for economic policy as a

[1]Except for cities of over 1,000,000 inhabitants where commutes by train or tube represent a sizeable fraction.

[2]Especially transportation of agricultural products and live animals, of foodstuffs and animal fodder, of crude, of manufactured minerals and of building materials.

[3]Transport Statistics Great Britain. We compare the length after 1999 because many changes between 1998 and 1999 corresponded to reclassification of the trunk roads, as explained in section 2.4.

whole. In the current and following chapters we investigate the causal impact of road improvements on aggregate and individual economic outcomes using micro data on firms and workers for Great Britain between 1998-2008.

The theoretical predictions on the net effects of transportation improvements on aggregate and individual economic outcomes are numerous. We analyse the effect on ward employment, number of plants and productivity (aggregate outcomes), firms employment, productivity, gross output and wages (firm outcomes) and workers wages, hours worked, employment status and travel time (labour market outcomes).

Better transport infrastructure brings places and people closer together. This has two effects on the actual size of the market. Firstly, for a given location of firms and workers, effective density increases, as it becomes easier to reach other locations using the improved transportation network. Secondly, new infrastructure changes the attractiveness of locations, which may boost spatial concentration if firms and workers relocate. These effects may reinforce each other and create positive agglomeration spillovers (as discussed by Ottaviano, 2008). On the other hand, improved access to markets also strengthens competition, forcing the exit of the less productive firms and thus increasing aggregate productivity (Melitz, 2003). Finally, firms use transport services as a production input, so changes in the supply and relative prices of transport affect the input mix used by the firms and their demand for other inputs, for example labour.

The effects on labour markets also operate through multiple channels (the effects of transport on labour market outcomes are discussed extensively in Gibbons & Machin, 2006). First of all, reduced transport costs bring employers and workers closer together. Transport improvements change commuting costs and hence affect the size of labour markets. According to the so called "spatial mismatch hypothesis", accessibility to jobs is an important determinant of labour market participation. For unemployed or inactive workers decreased commuting costs modify search costs and reservation wages, and in this way, they can help reduce frictional unemployment and increase the employment probabilities of those who are jobless. Secondly, improvements in the transport network can increase the scope of agglomeration economies, as for a given physical distance employers and employees are nearer to each other. Increased competition for jobs and workers (tighter labour markets) can foster productivity, and additionally we could see more and better job matches taking place in these enlarged labour markets. Finally, transport investments could be capitalised into non-labour prices and affect congestion, which in turn could affect location decisions of firms and workers.

Yet, the number of empirical studies that have tried to quantitatively assess these effects is still limited. Given the unclear net theoretical predictions, the size and direction of the effects of transport policy on economic outcomes is mainly an empirical question (Gibbons & Machin, 2006).

In this chapter we describe the methodology used in order to identify the effects that transport improvements might have on firm and worker economic outcomes. The empirical exercises are performed in the next two chapters. We also provide the theoretical background and review the existing empirical evidence that motivated the study. Then we describe the measure we constructed to capture the effect of transport improvements. The measure we use is an index of accessibility to employment or effective density. It captures the amount of employment which is reachable from a given location using the road network at every point in time. Improvements undertaken in the road network may affect travel times between locations, and thus directly affect effective density.

For this purpose, we construct a novel data set of road improvements carried out in Great Britain at different points in time during the period 1998-2008. We combine this data with road network data and use Geographical Information System (GIS) network analysis tools to calculate optimal minimum travel times by road between locations for every year in our sample. We use this information, jointly with information on the exact location of plants and employment, to calculate accessibility to employment in different years.

We analyse the effect of improvements carried out on major roads, which cover trunk roads[4], principal roads (class A) and motorways. Even if these roads only represent 13% of total road network length (table B.3), they correspond to 65% of driven kilometres[5]. We focus on major roads for two reasons. The first one is data availability: detailed data on road projects is only available for major schemes. The second reason is that transport policy is aimed at improving economic integration and reducing congestion of wide areas (Highways Agency, 2009), so we can expect the most substantial investments to be carried out on these types of roads.

Accessibility to employment measures the amount of employment which is reachable using the road network from a given location, inversely weighed by the travel time to reach other locations. One advantage of using this measure is that it is not constrained to artificial geographical boundaries like measures based on the density of roads. Moreover, it allows us to use variation due to transport improvements which affect optimal travel times, even if the location was previously connected to the network. This measure is, thus, very appropriate in a setting like Great Britain where road density was already high at the beginning of our period of study. The construction of this measure is explained in detail in this chapter.

The geographical unit we use as the basis for our analysis is the electoral ward.

[4]Roads over which the Department of Transport has direct control.
[5]Transport Statistics Great Britain.

Wards (as defined in 1998) are quite small areas and there are over 10,500 in Great Britain. The average area of British wards is 21 square kilometres (they are smaller in England, 15 km^2, and much bigger in Scotland, 62 km^2) and the average population estimate in 2001 (for England and Wales) is around 6,000 people[6]. Having such a large number of units implies that we have variation in accessibility changes due to road improvements even when we focus on areas close to where the road improvement took place. Therefore, the fine geographical detail of our data allows us to implement a careful identification strategy.

In order to add to the scarce empirical evidence, in the next two chapters we test the impact of changes in accessibility on aggregate and individual economic outcomes. We carry out regression analysis and estimate the effects of accessibility on firm and worker economic outcomes. To be able to infer a causal interpretation our estimates we use a novel instrumental variables strategy that exploits small scale spatial variation in the way road improvements change accessibility to employment. Chapter 3 analyses the results on firm level outcomes (employment and productivity) and on aggregate outcomes (ward employment, number of plants and aggregate productivity). Chapter 4 analyses the results on individual labour market outcomes (wages, hours worked, commuting time and employment status).

In the current chapter we motivate the research and provide details of the construction of the accessibility measures, which is be used in the empirical exercises of chapters 3 and 4. Section 2.2 summarises the main theoretical channels through which transport policy might influence firm and worker economic outcomes. Section 2.3 reviews the existing empirical evidence on the effect of transport policy on economic outcomes. Section 2.4 describes the policy context. Section 2.5 explains the details of the construction of the accessibility measure and provides summary statistics. Section 2.6 summarises the identification strategy, which is be developed in more detail in chapters 3 and 4. Finally, section 2.7 concludes.

2.2 Theoretical framework

The role of transportation in the spatial distribution of economic activity and economic performance has become of increased interest to researchers in the last years. Decreasing transport costs are considered to be a central driver of economic integration and of the rise of agglomeration externalities, but solid empirical evidence on the channels through which these effects operate is still needed.

[6] This is for 2001 CAS Wards, which are slightly different from ours. This is the only information available for ward population figures, and it is approximate as it is estimated. See `http://www.statistics.gov.uk/hub/population/population-change/population-estimates` for more information.

2.2.1 Effects of transport on firms and aggregate outcomes

In chapter 3 we study the effect of transport improvements on firm level employment and productivity (total factor productivity) and on aggregate economic outcomes (ward employment, number of plants and aggregate productivity). Firms maximise profits conditional on the demand for outputs, the supply of inputs and the production technology. The degree of competition in the outputs and inputs market also affects the maximisation problem faced by the firm. Changes in transport infrastructure can influence all these conditions and therefore impact firm outcomes through several inter-connected channels[7].

At the firm level, transport improvements affect the performance of firms. On the one hand, they may improve the logistics and the internal organisation of firms, and can change the optimal input mix choice. Transportation services are used as production inputs and, if there is a substitution effect between inventories, labour and transport services, the demand and input mix could be affected (Holl, 2006). Intermediate input prices could decrease because of reduced transport costs or increased competition between the suppliers. Wages could also change if productivity effects are capitalised into wages or if wages are set as a function of commuting costs, which are affected by the transport network (Gibbons & Machin, 2006). Changes in wages or input mix can have an influence in the number of workers a firm employs.

Due to the changes in input prices firms might change the demand for inputs, and depending on the internal returns to scale, this could affect their final output. If output increases with respect to inputs more than proportionally (due to increasing returns to scale), the output/inputs ratio changes. However, this is a scale effect and total factor productivity would remain unaffected. Therefore, we could observe an effect on labour producibility (the ratio of output to employment) but not on total factor productivity (a genuine productivity shifter: firms produce more output for the same amount of inputs).

Furthermore, better accessibility to consumers increases customer base (for example by increasing market area, as suggested by Lahr et al., 2005). This allows firms to expand production and exploit economies of scale. In addition to the potential scale effects discussed above, firms total factor productivity could be affected by the wider economic benefits of transport (Graham, 2007). These refer to agglomeration externalities (sharing, matching and learning – Duranton & Puga, 2004), which can be internal to the industry (localisation economies) or driven by the size of the market (urbanisation economies). Firms benefit from the presence of other firms nearby (in the same or different sector of production) and from the increased proximity to suppliers which arises from the improvements on the transport network.

[7]Our theoretical discussion draws mainly on Gibbons & Overman (2009) who provide an extensive analysis on the potential productivity and scale effects of transport infrastructure.

Agglomeration benefits act like a production function shifter, i.e. for a given amount of inputs, the firm is able to produce more.

Transport improvements can also influence the number and location of plants. If better transport infrastructure reduces the fixed cost of creating a firm or makes better-connected locations more attractive places to start a business, the creation and location of plants can be affected by transport policy. Related to this, if firms can take advantage of the increasing returns to scale or benefit from TFP improvements, their survival rates could be affected. Low-productivity firms which would have exited the market otherwise could be able to survive in the locations in which transport improvements take place. Moreover, entry and exit rates at the ward level could be affected as firms relocate to better benefit from scale effects and agglomeration externalities.

At the aggregate (ward) level, employment and productivity may consequently be affected also through firms entering and exiting the ward. This process may reinforce the scope for scale and agglomeration benefits. For example, aggregate employment could increase because firms input demands increase (scale effects) and because new firms move into the ward. And it could also increase because the existing plants in the ward are employing more workers to take advantage of scale or of the TFP effects.

Transport improvements can have different effects on firms and wards depending on the degree of competition in output markets. The greater the imperfection in competition, the more pro-competitive a transport investment will be and the greater the benefits for economic agents (Vickerman, 2007). In some cases the lack of good transportation could act as an effective trade barrier for less-productive or smaller regions. When transport infrastructure is improved, more-productive or larger regions, which can enjoy economies of scale, might start exporting to the previously isolated location, thus increasing competition. In the short run, this might harm the firms located in the smaller region, although it can benefit consumers because they have access to a wider range of goods and to cheaper final outputs. But in the long run, if only the most productive firms survive, aggregate productivity (and income) of the region would increase.

Hence, reduced transport costs can also increase spatial competition. If only the "fittest" firms survive (as suggested by Melitz, 2003), then the number of firms and aggregate productivity level could change. If we consider transport improvements equivalent to a decrease in transport costs, in the core-periphery new economic geography (NEG) model (Krugman, 1991b; Fujita et al., 2001) the reduction of transport costs can actually give rise to "catastrophic" agglomeration, in which all economic activity concentrates into one region.

In short, given that the theoretical predictions are many, the effect of transport improvements on the firm level and ward level outcomes remains an empirical question.

2.2.2 Effects of transport on labour market outcomes

While most of trade and location theory regards transport investments as having an effect on (goods) transport costs (Michaels, 2008, for example), urban labour theories (Zenou, 2009) consider transport policy as having an effect on commuting costs and therefore mostly operating through the labour market. Glaeser & Kohlhase (2003) document the declining role of goods-transportation costs in developed countries and highlight the increasingly important role of the mobility of workers in modern economies. Transport policy affects labour markets through multiple channels (for a review see Gibbons & Machin, 2006).

If transport improvements reduce commuting costs[8], employers and employees come closer together (spatial competition) and the effective size of the labour markets (scope) can be affected. This has a potential impact on multiple labour market outcomes.

The tightness of the labour market, i.e. the ratio of unemployed relative to the number of job vacancies, is affected by the accessibility and the size of the labour market. If labour markets are more accessible to unemployed workers residing outside the labour market area, or the labour market becomes effectively bigger because it is better connected, for a given number of vacancies the number of potential candidates would increase. This increases competition for jobs and might have two effects. Firstly, unemployed workers living further away from jobs might become employed due to the increase in accessibility. Secondly, due to increased competition, we could see an increase in the quality of job matches which could be translated into higher productivity and wages.

If employers require workers with specific characteristics or skills, employees can bargain to improve their labour market conditions (wages, hours worked, occupation) to compensate for longer commutes. However, if labour markets are thin (Manning, 2006), i.e. workers have access to a limited number of potential employers, longer commutes would not be fully capitalised into nominal wages. Longer commutes can also have an effect on wages through its effect on productivity, if shorter commutes are related to healthier or more motivated workers (van Ommeren & Gutiérrez-i-Puigarnau, 2011).

Related to the these points, the spatial mismatch theory predicts that longer dis-

[8]Commuting costs involve both time and monetary costs. As the transport improvements we study are driven by road construction, we focus on the effects of reduced commuting travel times as a proxy for commuting costs.

tance[9] between residential location and job location can have negative effects on the labour market outcomes of those living further away from employment centres. A reduction in commuting travel times induced by transport investments would thus affect search costs and reservation wages, therefore impacting observed wages and unemployment rates. All these effects might impact labour market outcomes like employment status (employed versus unemployed), unemployment spells, unemployment duration and nominal wages.

Transport policy can also affect labour supply. It can weaken barriers to participation in the labour market, encouraging the entry of disadvantages groups (female or low-skilled) into job search. It can affect hours worked if wages increase due to transport improvements, making work more attractive; or even if increased competition in denser areas induce young professionals to behave in a more rivalrous manner (Rosenthal & Strange, 2008, the "rat race" argument in). Reduced commuting costs could also affect hours worked if the changes in travel times and wages affect the optimal labour-leisure choice.

The emergence of agglomeration externalities might also have positive effects on workers earnings if urbanisation economies[10] and proximity to markets (Krugman, 1991b; Krugman & Venables, 1995) give raise to productivity gains which are capitalised into nominal wages.

Additionally, if better connectivity is regarded as an amenity of locations, improved infrastructure could be capitalised into house prices (some evidence of this is provided in Gibbons & Machin, 2005). In this sense, it would affect real wages, which would change because of changes in nominal wages and in local prices. Net changes in real wages would affect individuals real labour earnings, but if there is capitalisation of accessibility into house values, for home owners there could be a positive effect on their total wealth.

Real earning changes could also affect residential mobility and the commuting versus migrating choice. Higher earners would find it profitable to commute longer if they can afford better quality housing further away from their jobs. In addition, transport investment could affect residential sorting if capitalisation into house rents force low income workers (renters) to move away from areas where transport improvements have taken place.

As in the case of aggregate and firm outcomes, due to the multiplicity of channels through which transport policy and decreased commuting times can affect labour market outcomes, the estimation of robust causal effects is of considerable interest.

[9] Distance includes geographical distance but also bad public transport provision or poor access to private transport (car).

[10] Sharing, matching and learning in Duranton & Puga (2004) terminology; input-output linkages, labour market pooling and knowledge spillovers in the classical Marshall (1890) terminology.

2.3 Empirical evidence

Even if some authors have explicitly included the role of transportation into spatial economic analysis (Combes & Lafourcade, 2001; Puga, 2002; Behrens et al., 2004; Venables, 2007), there is still a need to empirically establish the causal link from transportation infrastructure to spatial economic performance, especially with regard to the effects on individual outcomes.

There exists a substantial amount of empirical evidence on the effects of infrastructure (including transport) on economic outcomes at the macro-level (for a review see Gramlich, 1994; Straub, 2011). This literature has focused on the impacts of investment in roads and public infrastructure on several outcomes, and the authors generally estimate an aggregated Cobb-Douglas production function where infrastructure or roads are treated as a factor of production (García-Milá et al., 1996). The effect of infrastructure has been estimated on outcomes such as aggregate productivity (Aschauer, 1989; Holtz-Eakin, 1994; Fernald, 1999), earnings (Chandra & Thompson, 2000) or employment (Jiwattanakulpaisarn et al., 2009). Some papers have tried to estimate the spillover effects on neighbouring regions to those where the infrastructure investment takes place (Boarnet, 1998; Moreno & López-Bazo, 2007).

A number of recent papers have estimated, using careful identification strategies, the effect of roads on other economic outcomes using data for the USA[11]. Burchfield et al. (2006) examine the effect of early public transport infrastructure and road density on urban sprawl. They find a positive relationship between public transit and density development, but no effect of road density on sprawl.

Using the 1947 planned Interstate Highway System as an exogenous source of variation, Baum-Snow (2007b) studies the effect of highways on the process of suburbanisation of American cities since the 50s, and on the changes in commuting patterns since the 60s (Baum-Snow, 2010). He finds that a sizeable fraction of city center losses in population between 1950 and 1990 can be attributed to highways, and the presence of highways can also explain the increase of commuting flows within suburban areas.

Michaels (2008) uses a similar source of exogenous variation to estimate the effect of reduced trade barriers on the demand for skills. The Interstate Highway System aimed to connect large cities, to serve national defense, and to connect with major routes in Canada and Mexico, but in order to connect these nodes the highways crossed many rural counties. A Heckscher-Ohlin theoretical model predicts changes in the relative demand of skills at the county level when trade barriers are reduced

[11] Turner (2009) provides a review of recent evidence and a discussion of the methodological issues involved in the estimation of the effects of transportation.

through the construction of road infrastructure. He captures the effects highways with an indicator which is activated if a highway crosses a county. He tests the effect of this indicator on on domestic trade (trucking and retail sales), demand of skilled labour (measured as wage-bill of nonproduction workers in manufacturing, relative to production workers) and industrial composition. Rural interstate highways had an important effect on trucking and retail sales, but no effect on changing the industrial composition of the counties. He also finds evidence in favour of the predictions of the model, i.e. that trade increases the relative demand for the abundant factor, so demand for skilled workers increased in skill-abundant counties.

Hymel (2009) uses a cross section of US metropolitan areas to assess the impact of traffic congestion on aggregate employment growth. The measure of congestion used is the annual aggregate amount of time lost due to congested driving conditions. He also uses the Interstate Highway plan and some other political variables as instruments in order to estimate causal effects. His empirical results suggest that increases in congestion significantly reduce subsequent employment growth.

Duranton & Turner (2011a) examine the effect of road construction on vehicle-kilometres traveled (VKT) in US cities. They estimate the effects on a panel of metropolitan statistical areas (MSAs) between 1983 and 2003 and exploit three instrumental variables to be able to infer causality in their estimates. They find that roads cause traffic and that the elasticity is very close to 1. Using a similar strategy the same authors investigate the effects of roads on urban growth (Duranton & Turner, 2011b) and on trade (Duranton et al., 2011). While an increase in the initial stock of highways has a causal positive effect on subsequent population growth, these authors find no effect of highways on the total value of exports.

Other papers (Faber, 2009; Donaldson, 2010) have focused on developing countries (highways in China and railroads in colonial India) to study the effect of the reduction of transport costs, due to transport network development, on trade integration and consequent economic development. These papers also rely on the use of instrumental variables to estimate causal impacts.

On the other hand, only a handful of papers have studied the effect of increased accessibility on firms' outcomes, and they have mostly focused on the analysis of firm relocation (Coughlin & Segev, 2000; Holl, 2004a,b) or firm birth (Holl, 2004c; Melo et al., 2010), all finding positive relationships between the presence of roads and firms' relocation and creation.

Our research is also related to the empirical testing of the effect of agglomeration economies and market-access on firm productivity (for example Ciccone & Hall, 1996; Henderson, 2003; Martin et al., 2011). Holl (2011) studies the relationship between market access and firm productivity when market access changes due to road investments and changes in population. She exploits data for a panel of firms during a period of intense road construction in Spain. When using plant fixed-effects the estimates are imprecise, so she relies on GMM techniques in order to overcome endogeneity problems, and she finds positive significant effects of market access on productivity. Li & Li (2010) use the construction of the Chinese highways system to evaluate the impact of improved transport infrastructure on the amount of inventories held by firms, arguing that the reduced inventories due to road construction improve efficiency and aggregate productivity.

A large fraction of the evidence of the effect of transportation on labour market outcomes has been provided by the empirical testing of the spatial mismatch hypothesis (SMH). The spatial mismatch literature has focused on explaining, both from an empirical (Rogers, 1997) and a theoretical perspective (Gobillon et al., 2007; Gautier & Zenou, 2010), the relationship between the poor access to jobs of disadvantaged groups (like low-skilled, ethnic minorities or females) and their poor labour market outcomes (mainly unemployment or inactivity).

Most of theoretical and the empirical evidence has been developed in the US context. In the last decades many low-skill jobs have been relocated to the city suburbs due to suburbanisation. Poorer individuals and ethnic minorities remained residentially segregated in city centres, due to mobility constraints and cheaper house prices. If these city centres are poorly connected to the suburbs (due to the lack of good public transport links) or if these groups have worse access to private transportation (like cars), then their performance in the labour market is endangered due to poor access to potential vacancies. There exists abundant empirical evidence on the spatial mismatch hypothesis (SMH) but mainly for the US (for a review see Inhanfeldt, 2006). Some recent papers have studied some aspects of the SMH using data for France (Détang-Dessendre & Gaigné, 2009; Gobillon et al., 2011), for the UK (Patacchini & Zenou, 2005) or for Spain (Matas et al., 2010).

Another channel through which transport policy affects labour market outcomes is by changing commuting costs (which can be proxied by commuting time or commuting distance). A reduction in commuting time and costs associated with transport improvements enables people to increase the scale of their job search and could also encourage potential workers to participate in the labour market (Vickerman, 2002; Jiwattanakulpaisarn et al., 2009). There are a few papers that have tested the impact of commuting distance on labour market outcomes. van Ommeren & Gutiérrez-i-Puigarnau (2011) and Gutiérrez-i-Puigarnau & van Ommeren (2010) look at the effect of a reduction in commuting distance induced by a establishment-relocation (within the same firm) on hours worked and absenteeism (productivity). Mulalic et al. (2010) use changes in distance commuted caused by relocation of firms to study the effects of commuting on wages. Search costs can also lead to thin local labour markets which may explain the positive gradients between wages and commuting times (Manning, 2006).

Finally our research is also related to the effect of agglomeration economies and proximity to markets on nominal wages (Combes et al., 2008a; Mion & Naticchioni, 2009; Hering & Poncet, 2010). Workers which are located in denser areas are able to capitalise positive agglomeration externalities on wages. The identification of the effects of agglomeration economies is not straightforward (Combes et al., 2011). Combes et al. (2010) point out two major problems with the identification of the effects of agglomeration on wages. Denser or better connected places might attract more workers and this in turn increases agglomeration (which they refer to as the "endogenous quantity problem"). They suggest the use of instrumental variables to deal with this problem. Also, more able workers might sort into dense areas in order to take advantage of the agglomeration externalities (which they refer to as the "endogenous quality problem"). To deal with this second problem, the authors suggest using individual panel data to control for unobservable individual and locational characteristics. In section 2.6 and in chapter 4, we use a similar strategy, but we focus on the effect on agglomeration on labour market outcomes through improvements on road accessibility.

We believe that our research contributes to the existing evidence in three ways. Firstly, in contrast to many studies mentioned above which use density of roads or connectivity, we measure the effect of transport policy using an index of accessibility to employment. This measure has the advantage of capturing small improvements in the network. It weights the importance of the different locations based on their economic size and on their position in the road network relative to the rest of the country. Even if the location was previously connected to the road network, any improvements in the network can affect the relative "connectivity" of the location, when the improvements reduced the travel time necessary to reach other points of the network. As discussed below (section 2.5.2), this measure is more appropriate in the UK context than other measures used in previous empirical studies.

Secondly, we make use of a careful identification strategy in order to overcome the endogeneity issues concerning the estimation of the effects of accessibility on economic outcomes. A large fraction of the papers summarised above rely on the use of historical, geographical or political instruments to tackle the endogeneity of transport project placements. In contrast, our strategy is based on the use of individual micro-datasets and small spatial scale. We use the variation in accessibility stemming from road improvements and we compare individuals and wards which are located very close to the improvements and which vary in the timing and the intensity of the transport treatment (more details are given in section 2.6).

Finally, we believe that our results are able to shed some light on the channels through which transport policy is affecting individual economic outcomes. For example, most of the evidence on the effect of market access on workers wages, even if causal, cannot separate the impacts attributable to changes in the economic size of the locations and the changes attributable to the "proximity" of these locations. As explained in detail below, we focus on the variation of accessibility (effective density) which stems from road improvements. If there are any potential causal benefits to firms or workers arising from changes in accessibility, these could be induced through transport policy. Therefore, our estimates inform us not only on the overall effect of accessibility on firm and worker economic outcomes but also on the channel through which policy can potentially impact these outcomes.

2.4 Policy context

Due to rising concerns about the increase in traffic flows and the limited capacity of existing networks, in 1997 the Labour Government carried out a reform of the management of major roads. Major roads comprise trunk roads, motorways and principal roads (A roads). The aim of this reform was to "radically change transport policy" (Department of Transport, 1998a). This reform involved the transfer of parts of the English Trunk Network to local authorities, while the Department of Transport kept control of the most strategic roads (those connecting major population centers, ports and airports, key cross-border links and the Trans-European Road Network). The management of these roads was transferred to the Highways Agency in England (which is part of the Department of Transport), to Transport Scotland and to the Welsh Assembly Government. This reform also included "a carefully targeted programme of larger scale improvements", mostly aimed at improving maintenance and focused on these strategic roads (Department of Transport, 1998b).

As stated in the white paper that inspired the 1997 reform (Department of Transport, 1998b), the main aim of the changes introduced by the transport policy reform were to "to create a better, more integrated transport system to tackle the problems of congestion and pollution". In this direction, new objectives were set for the Highways agency, which aimed "to give higher priority to better maintenance and to make better use of existing roads" and "to put greater emphasis to environmental and safety objectives". As of 2009, the main objectives of the Highways Agency are improving the quality and the economic efficiency of the service, reducing traffic congestion, improving road safety and enhancing the environment by mitigating the potentially adverse impact of the strategic road network (Highways Agency, 2009). Traditionally the evaluation and appraisal of transport improvements has been carried out using cost-benefit analysis and has focused on the assessment of the direct impacts on the users of the transport mode (Gibbons & Overman, 2009). In the UK context, appraisals carried out by the Department of Transport have usually estimated the social welfare benefits and costs of a project, relative to a scenario in which the scheme did not exist. In 1998 the New Approach to Appraisal (NATA) was introduced following the reform of transport policy and the main innovation with respect to previous evaluation methods was the quantification of environmental impacts. Transport projects are appraised in a sustainable development framework: all projects must set out their environmental, economic, safety, accessibility and integration effects. Evaluated welfare effects include journey time savings and reliability, environmental effects and other factors (Department of Transport, 2005).

The appraisal of trunk road investment suggests that the reduction of transport congestion benefits the economy through time savings and improvements in reliability (Department of Transport, 1998b). However, there may exist wider benefits that arise in the presence of imperfectly competitive markets or increasing returns to scale (Venables & Gasiorek, 1999) and which are not captured by the 1998 changes to appraisal methodology (NATA). In recent years there have been some attempts to include the "wider-economic" impacts of transport improvement as part of transport appraisal (Department of Transport, 2005). Wider impacts can be incorporated into appraisal through the calculation of three components that capture the effect of agglomeration economies (Graham, 2005, 2006), benefits associated with increasing output (Gibbons & Overman, 2009) and benefits arising from labour market effects (Gibbons & Machin, 2006).

The correct identification and measurement of the impact of road improvements on economic outcomes can provide guidance on the inclusion of wider-benefits into appraisal. The research carried out in chapters 2 to 4 aims to contribute to this debate by providing causal evidence on the economic effects of road improvements and by identifying the channels through which these effects operate.

2.5 Measuring road improvements

2.5.1 The road schemes

We collected information on completed road improvements for the British major roads network by combining information provided by the Department of Transport and other data sources[12]. We collected data on around 75 projects which were completed between 1998 and 2007.

As explained in section 2.4, the improvements aimed to improve traffic flows and road security, and indirectly affect the environment and the economy by reducing traffic congestion[13]. The nature of these projects is diverse. They cover construction

[12]Mainly The Highways Agency, the Motorway Archive, Transport Scotland and Wikipedia.

[13]For example, the objectives of scheme "A6 Clapham Bypass" of table 2.1 were to "improve

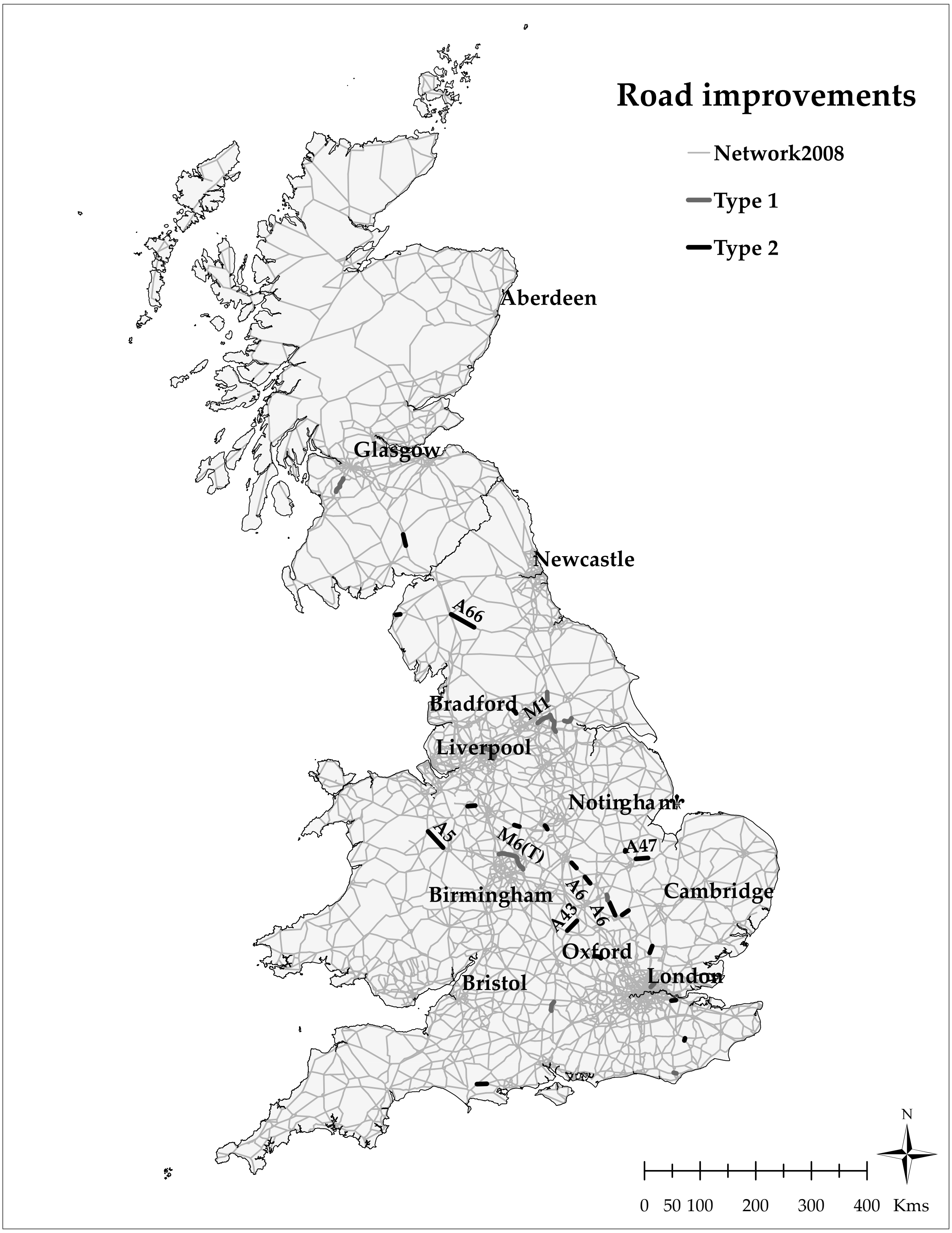


Figure 2.1: Major road network in 2008 and location of road schemes completed between 1998 and 2007

of new junctions, dualling, widening, upgrades and construction of new roads. We focus on construction of new roads and keep 31 road schemes, which are listed in table 2.1. The length of these schemes is also provided in table 2.1.

Opening year	Type	Road	Scheme	Length in kms
1998	Type 2	A16	A16 Market Deeping/Deeping St James Bypass	1.6
1998	Type 1	A34	A34 Newbury Bypass	9.3
1998	Type 2	A50	A50/A564 Stoke - Derby Link	5.1
1999	Type 1	A12	A12 Hackney Wick - M11 Contracts I-IV	4.7
1999	Type 2	A35	A30/A35 Puddleton Bypass	9.3
1999	Type 1	M1	M1/M62 Link Roads	16
1999	Type 2	M74	A74(M). Paddy's Rickle - to St Ann's (J16)	11.6
2000	Type 1	M60	M66 Denton - Middleton	15.3
2002	Type 1	A27	A27 Polegate Bypass	3.2
2002	Type 2	A43	A43 Silverstone Bypass	14.2
2002	Type 2	A6	A6 Clapham Bypass	14.57
2002	Type 2	A66	A66 Stainburn and Great Clifton Bypass	4.1
2003	Type 2	A41	A41 Aston Clinton Bypass	7.3
2003	Type 2	A5	A5 Nesscliffe Bypass	21.48
2003	Type 2	A500	A500 Basford, Hough, Shavington Bypass	7.7
2003	Type 2	A6	A6 Alvaston Improvement	4.7
2003	Type 2	A6	A6 Great Glen Bypass	6.8
2003	Type 2	A6	A6 Rothwell to Desborough Bypass	8.43
2003	Type 1	A6	A6 Rushden and Higham Ferrers Bypass	5.4
2003	Type 2	A650	A650 Bingley Relief Road	4.4
2003	Type 1	M6(T)	M6 Toll. Birmingham Northern Relief Road	29.7
2004	Type 2	A10	A10 Wadesmill to Colliers End Bypass	7
2004	Type 1	A63	A63 Selby Bypass	9.5
2005	Type 1	A1(M)	A1(M) Wetherby to Walshford	8.1
2005	Type 2	A21	A21 Lamberhurst Bypass	2.4
2005	Type 2	A47	A47 Thorney Bypass	10.7
2005	Type 1	M77	M77 Replaces A77 from Glasgow Road	18.25
2006	Type 1	A1(M)	A1(M) Ferrybridge to Hook Moor	19.2
2006	Type 2	A421	A421 Great Barford Bypass	7.6
2007	Type 2	A2	A2 / A282 Dartford Improvement	4.2
2007	Type 2	A66	A66 Temple Sowerby Bypass and Improvements at Winderwath	26.2
			TOTAL	**318.03**

Sources: Own authors calculations using information from the Department for Transport, the Highways Agency, the Motorway Archive, Transport Scotland and Wikipedia.

Table 2.1: Road schemes – 1998-2007

We define two types of projects. Type 2 correspond to roads for which there was an alternative route before, but the road was a minor road (not existing in the

road safety", "relieve congestion" and "provide the opportunity for environmental improvement in Clapham by removing through traffic". See `http://www.highways.gov.uk/roads/projects/6006.aspx` for more information and the evaluation report.

major road network) and an upgrade (which involves improvement and the construction of new lanes) was carried out so the road becomes part of the major road network. They mainly correspond to bypasses which relieve traffic congestion from villages and usually flow in parallel to an existing alternative minor road. Type 1 corresponds to "genuinely" new roads, i.e. roads for which we do not have an alternative minor road flowing in parallel. In practice, as detailed in section 2.5.2.3 below, we defined these two types of improvements in order to be able to calculate travel times given the characteristics of the road network data that we use. They are not substantially different, both involve road construction, except for the treatment we give them in them in the calculation of the travel O–D matrices.

	Length in kms	**Percentage**
MAJOR ROADS NETWORK IN 2008	50,093.73	
All improvements 1998-2007 (type 1 and type 2)	318.03	0.64%
of which:		
Type 1 improvements	138.65	43.6%
Type 2 improvements	179.38	56.4%

The right column reports the percentage of total improvements over the whole network (2^{nd} row) and the percentage of the improvements by type over total improvement length (3^{rd} and 4^{th} rows). Sources: Own authors calculations using information from the Department for Transport, the Highways Agency, the Motorway Archive, Transport Scotland and Wikipedia.

Table 2.2: Length of schemes – 1998-2007

Figure 2.1 displays the location of these projects and the major road network at the end of our period of analysis (2008). Projects are scattered all over Britain. We focus on new construction because these improvements are the ones we can expect to have a substantial effect on travel times between wards. Table 2.2 summarises the length of the improvements by type and displays the percentage of kilometres they represent with respect to the total network length[14]. Total improvements represent 0.64% of the network length, and the majority of them (56.4%) correspond to type 2 improvements.

2.5.2 Accessibility measures

2.5.2.1 Definition

Our aim is to estimate the causal effect of road improvements on economic outcomes. Therefore, the first challenge is to find a measure that captures changes in the road infrastructure. Some measures that have been used in the literature are connectivity to the network (Faber, 2009), kilometres of roads within a given area (Melo

[14] The length of this network slightly differs from the information in table B.3 – 50,250 kilometres –. This is due to the fact that this length corresponds to the GIS network we used in the calculation of the travel times which is just approximate. More details in section 2.5.2.3.

et al., 2010; Duranton & Turner, 2011b), distance to closest highway (Baum-Snow, 2007a), number of rays crossing a given area (Baum-Snow, 2007a, 2010), presence of highways in a given location in a particular year (Chandra & Thompson, 2000; Michaels, 2008), "lowest-cost route effective distance" (Donaldson, 2010) or even amount of public expenditure on road infrastructure (Fernald, 1999).

The current road network in Great Britain is very dense. At the beginning of our period of analysis, 1997, the length of the major road network is of almost 50,000 kilometres (table B.3) and during the period of analysis (1998-2008) the network length increased around 0.65% up to 50,250 kilometres in 2008. For this reason the measures used in the papers mentioned above would not be appropriate for our context. Instead, we use a measure of accessibility to employment "through the road network" (or effective-density) from each location at every point in time.

Accessibility to employment measures the amount of employment which is reachable using the road network from a given location, inversely weighed by the travel time to reach these other locations. One advantage of using this measure is that it is not constrained to artificial geographical boundaries like some of the alternative measures. Moreover, it allows us to use variation due to road improvements which affect optimal travel times, even if the location was previously connected to the network. Additionally, it captures the effects of transport improvements over the whole geography. This measure is, thus, appropriate in our setting where road density was already high at the beginning of our period of study.

Formally, accessibility to employment A_{rt} from a given location r at time t is defined as:

$$A_{rt} = \sum_{j \neq r}^{R} \left[a\left(c_{rjt}\right) * econ\_size_{jt} \right] \tag{2.1}$$

where $a(.)$ is the transport cost function, c_{rjt} are the transport costs between locations r and j at time t and $econ\_size_{jt}$ measures the economic size of the location at time t. In our analysis t corresponds to years 1998 to 2008.

This index is a measure of the economic mass accessible to a firm or a worker in a particular location, given the local transport network. At a given origin location r at time t, employment accessibility A_{rt} is a weighted sum of employment in all destinations j that can be reached from origin r by incurring a transport cost c_{rjt} along some specified route between r and j (for example straight line distance or minimum cost route along a transport network – measured in travel time or in distance). The function $a(.)$ determines how the weights enter in the calculation of A_{rt} (more details below).

Our definition of accessibility is different from other measures of accessibility used in the literature. Other authors, for example Ihlanfeldt & Sjoquist (1990); Raphael (1998); Ihlanfeldt (2002); Détang-Dessendre & Gaigné (2009) or Gobillon et al. (2011) have used indices of accessibility based on mean commuting time of employed workers within some given area, number of available jobs within a given distance band or similar weighted measures to A_{rt} but using different costs functions or distance decays, calculating skill-specific indices, or using employment growth as the measure of economic size of locations.

Instead, we measure the economic size of locations using total employment in a ward and we use travel time between wards as a measure of transport costs (more details on this below). Assuming inverse cost weights[15] and a cost decay equal to one[16], the final expression of accessibility to employment becomes:

$$A_{rt} = \sum_{j \neq r}^{R} \left[\left(1/travel\_time_{rjt}\right) * employment_{jt} \right] \tag{2.2}$$

The accessibility index as defined in (2.2) is identical in structure to market potential measures used in economic geography (e.g. Harris, 1954; Krugman, 1991b), and to the accessibility indices used more generally in the transport literature (e.g. Vickerman et al., 1999; El-Geneidy & Levinson, 2006). This measure is similar to those usually used in empirical tests of the spatial mismatch hypothesis (for example Rogers, 1997) and closely related to the market-access measure used in tests of the effect of agglomeration economies on wages (Mion & Naticchioni, 2009; Hering & Poncet, 2010) or on productivity (Ciccone & Hall, 1996; Martin et al., 2011). Apart from its comparability to these papers, we believe the measure captures most of the theoretical channels explained in section 2.2.

We use 1998 electoral wards (10,500 units) as our geographical unit. The calculation of equation (2.2) requires the construction of an origin–destination (O–D) matrix whose components are travel times between the locations. When computing the O–D matrix we apply a limit of 75 minutes drive time (1.25 hours). This limit facilitates O–D matrix computation but hardly affects the value of the accessibility index because wards beyond 75 minutes have negligible weights in the calculation of A_{rt}. Moreover, as shown in table B.4, more than 99% of commutes in the Great Britain are below 90 minutes.

Accessibility A_{rt} changes in a given origin r are driven both by changes in travel times between wards (stemming from road improvements) and by changes in the

[15]In the definition of A_{rt} the value of the weight $a(.)$ attached to any destination r is a decreasing function of the cost of reaching destination j from origin r. α is the cost decay. Potential weighting schemes include: "cumulative opportunities" weights $a\left(c_{rjt}\right) = 1$ if j is within a specified distance of r, zero otherwise; "exponential weights" $a\left(c_{rjt}\right) = \exp\left(\alpha c_{rjt}\right)$; "logistic weights" $a\left(c_{rjt}\right) = \left[1 + \exp\left(-\alpha c_{rjt}\right)\right]^{-1}$ or "inverse cost weights" $a\left(c_{rjt}\right) = c_{rjt}^{-\alpha}$. See Graham et al. (2009) for further discussion of these indices.

[16]Graham et al. (2009), using the inverse cost weighting scheme, estimate the parameter of distance decay functions for several sectors using similar British data to ours and find values between 1.8 and 1, depending on the sector. In chapters 3 and 4 we check the robustness of our results to different distance decays.

employment of origin ward r and in the different wards j around r. This may lead to endogeneity problems in the estimation of the effect of accessibility, if the employment changes near the origin are causally linked with changes in the economic outcomes in the origin or driven by the same unobserved factors. Moreover, as our focus is on the effect of road improvements, the examination of A_{rt} would be limited as we are not able to differentiate between the changes driven by employment relocation and the changes driven by changes in travel time between locations. It is then useful to construct an alternative accessibility measure $\hat{A}_{rt}$ that focuses on the changes in accessibility that stem only from transport improvements:

$$\hat{A}_{rt} = \sum_{j \neq r}^{R} \left[\left(1/travel\_time_{rjt}\right) * employment_{jt_0} \right] \tag{2.3}$$

where t_0 is some fixed period of time, before our period of analysis starts[17].

Fixing employment to its t_0 level ensures that changes in the accessibility index (2.3) over time occur only as a result of changes in the costs c_{rjt} (e.g. travel time) and not as a results of changes in wards employment. In the empirical work in the next chapters, we instrument A_{rt} with $\hat{A}_{rt}$ in order to only use the variation in accessibility which stems from the road improvements.

In the calculation of A_{rt} the travel time within location r is given the value 0.5 times the minimum travel time to any other ward within the 75 minutes radius. Alternatively, we could exclude location r from the calculation of the accessibility measures to lessen the potential reverse causality problem. This problem arises because our dependent variables include several economic outcomes of the wards or of the individuals located in the ward (firms and workers) and it is likely that these outcomes and the economic size of a given location are be jointly determined. As explained above (and in more detail in section 2.6 and in the following chapters), we use an instrument to address this issue, so the inclusion of the own economic-size in the calculation of A_{rt} should not be an issue. In fact, the qualitative interpretations of the results are robust to the inclusion or exclusion of own economic size.

2.5.2.2 Economic size of the locations

In the calculation of the accessibility indices (2.2) and (2.3) we use total employment in the ward as measure of economic size of the locations. The employment figures at the ward level were obtained using data from the Business Structure Database (BSD). The BSD is a yearly snapshot of the Inter-Departmental Business Register (IDBR), which is accessible to researchers through the Secure Data Service (SDS) delivered by the UK Data Archive (UKDA). We use data from 1997 to 2008.

[17] Due to data availability, t_0 corresponds to 1997 for the results in chapter 3 and to 2001 for the results in chapter 4.

This dataset is maintained by the Office of National Statistics (ONS) and contains a yearly updated register of the universe of businesses in the United Kingdom. It is drawn from administrative registers. It covers about 98% of business activity (by turnover) in Great Britain.

The smallest unit of observation is the establishment or plant (”local unit”), but there is also information on the firm to which the plant belongs (”reporting unit”) and the enterprise and enterprise group of the firm. The dataset provides detailed information on the location (postcode), the sector of production (up to 5 digits) and employment of the plant. This level of detail allows us to calculate employment at any geographical level aggregating up from postcodes. However, individual establishment identifiers are not stable over long periods of time, which makes calculations of entry and exit of plants problematic.

Alternatively we can define indices (2.2) and (2.3) using different measures of economic size. Specifically, we use number of residential addresses in the ward[18], obtained from the National Statistics Postcode Directory (NSPD) and the number of establishments at the ward level (local unit counts), calculated using the BSD. We use these alternative indices in the robustness checks carried out in chapter 4.

2.5.2.3 The calculation of travel times

The accessibility indices A_{rt} and $\hat{A}_{rt}$ are calculated using the ward employment data from BSD and the ward-to-ward travel times in years 1998-2008 using the expressions (2.2) and (2.3). The second component in the accessibility index is then an origin-destination (O–D) matrix containing the costs c_{rjt} (journey time) between each origin and destination (the ward centroid). This matrix is required for all the years of our sample. To calculate travel times we construct GIS-networks for every year between 1998 and 2008. We do this by combining the dataset on road schemes described in section 2.5.1 and two GIS-networks (for years 2003 and 2008) for Great Britain provided by the Department of Transport (DfT).

The 2008 GIS-network contains all the major road links existing at the beginning of 2008. It includes all major roads that, according to the DfT, cover roughly 65% of vehicle kilometres. The network includes information on several characteristics of the road links: the count point code (CP) of each road section (which helps is to identify the links and refers to the point where the traffic is counted), the grid reference for the traffic count point, a unique reference for the local and national transport authorities which manage the link, the road to which the link belongs to (number and type), the maximum permitted speed, the total length of network road link in kilometres and the traffic total flows[19]. Total flow is defined as the “Annual

[18]This is due to the lack of available data on population at the ward level on an annual basis.

[19]More information in *http://www.dft.gov.uk/matrix/estimates.aspx*.

Average Daily Flow" (AADF) and it is measured in terms of number of vehicles. It corresponds to the average over a full year of the number of vehicles passing a point in the road network each day.

We use road construction as the source of variation in travel times over time. We geo-locate all the road links belonging to each of the 31 schemes listed in table 2.1 and we match them to the 2008 road network based on their CP code. Starting from the 2008 network, in every year we remove the new links opened in that year in order to reconstruct the network as it was in years prior to 2008. We construct a network at the beginning of each year of the period 1998-2008. The road network consists of roughly 17,000 road links annually. These networks contain the links existing every year. In order to construct A_{rt} and $\hat{A}_{rt}$ we need to calculate the cost of crossing those links c_{rjt}, which we define based on travel times.

In order to calculate travel times, we use data on traffic speeds from the 2003 generalised primary road GIS-network provided by DfT. Traffic speeds are modeled from traffic flow census data using the Road Capacity and Costs Model (FORGE) component of the National Transport Model (NTM). The National Transport Model provides "a means of comparing the national consequences of alternative national transport policies or widely-applied local transport policies, against a range of background scenarios which take into account the major factors affecting future patterns of travel". It is used to produce forecasts on traffic flows in order to design transport policies[20]. The Road Capacity and Costs Model is one of the three sub-models included in the NTM and it corresponds to the highway supply module.

The Road Capacity and Costs Model (FORGE) is used to show the impact of road schemes and other road-based policies[21]. As explained in the documentation[22]: "The inputs to the Road Capacity and Costs Model are car traffic growth (based on growth in car driver trips) and growth in vehicle-miles from other vehicle types. This traffic growth is applied to a database of base year traffic levels to give future "demand" traffic flows. These are compared to the capacity on each link, and resulting traffic speeds are calculated from speed/flow relationships (which link traffic volumes, road capacity and speed) for each of 19 time periods through a typical week". One of the outputs of FORGE is therefore vehicle speeds by road type, and this is what we use in the calculation of travel times between wards.

We use journey times, obtained from FORGE, in the non-busy direction averaged over all time periods between Monday-Friday 08:00 and 18:00. We focus on non-busy travel directions because the busy travel directions are, in principle, more sensitive to changes in congestion induced by new travel links. In practice, there

[20]See http://www2.dft.gov.uk/pgr/economics/ntm/ for more information.
[21]See http://www.rudi.net/files/FORGE.pdf for more information.
[22]See http://www2.dft.gov.uk/pgr/economics/ntm/pdfntmoverview.pdf for more information

are only minor differences between the modeled journey times in the busy and non-busy directions, or between averages over the working day and full 24-hour periods (see Gibbons et al., 2011, for more details). Due to data limitations, we use journey times in 2003 for the whole period 1998-2008. These speeds are based on traffic flows for 2003, and we keep these constant for all years in the sample.

For links opened after 2003 we use estimated journey times from a regression model using a dataset of over 17,000 links in the 2003 network. We regressed link speeds from the 2003 FORGE network on speed limit dummies, traffic flows, traffic flows squared, road category dummies (six categories) and local authority dummies[23]. The regression predicts speeds from the FORGE reasonably well (R-squared = 0.76). We then used the results to predict travel times for links opened after 2003 for which no FORGE speed is available. To obtain these predictions we use the link characteristics and traffic flows.

We are interested in using the variation in accessibility which stems from road construction, which is how we capture changes transport investments. Therefore in the calculation of predicted traffic speeds we need to avoid using endogenous changes in traffic flows that could be induced by road construction. For the links existing in 2003 we use the FORGE modeled speeds provided by the Department of Transport, which are calculated using 2003 traffic flows. For the links after 2003 we use 2008 flows. These are available for all the links constructed after 2003 because they were provided in the 2008 GIS-network. Alternatively we could use annual traffic flows on the year of opening on the link[24]. However, in order to avoid endogenous traffic flow changes, we prefer to use traffic flows for the same year in all the new links, and the only year in which this is possible is 2008. For some of the links, the prediction exceeded travel time implied by the speed limit. We replaced predicted speed with the speed limit for these links.

Some of the road schemes in the Highways Agency data are bypasses around villages and small towns (type 2 improvements in table 2.1). Typically, before the bypass was opened there was a primary road through the village or town but after the introduction of the bypass the old road was downgraded. The downgrading of the old road implies that it is not present in our 2008 primary road network. Hence, using the method of deleting links based on their opening years would create an artificial break in the primary road network, when it comes to bypasses. Therefore, we keep the bypasses in the network in the pre-opening years and assume that travel time on the bypass before opening year was twice the post opening travel time. Scheme evaluation reports available to us support the assumption of significantly longer travel time through the village/town before the bypass is opened.

[23]The results of this regression are available on request.

[24]which are available in `http://www.dft.gov.uk/matrix/Search.aspx`.

After constructing the generalised traffic networks of each year, we use the network analysis algorithms in ESRI ArcGIS® to compute least-cost (minimum journey time) routes between each origin ward j and destination ward k in years 1998-2008. When computing the O-D matrix we apply a limit of 75 minute drive time. This limit facilitates O-D matrix computation but does not affect the value of accessibility index because wards beyond 75 minutes would have negligible weights in the calculation of A_{rt}.

Table 2.3 contains summary statistics for the travel times between wards for year 1998-2008. The number of O–D crossings is over 9,000,000 in every year. Over the period, the average travel time between two wards decreased 0.55%.

Year	Number of observations	Mean	Standard deviation	1^{st} percentile	10^{th} percentile	90^{th} percentile	99^{th} percentile
1998	9,170,886	49.826	17.534	6.685	24.148	70.958	74.605
1999	9,170,886	49.809	17.524	6.684	24.145	70.922	74.600
2000	9,170,886	49.788	17.518	6.684	24.139	70.905	74.598
2001	9,170,886	49.745	17.515	6.681	24.104	70.872	74.593
2002	9,170,886	49.745	17.515	6.681	24.104	70.872	74.593
2003	9,170,886	49.735	17.512	6.679	24.099	70.859	74.591
2004	9,170,886	49.613	17.476	6.674	24.049	70.724	74.573
2005	9,170,886	49.597	17.471	6.672	24.044	70.704	74.571
2006	9,170,886	49.591	17.469	6.672	24.043	70.696	74.569
2007	9,170,886	49.575	17.463	6.672	24.037	70.675	74.566
2008	9,170,886	49.554	17.457	6.669	24.025	70.645	74.560
Growth 1998-2008		-0.55%	-0.44%	-0.23%	-0.51%	-0.44%	-0.06%

Times in minutes. Min=0 and Max=75 mins. Times measured at the beginning of the year. Sources: Department of Transport and own authors' calculation.

Table 2.3: Summary statistics O–D matrices of travel times – 1998-2008

It should be noted that the network is highly generalised. Journeys via the minor road network are not modeled. Forbidden turns and one way systems are not modeled. All link intersections are treated as junctions. The changes in accessibility must therefore be regarded as approximate. This might have implications in the estimation of the effect of accessibility on economic outcomes if it induces measurement error in the accessibility indices calculation. Hence, our estimates could be too small due to attenuation bias. In this case, we should interpret any effect as a lower bound of the "real" effect.

2.5.2.4 Accessibility changes arising from transport improvements

Accessibility indices A_{rt} and $\hat{A}_{rt}$ can be applied to study the economic effects arising from changing accessibility by road when the costs c_{rjt} in (2.2) and (2.3) are calculated using routing along the transport network. This works because transport improvements change the structure of costs c_{rjt} along the transport network and the structure of costs along routes from r to potential destinations j. This in turn changes the accessibility index.

For example, consider a transport improvement that involves a journey time reduction on a road link between two nodes p and q. This scheme has a first order effect on the costs of the least-cost route between r and j if:

(a) the least-cost route between r and j passes along the link p–q in both the pre and post-improvement periods, such that the transport improvement reduces the cost of the journey along p–q and brings employment at destination j "closer" to origin r in cost terms.

(b) the least-cost route between r and j bypasses link p–q in the pre-improvement period, but switches to use the link p–q in the post-improvement period because of the reduction in costs; again this brings employment at destination j "closer" to origin r in cost terms.

There are also "second order" effects arising when:

(c) the least cost route between r and j bypasses link p–q in the pre-improvement and in the post-improvement periods. However, journeys between other origin and destination pairs have switched to using the link p–q, which reduces congestion on the alternative links in the network used by the routing between r and j; again this brings employment at destination j "closer" to origin r in cost terms.

In the empirical work of chapters 3 and 4 below we rely only on the first order effects of type (a) and (b) arising from new transport infrastructure. We have to ignore second order effects of type (c) because, as explained in the previous section, our road transport network data does not allow us to observe changes in travel time induced by changes in congestion occurring as a result of transport improvements (we have no information on traffic flows observed prior to the improvements).

Changes in cost of all these types imply changes in the accessibility indices (i.e. a change in effective density). The amount of change in the accessibility index at a location j depends on the likelihood that a route between r and j uses the improved link p–q, and on economic mass in j. Our methodology uses the changes in the accessibility index at each location r to estimate the extent to which wards and individuals in location r are "potentially" affected. In this sense, our measure (changes in accessibility) captures the "intention-to-treat".

2.5.3 Descriptive statistics

Table 2.4 summarises the changes in the log of accessibility between 1998 and 2008. In the upper panel the accessibility measure is calculated fixing employment at 1997 levels ($\hat{A}_{rt}$), so that all the variation in accessibility comes from changes in the road network. In the lower panel we calculate the accessibility allowing both employment and road infrastructure to vary (A_{rt}). The middle panel fixes travel times at the beginning of the period (1998) so the variation in accessibility comes from relocation of employment across space. This measure is similar to the standard "market-access" measures calculated using geodesic distance. The table shows the growth in three types of accessibility indices for all the wards (10,540) and for wards which are situated within 10, 20 and 30 kilometres of the road schemes carried out during the period of analysis.

	Wards	**Mean**	**Standard deviation**	**90^{th} percentile**	**Maximum value**	**Proportion of zeroes**
1997 employment and time-varying travel times						
All	10,540	0.470%	1.920%	0.940%	51.110%	31.870%
10 kms	1,389	2.110%	4.520%	5.370%	51.110%	5.030%
20 kms	3,350	1.230%	3.240%	2.800%	51.110%	5.820%
30 kms	4,853	0.950%	2.750%	1.960%	51.110%	5.680%
Time-varying employment and 1998 travel times						
All	10,540	7.170%	7.870%	13.500%	137.070%	0.000%
10 kms	1,389	5.280%	5.150%	10.350%	52.060%	0.000%
20 kms	3,350	5.480%	5.080%	10.470%	67.040%	0.000%
30 kms	4,853	5.690%	5.020%	10.810%	67.040%	0.000%
Time-varying employment and time-varying travel times						
All	10,540	7.640%	8.150%	14.080%	137.070%	0.000%
10 kms	1,389	7.410%	7.570%	14.450%	62.960%	0.070%
20 kms	3,350	6.720%	6.530%	12.540%	67.090%	0.030%
30 kms	4,853	6.650%	6.090%	12.390%	67.090%	0.020%

Sources: Department of Transport, BSD and own authors' calculations

Table 2.4: Summary statistics change log of accessibility between 1998 and 2008

The upper panel of table 2.4 shows that employment accessibility induced by road improvements was on average small, only 0.5% (which is very similar to the decrease in travel times observed in table 2.3 above). However, average accessibility change increases significantly when we focus on wards closer to improvements. For example, within the 10-kilometre distance band the mean change is 2.11% and the 90^{th} percentile is 5.37%. As we expand the sample away from the schemes changes in accessibility tend to fall. Within 20 kilometres, mean accessibility change is 1.23% and the 90^{th} percentile is 2.8%. Within 30 kilometres of the schemes mean accessibility change is 0.95% and the 90^{th} percentile is 1.96%. It is worth noting that the standard variation of the changes is relatively large.

The lower panel of table 2.4 shows that accessibility employment dynamics are a more important driver of variation in effective density than road improvements. This is confirmed if we observe the changes in the middle panel, which focus on the variation on accessibility driven by employment relocation. Nevertheless, variation due to road improvements is non-negligible relative to overall changes.

	Wards	Mean	Standard deviation	90^{th} percentile	Maximum value	Proportion of zeroes
2001 employment and time-varying travel times						
All	10,540	0.300%	1.630%	0.570%	52.370%	42.880%
10 kms	862	2.440%	5.140%	5.930%	52.370%	9.860%
20 kms	2,453	1.120%	3.230%	2.400%	52.370%	11.290%
30 kms	3,888	0.790%	2.620%	1.650%	52.370%	10.650%
Time-varying employment and 2002 travel times						
All	10,540	13.430%	5.270%	18.330%	80.230%	0.000%
10 kms	862	12.370%	3.800%	15.610%	32.180%	0.000%
20 kms	2,453	12.570%	4.090%	15.940%	56.830%	0.000%
30 kms	3,888	12.570%	3.810%	16.000%	56.830%	0.000%
Time-varying employment and time-varying travel times						
All	10,540	13.740%	5.470%	18.700%	80.230%	0.000%
10 kms	862	0.420%	6.450%	19.850%	72.910%	0.000%
20 kms	2,453	13.710%	5.220%	17.590%	72.910%	0.000%
30 kms	3,888	13.370%	4.630%	17.120%	72.910%	0.000%

Sources: Department of Transport, BSD and own authors' calculations

Table 2.5: Summary statistics change log of accessibility between 2002 and 2008

Due to limitations in the availability of some of the variables of the survey used in the obtention of the empirical results of chapter 4 we exploit data from 2002 to 2008. For this reason, we also computed indices (2.2) and (2.3) using 2001 as the base year for the calculation of $\hat{A}_{rt}$. The summary statistics for the change in log of accessibility between 2002 and 2008 are displayed in table 2.5. They are very similar to those of table 2.4: overall change in accessibility stemming from road improvements (top panel) is relatively small (0.3%), but it increases substantially when we restrict our sample to wards which are located close to improvements carried out between 2002 and 2008. As before, most of the variation in changes in overall accessibility (bottom panel) stems from employment changes across wards.

Tables 2.4 and 2.5 illustrate the magnitude of the changes in accessibility. In fact, we use annual variation in accessibility in the estimation of the empirical results in the following chapters. Tables B.5 to B.8 in the appendix provide summary statistics on the level of log accessibility and on the average within-ward variation in log accessibility for 1998-2008 and the sub-period 2002-2008. Tables B.7 and B.8 are specially useful because they show the variation that we exploit in the fixed-effects regression analysis undertaken in the following chapters.

Figures 2.2 and 2.3 illustrates the spatial relationship between road schemes and resulting accessibility increases. Figure 2.2 shows the changes in log accessibility between 1998 and 2008 using annual employment A_{rt}, while figure 2.3 shows the

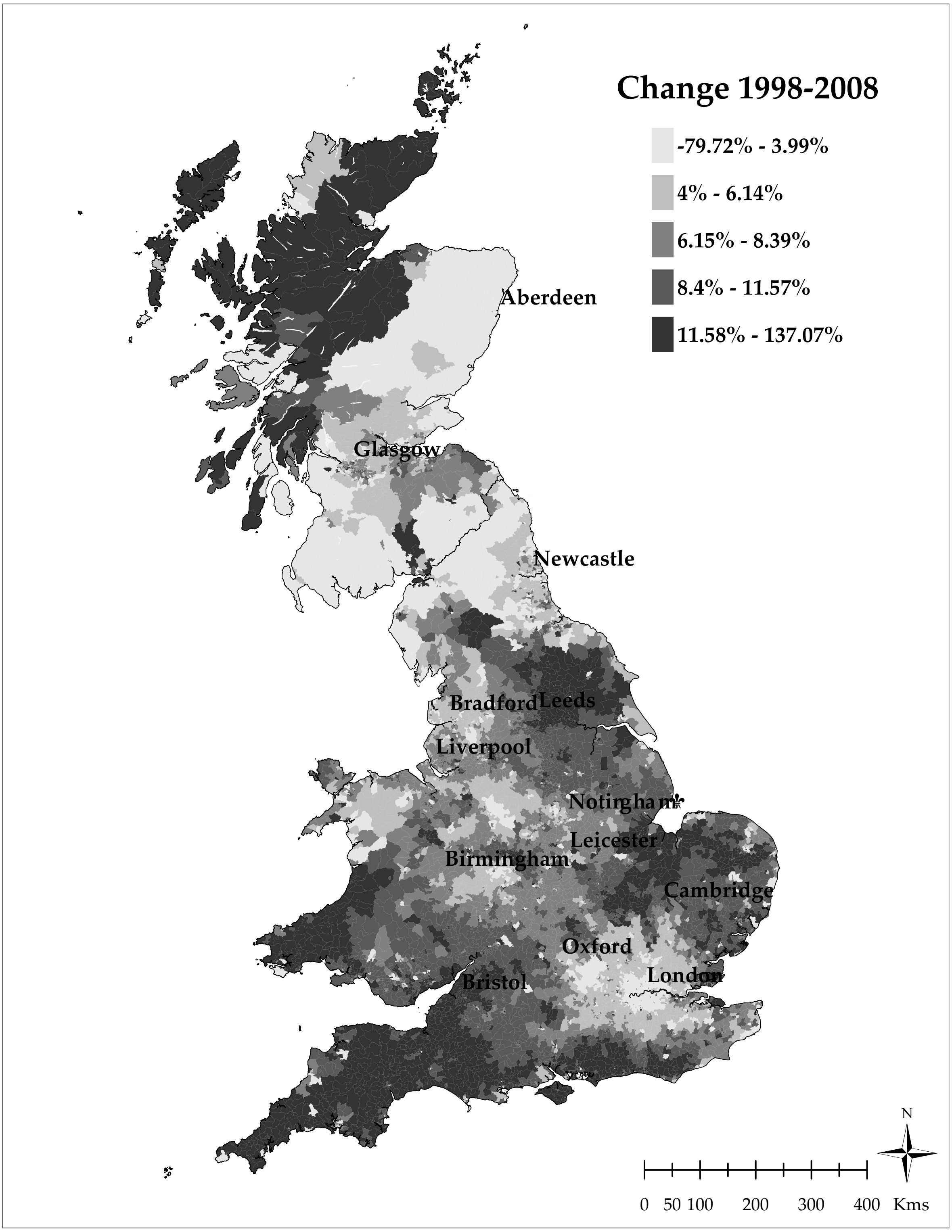


Figure 2.2: Changes in log accessibility (annual employment) – 1998-2008

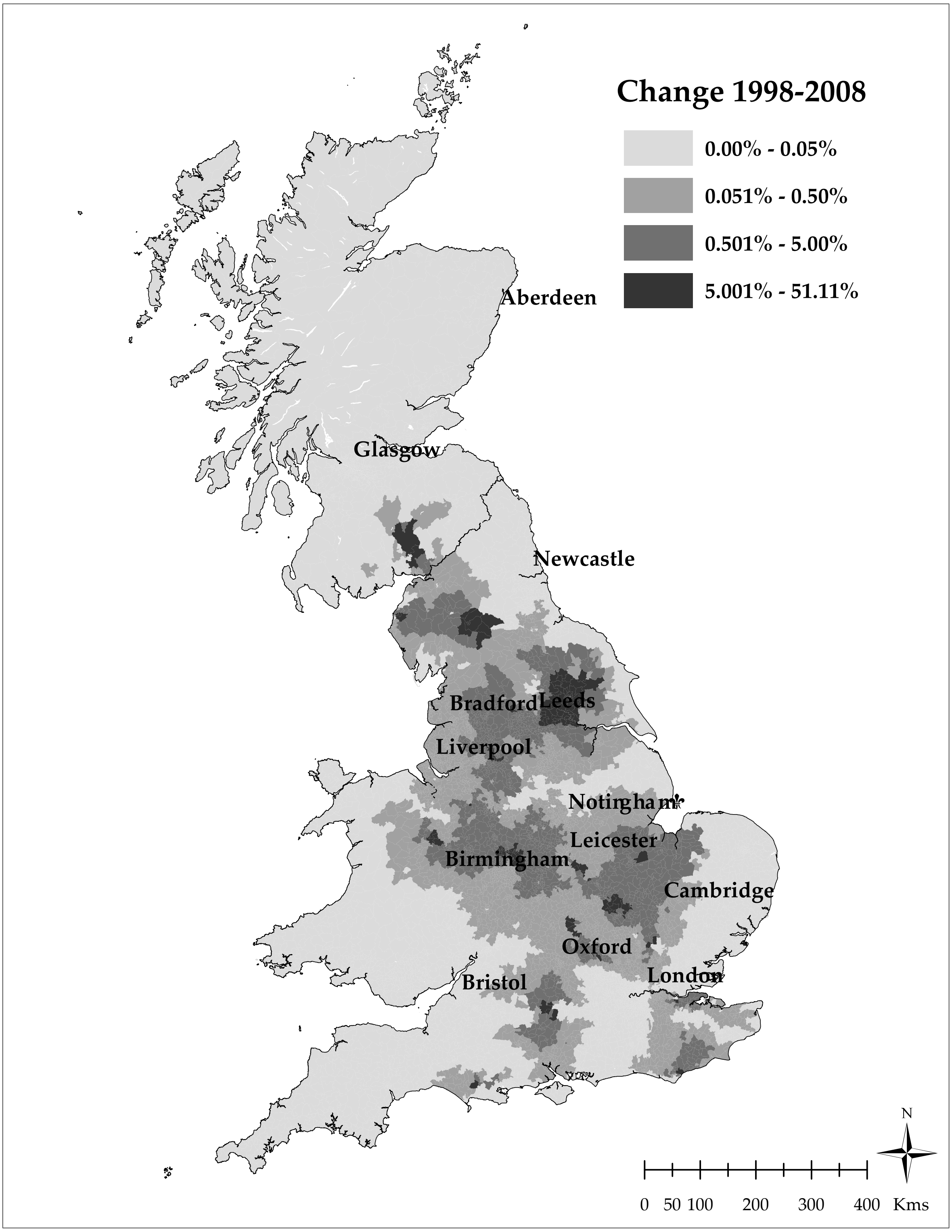


Figure 2.3: Changes in log accessibility (1997 employment) – 1998-2008

changes in log accessibility which stem only from road improvements ($\hat{A}_{rt}$). The biggest changes in accessibility are around the schemes plotted in figure 2.1, but there is substantial spatial variation across the country.

The amount of spatial variation is more evident in figure 2.4. It focuses on the Manchester-Leeds area and allows us to illustrate the identification strategy (which is explained below). The thin light white lines show the primary road network in 2008. New links and significant improvements between 1998 and 2007 are indicated by bold lines, with the year of opening and the name of the road labeled. The dark grey lines are ward boundaries. The map illustrates that the effects of road improvements on accessibility vary considerably across wards in the vicinity of the same improvement. We argue that these differences in accessibility changes across wards are coincidental and can be treated as exogenous, especially when controlling for differential time trends near different schemes[25]. This relates to the identification strategy used to tackle the endogeneity of transport investment placements explained below.

2.6 Endogeneity issues and identification

In the following two chapters we estimate the impact of accessibility on several ward-level and individual economic outcomes. As our aim is to be able to estimate causal relationships, a correct identification of the estimated parameters of the accessibility effects is crucial. In this section we illustrate the general identification strategy followed in chapters 3 and 4.

For example for the aggregate economic outcomes we estimate the specification:

$$y_{rt} = \beta A_{rt} + \mu_r + \tau_t + \varepsilon_{rt} \tag{2.4}$$

where y_{rt} is the economic outcome of ward r at time t, A_{rt} is the log of the accessibility to employment of ward r at time t, τ_t capture common shocks affecting accessibility and the economic outcome of study of all wards in a given year and ε_{rt} is the idiosyncratic error term.

For the correct identification of parameter β, there are three endogeneity concerns which may challenge the validity of the estimates. Firstly, cross-sectional estimates of the effect of accessibility on economic outcomes could be biased if the model does not capture underlying factors (such as place specific productive advantages) that affect both effective density and economic outcomes. We use a fixed-effects estimation method to address this problem (by including μ_r in the aggregate outcomes estimations, and individual firm or worker fixed-effect in the indi-

[25]For completeness, the same maps but using the changes between 2002 and 2008 are provided in the appendix (figures B.1, B.2 and B.3)

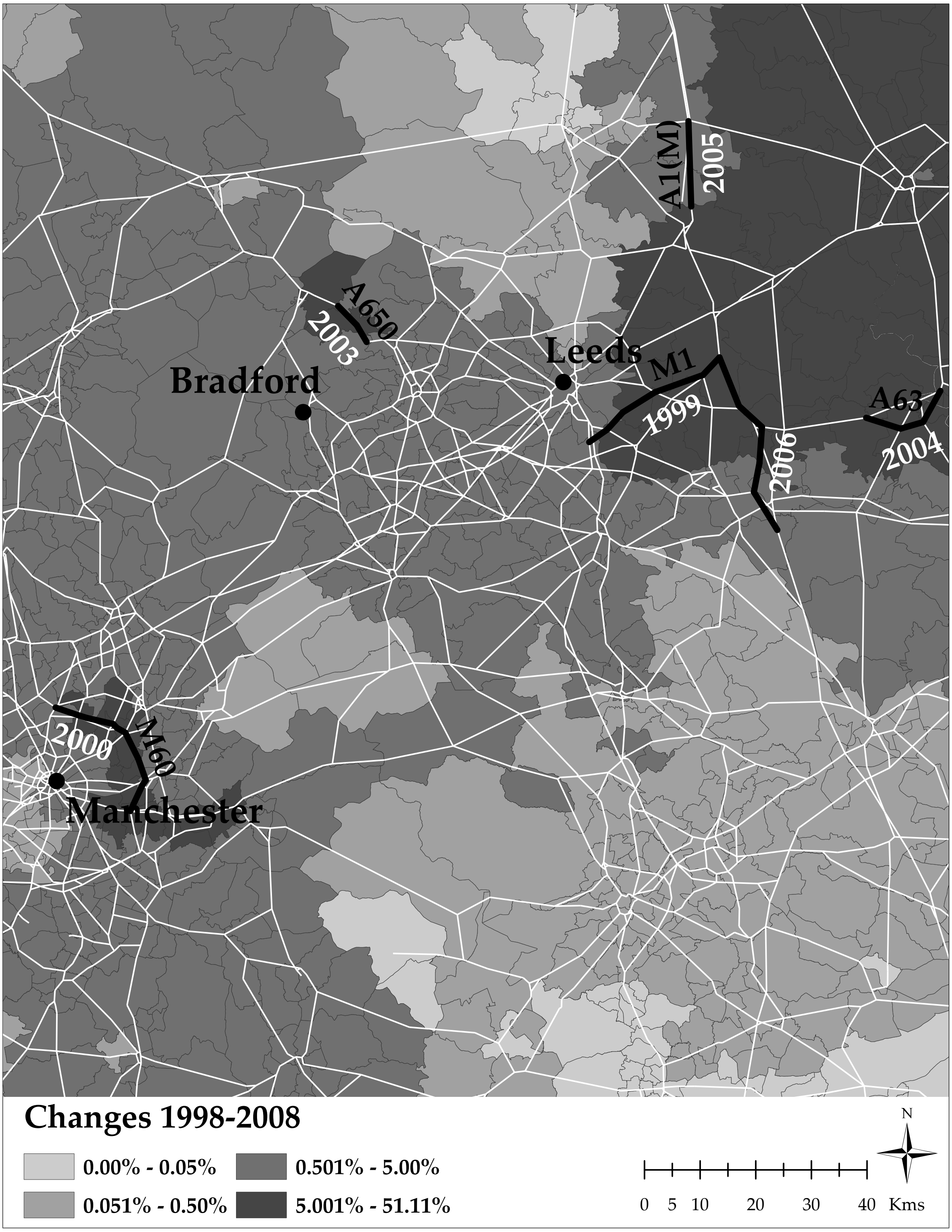


Figure 2.4: Changes in log accessibility (1997 employment) – 1998-2008

vidual outcomes estimations). When we use fixed-effects we exploit the within ward/individual variation over time. However, in the fixed effects framework, changes in accessibility can arise because of road improvements and because employment relocates.

The second endogeneity issue is that accessibility changes due to relocation of employment may be partly driven by the outcome variable studied or be correlated with the same unobserved shocks. As discussed above, to address this source of bias, we construct an instrument which uses accessibility changes stemming only from the transport improvements. In practice, we instrument A_{rt} with $\hat{A}_{rt}$.

Finally, in order to reduce the possible bias caused by the endogeneity of the placement of the transport investments, we focus on observations which are located in wards within 10, 20 or 30 kilometres of road schemes. This way, we compare firms and locations which are close to the improvements and we exploit the fact that the impact of the improvements varies considerably even within the distance band. As illustrated by figures 2.4 and B.3 and explained above, there is a substantial amount of variation in the changes of accessibility around the improvements. It is quite unlikely that the improvements are aimed at specific individuals or wards within those narrowly defined distance bands, specially after controlling for different growth trends around the schemes.

2.7 Conclusions

In this chapter we have set out the reasons why there is a need to provide causal estimates of the effects of roads on economic outcomes in the UK context. Transport infrastructure projects require large amounts of funding so the correct identification of their impacts is key. Traditional evaluation of transport projects does not fully account for the potential economic benefits of improvements. Moreover, given that the theoretical predictions of the impacts of transport on economic outcomes are ambiguous, robust empirical evidence is necessary to learn about the channels through which transport affects workers and firms.

We focus on road improvements and use a measure of accessibility to employment to capture the effects that road construction has on economic outcomes. In this chapter we have explained in detail why we consider this measure appropriate for the UK context and have provided details of its construction. The use of micro-data and small geography allows us to implement a careful identification strategy which helps overcome a number of endogeneity issues. In the following chapter we estimate the effect that changes in accessibility has on aggregate economic outcomes (at ward level) and on individual outcomes, both for firms (chapter 3) and for workers (chapter 4).

Chapter 3

Economic Impacts of Transport Policy: Aggregate and firm outcomes

3.1 Introduction

In this chapter we present and analyse the results of the impact of accessibility on aggregate outcomes and on firms. As explained in chapter 2, transport improvements can affect aggregate and firm outcomes through several channels. A reduction of transport costs due to road improvements can affect the demand for inputs and outputs that firms face, and, depending on the price response, the input mix and output might be affected.

If workers bargain down their wages due to the increased competition in the labour markets, firms might take advantage of the improved accessibility by employing more workers. If firms benefit from economies of scale, if the amount of inputs firms employ changes, labour productivity might be affected as output responds. Alternatively, firms could experience positive externalities from increased accessibility in the form of agglomeration economies, which could have an impact on total factor productivity (Gibbons & Overman, 2009).

At the aggregate level, ward employment could be affected if firms change the amount of labour they employ or if the number of plants in the ward changes. Transport improvements can affect the local number of plants in two ways. Better connectivity can improve firms productivity allowing the plants to survive when in other conditions they would have existed the market. Conversely, if spatial competition increases some firms might exit the market. If the firms that stay in the market are the ones which have higher productivity (Melitz, 2003), aggregate productivity might increase.

In this chapter we test these predictions using individual firm micro-data and aggregate data at a very small geographical scale (wards). We use British data during the period 1998-2008. We capture the effect of road improvements using a measure of accessibility to employment (effective density), as explained in detail in the previous chapter. Our main approach is to estimate aggregate effects of accessibility for small spatial units (wards) taking into account firm exit, entry and geographical relocation. In addition, we estimate firm level effects for existing firms exposed to different changes in accessibility. We take advantage of the rich datasets and the small geographical scale to tackle the endogeneity issues which challenge the causal interpretation of our results.

In section 3.2 we describe the data used in the empirical exercises. In section 3.3 we discuss the results for the local economic outcomes. We present results on ward employment and number of plants (section 3.3.2) and for ward productivity (section 3.3.3). In section 3.4 we present the results on the individual firms outcomes, focusing on employment (section 3.4.2) and productivity (section 3.4.3). Finally, in section 3.5 we conclude.

3.2 Data

To calculate aggregate employment and the number of establishments within each ward we use the Business Structure Database (BSD) for the period of analysis. These aggregates are used as dependent variables in the estimation of the ward-level employment effects discussed in 3.3.1, and aggregated employment is used to calculate the accessibility measures as described in the previous chapter. Details on this dataset are also given in the previous chapter.

For the productivity regressions, we use the Annual Respondents Database (ARD). The ARD holds responses to the Annual Business Inquiry (ABI). The ABI is a stratified random sample, extracted from the BSD. The ABI is a comprehensive business survey covering balance-sheet information like gross value added, gross output, wages, intermediate inputs, employment, industry, and investment. We use this information and the EU KLEMS Deflators (base 1995) to express the firm balance-sheet data in real terms. Although the ARD only contains a sample of small businesses, being a census of large businesses it contains information on firms which cover a large fraction of employment (for example 90% of UK manufacturing employment).

We combine the balance-sheet data from ARD with capital stock data in order to study firm-level productivity effects. The capital stock variable is built from the gross investment flows using a perpetual inventory method and allowing for differential depreciation rates across the three main asset classes (equipment, structures and vehicles), available at the ONS-Business Data Linkage laboratory and constructed by the Centre for Research into Business Activity (CeRIBA). Capital stock is available only until 2004, which limits the sample period in the productivity regressions to 1998–2004. For the plant level employment regressions we do not need data on capital and use a longer panel of the ARD which also covers more plants. We use data of plant employment between 1998 and 2008.

As noted by Criscuolo et al. (2003), a number of issues arise when deciding the level of aggregation at which to work. ARD reports information for both "local units" (LU) and "reporting units" (RU). Balanced-sheet data is available at the RU level, while location and employment is available at the LU level. Questions related to employment can be investigated at LU level, since reporting units with several plants report on several local units that may be located in different wards.

The correct aggregation unit for productivity analysis is more difficult. Productivity and technology of production might vary across local units, across reporting units or indeed within local or reporting units. Hence, strong assumptions would be needed in order to calculate TFP at the plant level (we would need to apportion the RU balance-sheet information across the LU which belong to it, for example based on their share of employment). This may be problematic, and hence, our main regressions use a sample of single plant firms for which the firm-level and the plant-level coincide. We test the robustness of the findings to the inclusion of multi-plant firms in the sample.

To construct the ward level background characteristics used as control variables we use information from the Census 2001 provided by CASWEB[1]. We calculate the share of population aged 15-64 with higher education, mean age of population, share of population living on social housing and the rate of unemployment. We also use straight line distance to the nearest improvement (undertaken at any point in time during our period of study) calculated using GIS and the dataset of transport improvements described in chapter 2.

3.3 Effects on aggregate outcomes

We use regression analyses to estimate the average effect of accessibility by road on the scale of economic activity and productivity at the ward level. We have detailed geographical information on the location of the plants (postcodes). A UK postcode unit corresponds to a limited number of addresses (around 14) or a single large delivery point.

We aggregate the data up to electoral ward level, which leaves us with around 10,500 geographical units. A ward in the Great Britain is an electoral district at subnational level and it is the primary unit of British administrative and electoral geography. In England, the London boroughs, the metropolitan boroughs and the non-

[1]CASWEB is provided by the Census Dissemination Unit (CDU), which is based within MIMAS at the University of Manchester. See http://cdu.mimas.ac.uk/index.htm for more information.

metropolitan districts (including most unitary authorities) are divided into wards for local elections. The same applies to Welsh communities and to Scotland. Wards include roughly the same number of electors. We use the ward as defined in 1998, the first year of our study.

This unit is small, especially in dense areas. For example, the City of London (which is a single local authority) contains approximately 25 wards. The advantage of using wards as the geographical units is that they are very small spatial units, which allows us to identify phenomena that would be unobservable at a higher geographical level. The detailed spatial scale is also crucial for our identification strategy that uses spatial variation in the accessibility increases in the vicinity of road schemes.

For most of the results we use data for 11 years, 1998 to 2008. For these years we have the relevant firm data and can construct the accessibility measures. For total factor productivity regressions we use 7 years of data (1998–2004) due to the lack of data on capital stock in later years.

3.3.1 Empirical specification and identification strategy

We estimate the effect of accessibility to employment on three aggregated outcomes: ward employment, ward number of plants and ward total factor productivity. For the two first outcome variables we use a standard panel data fixed effects approach. For total factor productivity we use a two-step approach in which we first estimate yearly ward level productivity shifters and then regress them on accessibility in the second step. This is discussed in details in this section.

3.3.1.1 Employment and number of plants

Initially, the empirical model for employment and the number of plants is:

$$y_{rt} = \beta A_{rt} + \tau_t + \varepsilon_{rt} \tag{3.1}$$

where y_{rt} is the economic outcome of ward r at time t (log of ward employment or log of number of plants), A_{rt} is the log of the accessibility to employment along the road network (as defined in the previous chapter) of ward r at time t. Year fixed effects τ_t capture general changes that influence all locations in the study area in a given year (e.g. macro shocks). Finally, ε_{irt} is the idiosyncratic error term.

We are interested in the correct estimation of parameter β, which is the elasticity of employment/number of plants with respect to accessibility. We use OLS as our first specification. Traditional estimates of the effects of accessibility are often based on ordinary least squares (OLS) estimates of models like (3.1).

The OLS estimates are biased if unobserved area effects are correlated with accessibility – if for example, and as seems likely, better transport connections and higher employment density have evolved in places with productive advantages. To account for this we include ward fixed effects μ_r in specification (3.1), which capture the time-invariant heterogeneity of wards:

$$y_{rt} = \beta A_{rt} + \mu_r + \tau_t + \varepsilon_{rt} \tag{3.2}$$

A first step to eliminating these biases is to eliminate fixed-over-time ward effects μ_r by time-demeaning the data within wards (so called within transformation). The within transformation is obtained by first averaging the equation (3.2) over time and then subtracting the ward averages from equation (3.2):

$$(y_{rt} - \bar{y}_{r.}) = \beta \left(A_{rt} - \bar{A}_{r.} \right) + \tau_t + \xi_{rt} \tag{3.3}$$

where $\bar{y}_{r.}$ and $\bar{A}_{r.}$ are ward averages of the outcome and of log accessibility over the time period (1998-2008), τ_t are year dummies which capture shocks affecting the variation over the ward means and $\xi_{rt} = (\varepsilon_{rt} - \bar{\varepsilon}_{r.})$. This way the ward fixed effects disappear and the estimates are robust to time-invariant ward heterogeneity that can be arbitrarily correlated with accessibility. This formulation is a starting point for evaluating the effects of transport policy on firms, because transport improvements generate changes in A_{rt} over time.

We prefer the within transformation to first differencing because first differencing assumes instantaneous responses to accessibility changes while within estimation allows for a more flexible time pattern for the effect. For example, if we have ten years of data and accessibility in a ward changes in the fifth year, the within estimator will be based on the comparison of the value of y in years 1-5 to year 6-10 whereas the first differenced estimator compares changes between years 5 and 6 to changes in other years. If the response to the accessibility change take longer than a year, the first differenced estimates will be biased downwards. The within estimator is better suited for capturing slow and gradual changes.

In general, the variation in accessibility over time within wards $(A_{rt} - \bar{A}_{r.})$ could come through changes in the spatial distribution of employment, or because of changes in the transport network. Changes in accessibility due to the relocation of employment across space may be directly affected by the outcome variable or correlated with unobserved shocks in the error term of (3.3), which may lead to bias in the estimation.

To address the issue of endogenous determination of accessibility, we instrument accessibility index A_{rt} (that varies within wards due to changes in both employment and transport network) with the measure, denoted by $\hat{A}_{rt}$, which only picks up changes in the transport network. We calculate the accessibility based on the pre-improvement spatial distribution of employment (year 1997). Details on the construction of these measures are given in chapter 2. We estimate (3.3) by two-stage least squares using this as an instrument for actual changes in accessibility.

The instrumental variable (IV) estimates from (3.3) could produce biased estimates of the aggregate effects of transport improvements, if areas with increasing or declining employment trends are those that experience the greatest accessibility changes due to road improvements. This implies that ξ_{rt} in equation (3.3) is correlated with the instrument. The usual reason to suspect this kind of problem is the possibility that transport policy is endogenous to the employment and productivity trends in the targeted locations, i.e. the decision to improve the transport network might be partly driven by productivity trends.

We address this potential source of bias by focusing the empirical analysis on places and firms that are close to the transport improvement sites. In the results section below we present estimates for samples within 20 kilometres of the sites of improvement (10 kilometres and 30 kilometres used in robustness checks). In this way we are comparing closely neighbouring places that differ incrementally in terms of the changes in accessibility they experience as a result of the road network improvements. We assume that these differences in changes in accessibility close to transport schemes are an incidental by-product of the scheme rather than its intended outcome. The main changes in mean travel times and employment accessibility occur close to the end points of new road schemes, although they are typically intended to improve the flow of traffic between cities or areas further away from the improvement. There are also often long delays between commissioning and opening of road schemes, which would weaken any link between pre-existing local employment or productivity trends and the decisions over where to site these projects.

In order to ensure that the instrument is uncorrelated with the underlying trends, we further control for differential trends in the vicinity of road schemes in our final specification. This is done by including a set of scheme dummies (31 schemes) interacted with year in equation (3.3). Lastly, we test for the robustness of the results to the inclusion of salient ward characteristics (straight line distance to closest road scheme, employment rate, average age, proportion of population aged 16-74 with higher education and proportion of population living on social housing). The area characteristics are measured in one year and do not change over time in our data but we interact them with time, which implies that the trend is allowed to differ by ward characteristics. We acknowledge that this robustness check is very demanding and may pick up some of the effect of road improvements, especially if the impact of road schemes is gradual in nature.

Note, that this estimation strategy ignores whether or not the specific firms or their employees or customers in fact make any use of the road network that has been put in place. The effects that are estimated are thus analogous to "intention to treat" estimates in the programme evaluation literature, and are the expected productivity changes for firms or areas exposed to the treatment (change in employment accessibility by road).

3.3.1.2 Aggregate productivity

The above strategy is used for employment and number of plants. When estimating the effect of accessibility on total factor productivity at the ward level we use a two-step procedure. The first step uses firm level micro data to estimate a production function model of the form:

$$y_{irt} = \alpha K_{it} + \delta L_{it} + \rho M_{it} + \gamma_{rt} + \nu_{irt} \tag{3.4}$$

The dependent variable y_{irt} is the log of gross output of plant i in ward t in year t, K_{it} denotes the log of the capital stock of the plant, L_{it} is the log employment of the plant, M_{it} is the log of intermediate inputs used, γ_{rt} are ward-by-year fixed effects to be estimated and ν_{irt} is the error term[2].

In the second step, we regress the estimated ward-by-year effects (averaged over all firms in a given ward in a given year) on the accessibility variable. Again, our first specification is a simple OLS regression. We estimate the equation:

$$\hat{\gamma}_{rt} = \beta A_{rt} + \tau_t + \varepsilon_{rt} \tag{3.5}$$

where $\hat{\gamma}_{rt}$ are the estimated ward-by-year TFP residual, A_{rt} is accessibility of ward r at time t, τ_t are year dummies and ε_{rt} is the error term.

As before, we add ward fixed-effects μ_r to control for unobservable time-invariant characteristics correlated with accessibility and ward productivity. Equation 3.5 becomes:

$$\hat{\gamma}_{rt} = \beta A_{rt} + \mu_r + \tau_t + \varepsilon_{rt} \tag{3.6}$$

Next, we use time demeaning to eliminate fixed-over-time area effects as in (3.3), i.e.

[2]The estimation of production functions like (3.4) can be affected by simultaneity bias. This problem would arise if the firm choice of inputs depends on unobserved productivity shocks, which are correlated with the inputs and the output at the same time. This would cause the input quantities K_{it}, L_{it} and M_{it} to be correlated with the error term ν_{irt}. Authors like Olley & Pakes (1996) and Levinsohn & Petrin (2003) have proposed methodologies to deal with this issue. We experimented with these methods, but the estimated α, δ and ρ changed very little so the results in the second step remained unchanged. Moreover, different methods tend to produce results where the correlation of TFP estimated by different methods is usually high (see **?**).

we estimate:

$$(\hat{\gamma}_{rt} - \bar{\hat{\gamma}}_{.t}) = \beta\,(A_{rt} - \bar{A}_{r.}) + \tau_t + \xi_{rt} \tag{3.7}$$

using a panel of wards between 1998 and 2008.

In order to address the potential endogeneity of effective density in (3.7), we use the IV strategy outlined above. Targeting of transport improvements is addressed by limiting the estimating sample to wards located within 20 kilometres of road schemes and controlling for differential trends around schemes.

The advantage of the two-step procedure compared with a regression with ward level aggregates is that, in the two-step method, the ward-year fixed effects γ_{rt} control for a wide range of time-varying ward specific factors and the coefficients on capital, labour and intermediate inputs will be less biased than in a ward level aggregate regression. As a result, also the coefficient on accessibility will be more reliable in the two-step method.

3.3.2 Results: employment and number of plants

The first regression results, presented in tables 3.1 and 3.2, are estimates of the effect of log accessibility to employment on ward log of employment and log of the number of plants. For these regressions we use the BSD data. Standard errors are clustered at the ward level to allow for arbitrary correlation across time.

The tables display the results using data on wards within 20 kilometres of road schemes and contain four different model specifications. The tables show the coefficient on employment accessibility, its standard error, the number of observations used, and the Kleibergen-Paap F-stat (first stage) for the IV specifications. Table B.9 shows descriptive statistics of the main variables used within the 10-20-30 kilometre distance bands (descriptives for the accessibility indices and for ward employment and number of plants).

In table 3.1, the first panel uses total employment (all sectors combined) and lower panels show results by five broad sectors (manufacturing, construction, consumer services, producer services and other)[3]. The first specification is a simple OLS regression which neglects all endogeneity issues. The OLS coefficients are positive and significant across the board indicating that wards with higher employment tend to have better accessibility. In the second column, we add ward fixed effects

[3] We use the 1992 Standard Industrial Classification (SIC) at 2 digits to define the 5 wide industrial categories. Manufacturing includes sector codes 15 to 37; construction includes sector codes 40, 41 and 45; consumer services includes sector codes 50 to 59; producer services includes sector codes 65 to 74. Other includes the rest, including primary activity, public sector, transport and other sectors. More information can be found in `http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/uk-standard-industrial-classification-1992--sic92-/index.html`.

	(1)	(2)	(3)	(4)
ALL SECTORS				
Coefficient	0.426***	0.260***	0.155	0.195*
Std. error	0.021	0.06	0.106	0.106
No. observations	36740	36740	36740	36740
Kleibergen-Paap F-stat			8732	13260
MANUFACTURING				
Coefficient	0.247***	0.635***	0.791***	0.473
Std. error	0.036	0.185	0.292	0.306
No. observations	36136	36136	36135	36135
Kleibergen-Paap F-stat			9004	14387
CONSTRUCTION				
Coefficient	0.217***	0.297*	0.541**	0.518**
Std. error	0.023	0.173	0.243	0.235
No. observations	36684	36684	36684	36684
Kleibergen-Paap F-stat			8789	13277
CONSUMER SERVICES				
Coefficient	0.435***	0.149*	0.217	0.076
Std. error	0.024	0.087	0.178	0.185
No. observations	36739	36739	36739	36739
Kleibergen-Paap F-stat			8732	13260
PRODUCER SERVICES				
Coefficient	0.734***	0.586***	0.979***	0.618**
Std. error	0.028	0.124	0.265	0.259
No. observations	36719	36719	36719	36719
Kleibergen-Paap F-stat			8729	13251
OTHER				
Coefficient	0.347***	0.126	0.065	0.13
Std. error	0.02	0.089	0.14	0.142
No. observations	36740	36740	36740	36740
Kleibergen-Paap F-stat			8732	13260
Distance band	20 kms	20 kms	20 kms	20 kms
Year FE	YES	YES	YES	YES
Ward FE		YES	YES	YES
IV			YES	YES
Scheme trends				YES

Sources: BSD-ONS via SDS. Standard errors clustered at ward level. * p<0.1, ** p<0.05, *** p<0.01

Table 3.1: Ward level employment results

	(1)	(2)	(3)	(4)
ALL SECTORS				
Coefficient	0.338***	0.105***	0.190***	0.135***
Std. error	0.016	0.03	0.055	0.05
No. observations	36762	36762	36762	36762
Kleibergen-Paap F-stat			8733	13256
MANUFACTURING				
Coefficient	0.323***	0.323***	0.522***	0.185
Std. error	0.02	0.101	0.161	0.169
No. observations	36167	36167	36166	36166
Kleibergen-Paap F-stat			9003	14381
CONSTRUCTION				
Coefficient	0.151***	0.220***	0.344***	0.226*
Std. error	0.017	0.074	0.119	0.117
No. observations	36709	36709	36709	36709
Kleibergen-Paap F-stat			8788	13271
CONSUMER SERVICES				
Coefficient	0.375***	0.099**	0.104	-0.012
Std. error	0.018	0.049	0.084	0.078
No. observations	36761	36761	36761	36761
Kleibergen-Paap F-stat			8733	13256
PRODUCER SERVICES				
Coefficient	0.582***	0.269***	0.732***	0.406***
Std. error	0.02	0.08	0.149	0.141
No. observations	36742	36742	36742	36742
Kleibergen-Paap F-stat			8730	13247
OTHER				
Coefficient	0.156***	0.034	0.059	0.147*
Std. error	0.016	0.042	0.077	0.075
No. observations	36762	36762	36762	36762
Kleibergen-Paap F-stat			8733	13256
Distance band	20 kms	20 kms	20 kms	20 kms
Year FE	YES	YES	YES	YES
Ward FE		YES	YES	YES
IV			YES	YES
Scheme trends				YES

Sources: BSD-ONS via SDS. Standard errors clustered at ward level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 3.2: Ward level number of plants results

to control for time-invariant ward specific factors. All the coefficients remain significant and positive except the one for other sectors. However, the coefficients for construction and consumer services become significant only at the 10% significance level. The coefficients for total employment, construction, consumer and producer services are reduced with respect to the OLS estimates, while for manufacturing the coefficient more than doubles in size.

The third column uses the instrumental variable strategy. Accessibility index $\hat{A}_{rt}$, which keeps ward employment fixed at the 1997 level and only varies due to road improvements, is used as an instrument for actual accessibility A_{rt}. The first stage F statistics indicate that the instrument is very strong. Accessibility is now positive and significant for manufacturing, construction and producer services, but insignificant for total employment. The estimated coefficients are larger than those in column 2. This would suggest that the fixed-effect estimates of column 2 are downward biased and the instrument is correcting for this. The downward bias could be due to accessibility growing faster in wards experiencing negative shocks to employment. If schemes are aimed at wards with particular employment growth trends, this could be biasing our results. For this reason, in column 4, we control for differences in trends near different schemes. The results show a positive and significant employment effect for total employment and employment in the construction and producer services sectors. The coefficients imply that a one percent increase in the accessibility index leads to a 0.6 percent increase in employment in producer services and 0.5 in construction. For total employment the effect is less significant and the elasticity is lower (0.2).

The results on the number of plants in table 3.2 are qualitatively similar to the employment regressions in table 3.1. The main difference is that the effect of accessibility on total number of plants is consistently positive and highly significant in all specifications, even in the most demanding one. The OLS regressions show a positive relationship between the number of plants and accessibility. Adding fixed effects decreases most of the coefficients and the "other" sector becomes insignificant. Instrumenting accessibility index A_{rt} with $\hat{A}_{rt}$ also makes consumer services coefficient insignificant, but the overall effect and three out of five sector effects are positive and highly significant. After controlling for scheme trends in the fourth column, the coefficient is positive and highly significant when all sectors are combined (elasticity 0.14), and positive and significant in producer service (elasticity 0.4). In addition, accessibility is positive and significant at the 10% level for construction and the "other" sector. The results suggest that the positive employment effect found in table 3.1 is to some extent due to new plants entering the ward attracted by greater accessibility.

Tables 3.3 and 3.4 test the robustness of the results to different distance bands

	(1)	(2)	(3)	(4)	(5)
ALL SECTORS					
Coefficient	0.02	0.191*	0.126	0.04	0.012
Std. error	0.118	0.104	0.108	0.065	0.066
No. observations	15191	53273	36740	15191	15191
Kleibergen-Paap F-stat	12675	15197	10856	3415	4582
MANUFACTURING					
Coefficient	0.407	0.483	0.239	0.534***	0.214
Std. error	0.344	0.294	0.315	0.189	0.196
No. observations	14942	52418	36135	14942	14942
Kleibergen-Paap F-stat	12724	15967	11414	3384	4506
CONSTRUCTION					
Coefficient	0.611**	0.422*	0.557**	0.383**	0.316*
Std. error	0.279	0.231	0.242	0.156	0.162
No. observations	15149	53176	36684	15149	15149
Kleibergen-Paap F-stat	12717	15185	10959	3435	4587
CONSUMER SERVICES					
Coefficient	0.01	0.043	-0.026	0.059	-0.013
Std. error	0.218	0.181	0.19	0.106	0.122
No. observations	15191	53272	36739	15191	15191
Kleibergen-Paap F-stat	12675	15196	10856	3415	4582
PRODUCER SERVICES					
Coefficient	0.409	0.590**	0.575**	0.349**	0.2
Std. error	0.296	0.257	0.256	0.167	0.166
No. observations	15183	53229	36719	15183	15183
Kleibergen-Paap F-stat	12676	15189	10855	3415	4582
OTHER					
Coefficient	-0.163	0.126	0.103	-0.081	-0.095
Std. error	0.164	0.139	0.144	0.084	0.089
No. observations	15191	53273	36740	15191	15191
Kleibergen-Paap F-stat	12675	15197	10856	3415	4582
Distance band	10 kms	30 kms	20 kms	20 kms	20 kms
Year FE	YES	YES	YES	YES	YES
Ward FE	YES	YES	YES	YES	YES
IV	YES	YES	YES	YES	YES
Scheme trends	YES	YES	YES		YES
Controls			YES		
Decay -1.5				YES	YES

Sources: BSD-ONS via SDS. Standard errors clustered at ward level. * p<0.1, ** p<0.05, *** p<0.01

Table 3.3: Robustness results ward employment

(10 kilometres and 30 kilometres) in columns 1 and 2, to introducing further control variables interacted with the trend (straight line distance to closest road scheme, employment rate, average age, proportion of population aged 16-74 with higher education and proportion of population living on social housing) in column 3 and to a different distance decay in columns 4 and 5.

In both tables, limiting the sample to the 10-kilometre band from road schemes reduces sample sizes and increases standard errors and some of the coefficients that are significant with 20-kilometre band become insignificant. The 30-kilometre distance band results are very similar to the 20-kilometre band: we find similar estimates for the effect on total employment and total number of plants, and positive and significant effects of accessibility on construction and producer services. Overall, the positive employment and number of plants effects for construction and producer services seem fairly robust to changes in the distance band and most estimates remain unchanged even after the inclusion of the very stringent set of additional controls.

Adding further controls (ward level 2001 characteristics interacted with a time trend) reduces the coefficient of accessibility for the number of plants when all sectors are pooled (row 1 in column 3) and it is only significant at the 10% level. However, the sector specific estimates for construction and producer services stay positive and significant indicating that the spatial distribution of plants is affected by accessibility changes induced by road improvements. The effect of accessibility on total ward employment becomes insignificant when we add census controls, but they remain significant for construction and producer services.

In columns 4 and 5 of tables 3.3 and 3.4 we use the accessibility measure with the same cost function but a higher distance decay. Column 5 includes scheme trends while column 4 displays the results for the IV fixed-effect estimates. As explained in the previous chapter, we calculate the accessibility index using a cost function defined as the inverse distance function with distance decay equal to one, i.e. $a\left(c_{jrt}\right) = c_{jrt}^{-\alpha}$ with $\alpha = 1$. If α increases to 1.5 this implies that in the calculation of A_{rt} the weights we allocate to wards which are further away from r decrease faster, so we give more importance to wards which are closer to r. In the case of the effect on the number of plants, the coefficients change very little. But in the case of the ward employment, the coefficients become almost all insignificant, specially when we include scheme trends (in column 5). This suggests that the effects on total employment we found in table 3.1 are partially driven by the employment in wards which are further away from r, and when the weight to these is reduced the effect disappears.

Tables 3.1 to 3.4 suggest that increased accessibility leads to increased employment and number of plants, overall and for some sectors. Overall, the elasticities of

	(1)	(2)	(3)	(4)	(5)
ALL SECTORS					
Coefficient	0.079	0.135***	0.090*	0.094***	0.064**
Std. error	0.053	0.05	0.049	0.031	0.029
No. observations	15202	53284	36762	36762	36762
Kleibergen-Paap F-stat	12681	15196	10856	3697	3777
MANUFACTURING					
Coefficient	0.096	0.254	0.14	0.235**	0.061
Std. error	0.189	0.162	0.169	0.099	0.097
No. observations	14958	52440	36166	36166	36166
Kleibergen-Paap F-stat	12736	15964	11412	3682	3840
CONSTRUCTION					
Coefficient	0.221	0.229**	0.244**	0.210***	0.149**
Std. error	0.136	0.115	0.123	0.075	0.076
No. observations	15162	53190	36709	36709	36709
Kleibergen-Paap F-stat	12731	15184	10957	3707	3778
CONSUMER SERVICES					
Coefficient	-0.036	-0.019	-0.069	0.063	-0.001
Std. error	0.084	0.077	0.082	0.048	0.045
No. observations	15202	53283	36761	36761	36761
Kleibergen-Paap F-stat	12681	15195	10857	3698	3777
PRODUCER SERVICES					
Coefficient	0.359**	0.399***	0.347**	0.378***	0.217**
Std. error	0.157	0.142	0.139	0.098	0.087
No. observations	15194	53242	36742	36742	36742
Kleibergen-Paap F-stat	12683	15187	10855	3697	3779
OTHER					
Coefficient	0.065	0.137*	0.134*	0.012	0.058
Std. error	0.085	0.075	0.074	0.046	0.045
No. observations	15202	53284	36762	36762	36762
Kleibergen-Paap F-stat	12681	15196	10856	3697	3777
Distance band	10 kms	30 kms	20 kms	20 kms	20 kms
Year FE	YES	YES	YES	YES	YES
Ward FE	YES	YES	YES	YES	YES
IV	YES	YES	YES	YES	YES
Scheme trends	YES	YES	YES		YES
Controls			YES		
Decay -1.5				YES	YES

Sources: BSD-ONS via SDS. Standard errors clustered at ward level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 3.4: Robustness results ward number of plants

employment were somewhat higher than the elasticities of the number of plants. Taken at face value, and disregarding the wide confidence intervals, the higher elasticity of employment could arise if existing plants increase their employment or entering plants are on average larger than old ones. In section 3.4.2 we shed some light on the former channel by studying employment responses at plant level with the ARD plant data.

3.3.3 Results: productivity

In this section we report results on the effect of accessibility on productivity at the ward level. Tables 3.5 and 3.6 show the results for the two-step methodology explained in section 3.3.1.2.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep variable: log of gross output	**ALL SECTORS**	MANUFACT.	CONSTRUCT.	CONSUMER SERVICES	PRODUCER SERVICES	OTHER
Log of employment	0.348***	0.370***	0.365***	0.231***	0.397***	0.363***
	0.004	0.006	0.011	0.007	0.008	0.009
Log of capital	0.065***	0.084***	0.092***	0.032***	0.135***	0.088***
	0.002	0.004	0.012	0.002	0.008	0.007
Log of intermediates	0.585***	0.547***	0.523***	0.743***	0.449***	0.544***
	0.003	0.007	0.014	0.006	0.009	0.01
No. observations	227377	51210	22491	72470	43453	37753
R^2	0.939	0.967	0.95	0.959	0.879	0.922
Ward*year FE	YES	YES	YES	YES	YES	YES
2 digit sector FE	YES	YES	YES	YES	YES	YES

Sources: BSD and ARD-ONS via SDS. Standard errors clustered at firm level. * p<0.1, ** p<0.05, *** p<0.01

Table 3.5: Local unit production function estimates

We first estimate ward productivity shifters for every year using all single-plant firms within 20 kilometres of the road schemes. The results for the first step are in Table 3.5. The dependent variable is log of gross output and we control for inputs (log of labour, capital and intermediate inputs). We predict the ward-by-year fixed effects and average them up to each ward in every year. In the second step, we then regress these estimated productivity effects on the accessibility variable.

Table 3.6 shows the results of the second step. The coefficient on accessibility is positive and significant using OLS, but turns negative and insignificant when we add ward fixed effects, apart from consumer services where the coefficient is positive and insignificant. Columns 3-4 show the IV results. Even if the instrument is very string, the IV estimates are imprecise and none of the IV estimates is statistically significant. These results seem to suggest that there is no effect of accessibility on ward aggregate productivity, although the coefficients are too imprecise to draw strong conclusions.

	(1)	(2)	(3)	(4)
ALL SECTORS				
Coefficient	0.059***	-0.103	-0.136	-0.311
Std. error	0.005	0.106	0.208	0.225
No. observations	19667	19609	19609	19609
Kleibergen-Paap F-stat			3771	8491
MANUFACTURING				
Coefficient	0.039***	-0.076	-0.242	-0.364
Std. error	0.006	0.133	0.277	0.299
No. observations	9145	8572	8572	8572
Kleibergen-Paap F-stat			1784	3086
CONSTRUCTION				
Coefficient	0.051***	-0.161	0.339	0.423
Std. error	0.008	0.235	0.423	0.453
No. observations	6197	5319	5319	5319
Kleibergen-Paap F-stat			890	1683
CONSUMER SERVICES				
Coefficient	0.042***	0.057	0.13	0.047
Std. error	0.006	0.169	0.25	0.274
No. observations	13012	12632	12632	12632
Kleibergen-Paap F-stat			2191	4500
PRODUCER SERVICES				
Coefficient	0.104***	-0.058	-0.163	-0.631
Std. error	0.01	0.285	0.85	0.918
No. observations	9169	8499	8499	8499
Kleibergen-Paap F-stat			889	1309
OTHER				
Coefficient	0.108***	-0.086	-0.141	-0.258
Std. error	0.015	0.356	0.68	0.734
No. observations	3561	3285	3285	3285
Kleibergen-Paap F-stat			1140	1262
Distance band	20 kms	20 kms	20 kms	20 kms
Year FE	YES	YES	YES	YES
Ward FE		YES	YES	YES
IV			YES	YES
Scheme trends				YES

Sources: BSD and ARD-ONS via SDS. Standard errors clustered at firm level. * p<0.1, ** p<0.05, *** p<0.01

Table 3.6: Ward level productivity results

3.4 Effects on firm outcomes

3.4.1 Empirical specification and identification strategy

For the estimation of employment and productivity effects at the firm level we use the Annual Respondent Database (ARD) data. Employment regressions use all plants in the sample, including multi-plant firms. To obtain these results we are able to use a larger version of the ARD for years 1998 to 2008[4]. For the productivity regressions we use a smaller sample for three reasons. Firstly, the capital data from CeRIBA is only available for the period 1998-2004. Secondly, we need to use more information from the balance-sheet data (namely gross output and intermediates) and given the number of missing values, this reduces considerably the number of observations we can use in the regressions. Finally, as discussed above, there is no good way to apportion value added and capital stock of a multi-plant firm to its plants so we only use the single-plant firms for which we have all the necessary information.

Given the structure of the data used, if a plant changes its location (ward) it is labeled as a different plant. Therefore plant identifier is location-specific. The firm level regressions (with fixed effects) essentially compare firms that remain in situ in the same location over time and experience larger or smaller accessibility changes. Thus, the plant level analysis differs conceptually from the ward level regressions that estimate aggregate effects allowing for entry and exit.

The underlying model is identical to (3.2) with the exception that the units of observation are plants instead of wards r. Starting from an OLS model we estimate the following relationship:

$$y_{irt} = \beta A_{rt} + \tau_t + \varepsilon_{irt} \tag{3.8}$$

where A_{rt} is the accessibility of ward r in which the plant is located at time t. In the plant employment regressions y_{irt} is log of employment of plant i located in r at time t. In the single-plant firm productivity regressions y_{irt} is log real gross output and the model further includes input variables (log of employment, capital and intermediate inputs). Year fixed effects τ_t capture general changes that influence all firms and locations in the study area in a given year (e.g. macro shocks).

In order to control for unobservable time-invariant plant/ward characteristics correlated with accessibility and our outcome variables, we include plant fixed effects μ_i. The plant fixed effects μ_i are ward-specific and, thus, include unobserved time invariant productivity advantages for all firms located in ward r. Specification

[4] We use the balance-sheet employment data in ARD (which is slightly different from the BSD figures) because it is more reliable as it is based in a survey.

(3.8) becomes:

$$y_{irt} = \beta A_{rt} + \mu_i + \tau_t + \varepsilon_{irt} \tag{3.9}$$

Again, we use time-demeaning to control for time-invariant plant (and ward) heterogeneity that can be correlated with accessibility.

$$(y_{irt} - \bar{y}_{i.}) = \beta \left(A_{rt} - \bar{A}_{r.}\right) + \tau_t + \xi_{irt} \tag{3.10}$$

where ξ_{irt} = where $\bar{y}_{i.}$ and $\bar{A}_{r.}$ are ward averages of the firm outcome and of log ward accessibility over the time period (1998-2008), τ_t are year dummies which capture shocks affecting the variation over the firm means and $\xi_{rt} = (\varepsilon_{irt} - \bar{\varepsilon}_{ir.})$.

In our dataset when plants relocate they receive a new identifier, so we only use the variation over time for plants that remain in the same location. Therefore, variation of accessibility around the ward mean is not driven by mobility of plant i because plant fixed effects are ward-specific and accessibility is measured at the ward level. All the variation in accessibility is due to restructuring of employment (in other firms) and transport improvements. To address the issue of endogenous determination of accessibility, we once again instrument accessibility index A_{rt} with accessibility index $\hat{A}_{rt}$, which only picks up changes in the transport network.

Estimation of (3.10) using within-plant changes in a panel of plants is only feasible using plants that exist, and appear in the data, both before and after the opening of the transport schemes that are used as the source of identifying variation in accessibility. This introduces sample selection issues. Firstly, firms that stay in the location of the transport scheme are likely to be those that can benefit most from it. Secondly, the method does not capture changes in employment or productivity associated with the opening of new plants. In addition, there are sampling-related reasons why some firms appear in our data in multiple years whilst others do not. These caveats aside, the IV estimation of β from the changes within plants over time provides guidance on the micro-level impacts of transport improvements for firms, which is one of the components of the aggregate ward level effects, and it is interesting in its own right.

3.4.2 Results: employment

Table 3.7 reports the results from the plant level regressions which essentially compare firms that stay in the same location over time and experience larger or smaller accessibility changes. The structure of the table is similar to tables 3.1 and 3.2.

The first column showing the simple OLS results indicates that more accessible places have smaller plants, apart from producer services where the relationship between accessibility and plant size is positive. In the second column we intro-

	(1)	(2)	(3)	(4)
ALL SECTORS				
Coefficient	-0.028***	0.026	0.02	-0.059
Std. error	0.004	0.028	0.071	0.072
No. observations	862302	862302	861811	861811
Kleibergen-Paap F-stat			112322	180275
MANUFACTURING				
Coefficient	-0.164***	0.149***	0.141*	0.161**
Std. error	0.002	0.031	0.079	0.081
No. observations	735500	735500	734342	734342
Kleibergen-Paap F-stat			122229	193050
CONSTRUCTION				
Coefficient	-0.017***	0.094***	0.064	-0.022
Std. error	0.002	0.031	0.069	0.071
No. observations	777390	777390	774945	774945
Kleibergen-Paap F-stat			148547	233255
CONSUMER SERVICES				
Coefficient	-0.022***	0.038**	-0.019	-0.038
Std. error	0.001	0.018	0.046	0.048
No. observations	1808808	1808808	1807020	1807020
Kleibergen-Paap F-stat			214606	359424
PRODUCER SERVICES				
Coefficient	0.023***	0.072***	0.034	-0.081
Std. error	0.001	0.021	0.057	0.06
No. observations	1702561	1702561	1699223	1699223
Kleibergen-Paap F-stat			136929	191720
OTHER				
Coefficient	-0.084***	0.042*	0.037	-0.039
Std. error	0.002	0.024	0.062	0.065
No. observations	1196191	1196191	1189212	1189212
Kleibergen-Paap F-stat			136776	204641
Distance band	20 kms	20 kms	20 kms	20 kms
Year FE	YES	YES	YES	YES
Ward FE		YES	YES	YES
IV			YES	YES
Scheme trends				YES

Sources: BSD and ARD-ONS via SDS. Standard errors clustered at firm level. * p<0.1, ** p<0.05, *** p<0.01

Table 3.7: Local unit employment results

duce plant fixed effects and, interestingly, the coefficients become positive and significant in all of the five sectors. This suggests that an increase in effective density is associated with increasing plant employment. However, addressing endogeneity concerns through the IV method increases standard errors substantially and decreases the size of the coefficients which become mostly insignificant. However, for manufacturing, the coefficient on accessibility stays positive and significant even in column 4.

These results suggest that the aggregate effects on ward employment found in table 3.1 might be mostly driven by increasing number of plants at the ward level[5], as individual plant employment does not seem to be responding to changes in accessibility. Thus, the employment growth at the ward level could be due to newly created plants in the ward or to plants moving into the ward from other locations.

This result relates to the theoretical channels highlighted in chapter 2. Increased effective density could make locations more desirable as consumers and producers become more accessible, driving plants to locate in these areas. Also, increased accessibility might reduce the fixed-cost of creating a new plant, as prospective benefits might be higher due to better access to consumers and producers, so the number of plants created in the ward might increase. Additionally, multi-plant firms might be relocating plants from other locations to the wards in which higher accessibility changes are taking place to take advantage of the better connectivity. Finally, this finding is consistent with previous evidence found on the effect of transport infrastructure on the creation of new plants, provided for example by Holl (2004c) and Melo et al. (2010).

3.4.3 Results: productivity

Results from the firm level production function regressions are reported in Table 9. All models control for capital, labour and intermediate inputs. As explained above, the sample is based on a panel of single plant firms for years 1998 to 2004. The dependent variable is log of gross output. Because the regressions control for inputs, the coefficients correspond to the effect of accessibility on Total Factor Productivity (TFP).

OLS results show elasticities of 0.03 – 0.19. Adding firm fixed effects increases standard errors and the estimates become insignificant. Turning to the IV estimates in columns 3 and 4, construction sector has a positive and significant coefficient in column 3, but it becomes insignificant in column 4 where scheme trends are con-

[5] We experimented with the BSD data in order to investigate if the ward employment effects came from increased employment from existing plants or from employment growth due to entry of new plants. However, BSD individual plant identifiers are quite noisy over time making this analysis difficult to undertake.

	(1)	(2)	(3)	(4)
ALL SECTORS				
Coefficient	0.092***	-0.036	0.058	0.074
Std. error	0.005	0.066	0.134	0.149
No. observations	39800	34364	34364	34364
Kleibergen-Paap F-stat			4189	7746
MANUFACTURING				
Coefficient	0.025***	0.000	-0.227	-0.279*
Std. error	0.007	0.083	0.158	0.16
No. observations	15419	13894	13894	13894
Kleibergen-Paap F-stat			2385	5425
CONSTRUCTION				
Coefficient	0.080***	0.347	0.715**	0.507
Std. error	0.017	0.24	0.292	0.321
No. observations	2213	1886	1886	1886
Kleibergen-Paap F-stat			472	790
CONSUMER SERVICES				
Coefficient	0.081***	0.101	0.361	0.493*
Std. error	0.008	0.123	0.253	0.284
No. observations	9743	8042	8042	8042
Kleibergen-Paap F-stat			610	1410
PRODUCER SERVICES				
Coefficient	0.188***	-0.263	-0.709	-1.042
Std. error	0.014	0.161	0.532	0.883
No. observations	6622	5548	5548	5548
Kleibergen-Paap F-stat			426	233
OTHER				
Coefficient	0.114***	-0.127	0.667	0.689
Std. error	0.013	0.228	0.463	0.447
No. observations	5803	4994	4994	4994
Kleibergen-Paap F-stat			393	748
Distance band	20 kms	20 kms	20 kms	20 kms
Labour, capital & intermediates	YES	YES	YES	YES
Sector-Year FE	YES	YES	YES	YES
Ward FE		YES	YES	YES
IV			YES	YES
Scheme trends				YES

Sources: BSD and ARD-ONS via SDS. Standard errors clustered at firm level. * p<0.1, ** p<0.05, *** p<0.01

Table 3.8: Firm productivity results

trolled for. However, in columns 3 and 4, we have some indication of a negative effect for manufacturing and positive affect for construction.

The lack of effects at the plant level is consistent with the lack of effects at the aggregate level found in section 3.3.3. Firms might be adjusting other outcomes (studied below) but our results suggest that increased accessibility is not acting as a productivity shifter neither at the aggregate (ward) not the individual level (plant).

3.4.4 Results: other firm outcomes

Due to the lack of evidence and the imprecision of the estimates in table 3.8, in tables B.10 to B.12 we experiment with alternative outcome variables at the firm level. We estimate similar specifications to 3.9 but changing the outcome variable y_{irt}. For these regression we use sample of both single and multi-plants but restrict it to the 1998-2004 period for comparability with the plant TFP estimates.

In table B.10 we use gross output per worker as the dependent variable, in table B.11 we use the average wage per worker (firm wage bill over the firm employment) and in table B.12 we test the effect of accessibility on firm gross output.

For all three outcome variables the results are very similar: using OLS we find positive relationships between accessibility and the economic outcome, but these become insignificant once we add fixed effects and imprecise when we instrument and add scheme dummies.

These results suggest that plants are not adjusting their output, and given the lack of response of plant employment found in section 3.4.2, the lack of effect on labour productivity (output per worker) is consistent with what we would expect. Also, the absence of TFP and employment effects at the plant level is consistent with the lack of effect of accessibility on average wage per worker, because neither productivity nor demand for workers is changing at the plant level. However, the estimates are quite imprecise as in the case of the TFP results.

3.5 Conclusions

This chapter uses a novel methodology in order to assess the productivity and employment effects from transport improvements at a very detailed geographic scale. We construct employment accessibility measures at the ward level using GIS network analysis with data on major road schemes and the register of local business employment. These accessibility indices are used for the analysis of employment and productivity at the firm level and ward level. We propose an instrumental variables strategy addressing the likely endogeneity of agglomeration by relying solely on transport improvements for identification in a panel data fixed effects setting.

Furthermore, we address the potential endogeneity of the road construction placement by focusing on wards and firms which are close to schemes. We argue that methods using cross-sectional variation in effective density, or variation due to spatial restructuring of employment, do not have a causal interpretation.

Overall, when we examine the effect of all major road transport improvements with firm level data focusing on firms and plants that remain in situ before and after the opening of new road links, we find insignificant effects on the employment and productivity of firms. Neither do we find significant productivity effects when we examine total factor productivity at ward level allowing for start-ups and closures. However, ward level employment and local unit count regressions give some support to the idea that increased accessibility caused by a better road network may lead to increased economic activity. The fact that plant level employment is unaffected, apart from manufacturing, suggests that the positive ward level employment effect is mainly attributable to increased entry or decreased exit.

The lack of robust evidence at the plant level could be driven by measurement error in the accessibility indices due to the features of the road network data used, as we might be failing to capture all the effects of road construction (see chapter 2 for more details). However, it should be noticed that, to our knowledge, this research is the first attempt to construct an accessibility index at such a detailed geographical level for a long period of time (11 years). Even if limited, our measure allows us to focus on a specific channel, road construction, through which policy can affect economic performance.

Our results add substantially to the existing evidence on the effects of transport policy on aggregate and firm economic outcomes. Our analysis highlights the importance of addressing endogeneity issues in a convincing way. We argue that utilizing small scale spatial variation in the impact of transport improvements on effective density offers a promising quasi-experimental setting, even though data requirements are high. We provide evidence both at the aggregate and at the individual level, which allows us to investigate the micro channels driving the aggregate results. Furthermore, we test the effects on a variety of outcomes (employment, number of plants, total and labour productivity, gross output and wages), which sheds some light on the impacts of transport on economic outcomes for which the existing evidence is scarce.

Chapter 4

Economic Impacts of Transport Policy: Labour market outcomes

4.1 Introduction

In this chapter we present and analyse the results of the impact of accessibility on individual labour market outcomes. We use microdata from two British datasets for the period 1998-2008 and 2002-2008 and test the effect of changes in accessibility on several outcomes: nominal wages, hourly earnings, hours worked, employment status and commuting time.

As discussed in chapter 2, transport policy can affect labour markets mainly through its impact on reducing commuting costs (measured as travel time to workplace in our context). Road improvements bring firms and workers closer together and this could have an impact on the quantity and quality of the interactions in the labour market (job matches), on the size of the markets (spatial competition for workers and for jobs) and on the emergence and scope of agglomeration economies (productivity effects on wages). Additionally, if firms react to road improvements by changing their demand for workers or by increasing wages, this would have an effect on worker outcomes. However, results from chapter 3 suggest that the effect of accessibility on firm employment and wages is nil. Given the lack of employment and wage effects for individual firms, it becomes of increased interest to study the potential responses of the labour markets to transport policy.

In order to carefully identify these effects we make use of large worker and firm microdata datasets, which were provided by the UK Office for National Statistics (ONS)[1]. We capture the effect of road improvements using a measure of accessibility to employment (effective density), as explained in detail in the previous chapters. We calculate the measure of accessibility to employment at the ward level using data

[1] Through the "Virtual Microdata Laboratory" service and the "Secure Data Service".

from the census of establishments (Business Structure Database) – which provides information on the universe of British plants reporting their location (postcode), their sector of activity and their number of employees– and data on road improvements and travel times in Great Britain.

For the individual labour market outcomes analysis we use two additional datasets: the Annual Survey of Hours and Earnings (ASHE) and the Labour Force Survey (LFS). The first survey is a panel of employees, and it provides information on their earnings, labour supply, job characteristics, and the postcode where the firm they work for is located and where they they live. The second dataset contains quarterly information on a sample of households and includes information on a wide range of labour market outcomes, as well as a large list of personal and household characteristics.

We exploit data on a panel of employees to estimate the effect of accessibility both from home and from work on individual earnings and hours worked. Due to data restrictions, for these regressions we use data for years 2002 to 2008. In what we consider our most reliable results, in order to isolate the effect of accessibility that stems only from transport improvements, we use the variation over time for individuals for a given ward of work and ward of home location. As before, we also make use of instrumental variables to reduce the potential endogeneity caused by employment relocation across the space.

To investigate the effect of accessibility on employment status and travel times we use repeated cross sections from the LFS for years 1998 to 2008. We control for a large set of individual and location characteristics as well as age-gender-ethnicity specific fixed-effects and time and ward dummies.

We believe we contribute to the existing literature in three ways. First, we use very rich worker microdata and a novel dataset on road improvements, which allow us to study the relationship between transport policy and labour market outcomes at a very detailed geographical scale and to focus on different groups of individuals. Secondly, given the quality of our data, we can adopt a careful empirical strategy that allows us to tackle several identification issues which could undermine the validity of our results. This makes it possible to examine some under-investigated channels through which transport could affect labour market outcomes. Finally, we provide empirical evidence on some strands of the existing theoretical urban labour economics literature, in particular the spatial mismatch theory or the effects of agglomeration on wages.

The rest of the chapter is organised as follows. Section 4.2 focuses on the results on individual wages and hours using the ASHE data. We present the specification and the identification strategy used (section 4.2.1), then we describe data used (section 4.2.2), and finally the empirical results (section 4.2.3). Section 4.3 has presents the results for the employment status and travel times, obtained using LFS data. Finally, section 4.4 concludes.

4.2 Effects on wages and hours worked

4.2.1 Empirical specification and identification strategy

To obtain results on individual earnings and hours worked we use the Annual Survey of Hours and Earnings (more details in section 4.2.2.1). We use a panel of workers to estimate the effect of accessibility to employment, both from work and from home, on individual labour market outcomes. In our main regressions we use a panel of employees surveyed between years 2002 and 2008. We study three outcomes: weekly wages, weekly hours worked and hourly earnings. We look both at basic and at total (which includes overtime) outcomes.

A worker i living in ward h and working in ward w at time t has labour market outcome y_{ihwt}. Given that accessibility is measured at the ward level, we ignore changes of home or job within the wards[2]. At each point in time, workers live and work in specific wards. Over time, the worker location can change, if he changes jobs, he changes home or he changes both. Initially, we use the variation in accessibility and labour market outcomes for an individual which are also driven by relocations (more on this below). The relationship we estimate is:

$$y_{ihwt} = \beta_0 + \beta_1 A_{iht} + \beta_2 A_{iwt} + \beta_3 c_{ihwt} + \theta X_{it} + \delta Z_{ht} + \lambda W_{wt} + \mu_i + \xi_t + \varepsilon_{ihwt} \quad (4.1)$$

where y_{ihwt} is the individual labour market outcome of worker i at time t (wages or hours), A_{iht} is accessibility to employment from home ward h at time t, A_{iwt} is accessibility to employment from work ward w at time t, c_{ihwt} denotes the commuting costs between home and work at time t, X_{it} is a vector of personal and job characteristics, Z_{ht} is a matrix of home ward characteristics and W_{wt} is a matrix of work ward characteristics. ξ_t are year fixed effects that control for common shocks affecting all wards in a given year. ε_{ihwt} is the idiosyncratic error and β_0 is a constant term. Both the labour market outcomes and the accessibility indices are transformed to natural logarithms so we can interpret the coefficients as elasticities.

The dataset used does not provide information on travel time or distance to workplace. We use the road networks between 2002 and 2008 (created as explained in the methodology chapter) to calculate optimal travel time between ward of home and ward of work along the road network. We do not have any information on the

[2]In fact, most of workers change ward of work when they change job (enterprise). In the robustness checks we control for job and home changes within wards and the results remain unchanged.

travel mode of the workers. It could be the case that they are not commuting by road or not commuting using the optimal route predicted by ArcGIS®. However, using optimal travel time to proxy for commuting costs has the advantage of getting rid of potential measurement error on self-reported travel time[3]. In practice, we estimate a reduced-form effect of commuting costs, in which our measure of commuting cost is the optimal travel time through the road network. Moreover, as this travel time measure changes over time due to the road improvements, it does not drop out when we include individual-home ward-work ward fixed-effects below, as opposed to straight geodesic distance (used for example by Graham & Melo, 2009, to approximate commuting costs).

We include this information in the estimated specification (4.1) in order to capture the effect of commuting costs on labour market outcomes. This way we can estimate the effect of accessibility from home and work conditional on commuting costs. This helps to interpret the results given the numerous theoretical channels through which transport policy can affect labour market outcomes, as discussed in chapter 2. For example, some of the effects of transport policy on wages could be due to employers compensating workers for longer commutes and some could come through increased spatial competition or agglomeration externalities. By controlling for commuting costs in equation (4.1) we can be more certain that the effects of accessibility are not due to compensation for longer commutes as we are explicitly controlling for that. Moreover, the interpretation of the coefficient β_3 also informs us about the relationship between commuting costs and labour market outcomes.

We are interested in parameters β_1 and β_2. We may be worried that there are (time invariant) unobservable individual characteristics that affect both individual labour market outcomes and accessibility indices (in levels) at the same time, and that are not included in X_{it}. For example, more able individuals may live or work in areas where accessibility and wages are higher. We include worker fixed effects μ_i to control for this (which in practice is equivalent to estimate the demeaned model). We can additionally control for time-invariant home ward (μ_h) and work ward (μ_w) characteristics, which are potentially correlated with individual wages, accessibility from home and from work and the individual time-invariant characteristics.

By introducing the individual fixed effects we estimate coefficients β_1 and β_2 using the variation of accessibility over time with respect to the average individual accessibility. At each point in time the worker may hold different jobs in different locations (w) and live in different places (h). Therefore the accessibility to which the worker is exposed at home and at work varies over time for three reasons: when

[3] Which is common, as noted for example by Gutiérrez-i-Puigarnau & van Ommeren (2010). Indeed, when tabulating travel times reported in LFS responses are disproportionably cumulated in values such 5, 10, 15, 20 and 30 minutes, likely due to rounding-up.

home-work location changes, when employment changes or when transport improvements (that change travel times) take place.

If workers are sorting spatially in order to take advantage of the changes in accessibility we would not be able to identify the separate effects on labour market outcomes which stem from (endogenous) sorting and those which are due to changes in accessibility for a given location (externalities or spatial competition). Sorting could be an outcome of accessibility or could be due to other unobservable reasons correlated with the labour market outcomes. For example, if workers with higher ability move to areas where accessibility and wages are growing faster, then the correlation between the changes in accessibility and the (demeaned) error term could be different from zero. The same could occur if more able workers choose jobs in areas where wages and accessibility are growing pro-cyclically. For these reasons, even after controlling for unobservable time invariant characteristics of the individuals, there would still be reasons to think the estimates of β_1 and β_2 could be biased.

To investigate this issue we define individual-home ward-work ward specific fixed effects, μ_{ihw}. Replacing the individual fixed effects in (4.1) with μ_{ihw} and we obtain:

$$y_{ihwt} = \beta_0 + \beta_1 A_{iht} + \beta_2 A_{iwt} + \beta_3 c_{ihwt} + \theta X_{it} + \delta Z_{ht} + \lambda W_{wt} + \mu_{ihw} + \xi_t + \varepsilon_{ihwt} \quad (4.2)$$

As we did in chapter 3, we can rewrite equation (4.2) in demeaned terms by subtracting the individual-fixed location means across time (focusing only on the accessibility measures):

$$(y_{ihwt} - \bar{y}_{i.}) = \beta_1 (A_{iht} - \bar{A}_{i.}) + \beta_2 (A_{iwt} - \bar{A}_{i.}) + (\varepsilon_{ihwt} - \bar{\varepsilon}_{i.}) \quad (4.3)$$

By estimating (4.2) we are identifying the effects of accessibility on individual labour market outcomes exploiting the changes of accessibility over time for an individual while keeping their home and work locations constant.

To implement this, in our panel, each individual is allocated a different fixed-effect depending in the pair work-home location. Therefore, we have two types of individuals: those that never move location pair (at least while observed in the data) and those that eventually change ward of work, ward of residence or both. Individuals in this last group may be sorting due to changes in accessibility, so in contrast to (4.1), (4.2) does not use this variation for the estimation of the effects of accessibility[4]. That is, when including μ_{ihw}, coefficients β_1 and β_2 are not capturing the effect of endogenously determined relocations after the changes in accessibility take place (that could be cause by increased accessibility or other unobserved reas-

[4]In the robustness checks we obtain the results focusing on the group of non-movers and the results remain unchanged.

ons). The comparison of the estimates obtained by estimating (4.1) and (4.2) thus inform us about the role played by sorting.

However, there are still other possible sources of bias in the estimation of the effect of accessibility. Even when keeping the home-work location fixed, accessibility changes around the individual means due to changes in employment (numerator) and in travel times due to transport improvements (denominator). We have the same endogeneity sources as in the case of firms and ward outcomes discussed in chapter 3. If changes in labour outcomes, for example wages, increase accessibility from home by means of attracting workers to the ward or around the ward in which the worker lives or works, we could have a reverse causality problem which would challenge the validity of the estimates. There could also exist unobservable trends which affect both the location of employment and the labour market outcomes which could bias the estimates.

To overcome this, as before, we instrument accessibility A_{rt} using $\hat{A}_{iht}$, the accessibility measure that keeps employment fixed before the period of analysis. In the case of the estimations using 2002-2008 ASHE data this year is 2001[5]. By doing this, we only use the changes in accessibility that come from transport improvements and not from employment growth. The instrument is strong by construction, because both the instrument and the instrumented variable use the variation which comes from road improvements. For the instrument to be valid, we need the instrument to be uncorrelated with unobservable shocks that affect changes in employment.

This relates to the issue of the potential endogeneity of the placement of transport investments (which affect travel times used in the calculation of A_{rt}). It could be argued that the instrument is not valid if the transport improvements are aimed at areas which are experiencing unobservable shock which are correlated with individual labour market outcomes. Transport investments may be taking place in areas in which workers would have done better anyway. Nevertheless, as explained in previous chapters, improving economic outcomes is not on of the key objectives of the transport investments carried out by the Government. Transport projects are generally aimed at a higher spatial scale and designed to improve safety or reduce congestion within a wider area.

However it could still be the case that transport improvement placements are endogenous to unobservable trends in labour market outcomes. To correctly identify the effects we once again define three distance bands which indicate if the ward is at 10, 20 or 30 kilometres from any improvement undertaken during the study period. Individuals placed within a specific distance band are more likely to be exposed to

[5] As compared to 1997 employment used in the calculation of $\hat{A}_{iht}$ in the previous chapter, which is also used in the estimations using LFS data below.

similar shocks and improvements are quite unlikely to be aimed at specific individuals within these narrowly defined distance bands. Within these bands we compare workers who are close to improvements at some point in time but that vary in the intensity and the timing of the treatment. The main results are obtained using the 30 kilometres band from both home and work wards[6]. These bands could overlap or not, depending how separated are the locations of work and residence. Given the size of the sample, we can drop the individuals which are within 1 kilometre of the improvements without losing a substantial fraction of the observations.

As in the case of the aggregate and firm outcomes, in order to ensure that the instrument is uncorrelated with the underlying trends, we further control for differential trends in the vicinity of road schemes in our final specification. As in chapter 3, this is done by including a set of scheme dummies (31 schemes) interacted with year in equation (4.2). In some specifications we also control for the distance to the closest improvement within the distance bands (trends and levels).

We also control for differential trends of the wards (home and work) based on 2001 characteristics[7]. We used CASWEB data to calculate the share of population aged 15-64 with higher education, mean age of population, share of population living on social housing, the rate of unemployment, proportion of workers commuting using motor vehicles and the average distance traveled to work. We also calculated a residential density measure, using address counts data in 2001 from the National Statistics Postcode Directory (NSPD) and the area of the wards in square kilometres, obtained from EDINA-UKBORDERS. We interact these characteristics with a linear trend. They control for differential growth in labour market outcomes, e.g. wages, depending on the level of these characteristics in 2001, before our period of analysis which starts in 2002. In the estimation of (4.1) we also introduce these characteristics in levels.

We furthermore control for individual personal or job characteristics. Some of these characteristics, for example full time status or occupation, could be regarded as "bad controls" because they could be outcomes of the transport policy. To help address this issue, we define the level of the characteristics at the beginning of the period (the first time the individual is observed within each of the two panel definitions) and we interact that level with a time-trend. By doing this we control for differential trends in the evolution of the labour outcome depending on the initial

[6]The choice of the 30-kilometre distance band for the baseline results in this chapter is different from the 20-kilometre distance band used in the previous chapter. We choose 30 instead of 20 because, as shown in section 4.2.3.1, we need a larger sample size in order to obtain a larger set of significant coefficients. In any case, for the main results which use the individual-home-ward fixed effects discussed in section 4.2.3.2, the coefficients are very similar and significant across different distance bands.

[7]We do this instead of including home-ward and work-ward dummies due to the large number of fixed-effects to be estimated, which is technically difficult in our data.

level of the job and personal characteristics. We use occupation trends (9 categories defined above), age group (10 year groups, from 16 to 65), full-time status and gender trends. In the estimation of (4.1) we also introduce these characteristics in levels.

Finally, we implement 2-way clustering of the standard errors to correct for arbitrary within-group correlation of the individual shocks in two distinct non-nested categories. There are 4 dimensions in our data: individuals, time, ward of home and ward of work. In the case of the estimation of (4.1) the categories are year and individual: the resulting standard errors are robust to arbitrary within-panel autocorrelation (clustering on panel id) and to arbitrary contemporaneous cross-panel correlation (clustering on time). Ideally we would have clustered at the level of the "treatment" (accessibility changes), which are the wards of location. But these categories are nested because individuals change location over time. In the case of the estimation of (4.2) we can cluster the standard errors at the home-ward and work-ward level, because now these categories are not-nested across time, i.e. individuals keep the same home-work pair over time in the panels as defined in (4.2). By doing this, we allow the errors of the workers to be correlated at the treatment level (wards). In addition, the standard errors are also robust to arbitrary heteroskedasticity.

4.2.2 Data

4.2.2.1 The Annual Survey of Hours and Earnings

The ASHE is an annual survey of the earnings of employees in Great Britain, which from 2004 replaced the New Earnings Survey (NES). Its primary purpose is to obtain information about the levels, distribution and make-up of earnings, and for the collective agreements that cover them. It is designed to represent all categories of employees in businesses of all kinds and sizes. The questionnaire is directed to the employer, who completes it on the basis of payroll records for the employee. The earnings, hours of work and other information relate to a specified week in April of each year.

ASHE is based on a survey of a 1% sample of employees on the Inland Revenue PAYE register (Pay As You Earn). The information is provided by the employer. The sample consists of employees whose National Insurance numbers end with two specific digits. It covers approximately 160,000 individuals a year. The survey is designed as a panel of workers, in which the same workers are observed for multiple years. The sample is replenished as workers leave the PAYE system (e.g. to self employment, retirement, overseas or death) and new workers enter it (e.g. from school, self-employment, immigration).

ASHE contains information on the make-up of weekly earnings and hours worked (basic, total and overtime), occupation (using Standard Occupational Classification – SOC), industrial sector (using Standard Industry Classification – SIC), collective agreement status, whether the job is private or public sector, age, gender, postcode of workplace, and from 2002, postcode of residence. We use the information on the postcodes to allocate accessibility from home and from work wards using the National Statistics Postcode Directory (NSPD).

In order to clean the data we drop the 0.5% top and bottom extreme values of the labour market outcome variables (wages and hours) and of the commuting times. We define total pay consistently over the whole period 2002-2008[8]. We also removed observations with negative values of the variables and individuals which show inconsistency in their age or gender over time. We only keep main jobs for those individuals that have more than one job in the same year. Finally, we drop the individuals for which earnings were affected by absence (loss of pay) and those paid at trainee/junior rates. In the main results, we focus on employees working in the private sector. We believe that the flexibility of wages and hours would be greater in the private than in the public sector, because the latter might be more regulated or constraint to specific types of jobs. That said, in the robustness check we also include public sector workers and the results remain unchanged.

ASHE provides information on the occupation level of the individuals using SOC 2000 codes from 2002. We define broader occupation codes using the first digit of the code[9]. We also define five broader industrial categories based on 2-digit codes from the SIC 2003 classification, and which are the same as the ones used in the analysis of local and firm outcomes[10].

The great advantage of this data is the good quality of the earnings and hours information and the detailed information on the geographical location of both the workplace and the place of residence. Furthermore, its panel structure allows us to control for unobservable time invariant characteristics of the workers which might be correlated with our variable of interest. However, the survey contains information only on workers who are employed, so we are unable to observe unemployment spells. It also contains no information on household characteristics (for example housing tenure status, civil status, number of children) which might be relevant in shaping the response of labour market outcomes to transport policy. For these

[8]Different stratifications of ASHE define gross pay differently. We use the most recent definition: *total (gross) pay=basic pay + incentive pay + shift and premium payments + overtime pay + other pay*.

[9]They correspond to: 1 Managers and Senior Officials; 2 Professional Occupations; 3 Associate Professional and Technical Occupations; 4 Administrative and Secretarial Occupations; 5 Skilled Trades Occupations; 6 Personal Service Occupations; 7 Sales and Customer Service Occupations; 8 Process, Plant and Machine Operatives; 9 Elementary Occupations.

[10]They correspond to: 1 Manufacturing, 2 Construction, 3 Consumer services, 4 Producer services and 5 Other sectors.

reasons, in section 4.3 we use an alternative dataset, the LFS, to study other labour market outcomes for which information is not provided in ASHE.

4.2.2.2 Descriptive statistics

Table 4.1 displays the number of observations by year and gender. We restrict the sample to those individuals for which the home-ward and the work-ward are situated within 30 kilometres of any improvement undertaken during the period 2002-2008. We exclude those which are located too close to the improvements (1 km). We have around 240,000 observations, even after restricting the sample to locations close to the road improvements. As expected, we have more male workers than female workers. Table B.13 displays the number of observations by year and industrial sector. We observe that most of the workers are concentrated in the service sectors and a large fraction in the manufacturing sector. Between 2002 and 2008 there is a decrease on the importance of the manufacturing sector while the proportion of workers employed in the producer services sector increases over time.

YEAR	Males	Female	**Total**
2002	17,756	13,457	31,213
2003	19,451	15,355	34,806
2004	19,248	15,687	34,935
2005	20,100	17,012	37,112
2006	20,111	17,178	37,289
2007	16,846	14,945	31,791
2008	16,732	14,658	31,390
Total	130,244	108,292	238,536

Source: ONS. Observations within 30 kms of home and work wards.

Table 4.1: Number of observations by year and gender

Table B.14 displays the number of individuals we have in each of the panel definitions. If individuals are allowed to change location over time, μ_i, they appear an average of 3 times over the 7-year panel, and there are almost 80,000 individuals. When we define the fixed-effects as an individual-home-work combination, μ_{ihw}, we have over 114,000 fixed-effects, which appear an average of 2.09 times in the panel.

Tables B.15 and B.16 display the summary statistics of the variables used in the analysis. The tables show the mean and the standard deviation overall and within and between the two panels. Average basic weekly earnings are £373, while the average number of basic hours worked a week is slightly below 34. Total pay and hours are slightly over the basic figures, around £25 a week are earned based on incentive or overtime pay, and around 1.5 hours a week is devoted to overtime.

Average total hourly earnings are almost 50 pence higher than basic hourly earnings, which suggest that overtime and incentive pay per hours is higher than hourly basic pay, as we would expect. Table B.15 displays the within variation of accessibility in the two panels. When we only use the variation over time for individuals while they remain in a given work-home location the within variation is reduced notably. However, as we see in section 4.2.3.2 this variation is still sufficient to identify the parameters, at least for accessibility from workplace.

4.2.3 Results

4.2.3.1 Individual fixed effects

In this section we present the results for the estimates of equation (4.1). The main results of the tables (columns 1-7) are obtained using the 30 kilometre band both from work-ward and home-ward. Columns 8 and 9 replicate the results of column 7 for the 20 and 10 kilometre distance bands. All the specifications include sic-year dummies (5 wide sector definition explained above) to control for year-specific shocks in each sector. These dummies also control for nation-wide changes in economic conditions, for example CPI levels. The standard errors are robust to heteroskedasticity and clustered at the individual and year levels.

We report results for basic outcomes (table 4.2) and total outcomes (table 4.3). The difference between basic and total outcomes are incentive and overtime pay (in the case of wages) and over time (in the case of hours worked). The results are presented for weekly nominal wages, weekly hours worked and weekly hourly earnings. The relationship between the three outcomes is as follows:

$$wage_{ihwt} = \frac{wage_{ihwt}}{hours_{ihwt}} * hours_{ihwt} = hourly\_wage_{ihwt} * hours_{ihwt} \tag{4.4}$$

In both tables PANEL A displays the results for log weekly wages, PANEL B the results for log hourly earnings and PANEL C the results for log hours worked. Given the log-log specification, the parameters can be interpreted as elasticities. The rows at the bottom of the tables provide details on the estimated model.

Tables 4.2 and 4.3 have the same structure. Only the coefficients on log accessibility from work, log accessibility from home, log of travel time between home and work and the log of firms size are displayed. The coefficients on the dummies and controls are not provided in order to improve the clarity in the exposition of the results.

In column 1 we present the results obtained by OLS, where we only include sic-by-year dummies. The coefficients are estimated precisely due to the large number of observations (above 200,000). In all cases we find a positive and signific-

PANEL A: **Log of basic weekly pay**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.233***	0.072***	0.074***	0.072***	0.066***	0.045***	0.037***	0.014	0.015
from work ward	[0.007]	[0.008]	[0.008]	[0.009]	[0.010]	[0.010]	[0.010]	[0.013]	[0.033]
Log of accessibility	-0.087***	0.042***	0.042***	0.030***	0.031***	0.027***	0.021***	0.005	-0.031
from home ward	[0.006]	[0.009]	[0.009]	[0.009]	[0.009]	[0.010]	[0.008]	[0.010]	[0.037]
Log of travel time	0.111***	0.029***	0.029***	0.029***	0.029***	0.028***	0.024***	0.025***	0.019***
between work and home	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.002]	[0.003]	[0.005]
Log of firms							-0.002	-0.002	0.000
employment							[0.001]	[0.001]	[0.003]
Observations	245,883	221,066	221,066	221,066	221,066	221,066	216,324	123,468	30,392
Kleibergen-Paap F-stat			184,524	280,260	282,681	221,022	231,113	94,225	18,883
Individual no of clusters	81,330	56,513	56,513	56,513	56,513	56,513	55,723	32,086	8,077
Year no of clusters	7	7	7	7	7	7	7	7	7
PANEL B: **Log of basic hourly earnings**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.212***	0.047***	0.048***	0.044***	0.041***	0.023***	0.021***	0.013	0.011
from work ward	[0.005]	[0.004]	[0.004]	[0.004]	[0.004]	[0.006]	[0.006]	[0.012]	[0.016]
Log of accessibility	-0.073***	0.032***	0.032***	0.024***	0.024***	0.024***	0.022***	0.011	0.022*
from home ward	[0.005]	[0.005]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.007]	[0.012]
Log of travel time	0.072***	0.015***	0.015***	0.014***	0.014***	0.014***	0.012***	0.012***	0.006**
between work and home	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]	[0.002]	[0.003]
Log of firm's							0.006***	0.005***	0.003**
employment							[0.001]	[0.001]	[0.001]
Observations	244,166	219,347	219,347	219,347	219,347	219,347	214,628	122,513	30,132
Kleibergen-Paap F-stat			178,198	278,318	283,270	223,385	234,337	91,067	18,564
Individual no of clusters	80,984	56,165	56,165	56,165	56,165	56,165	55,378	31,889	8,023
Year no of clusters	7	7	7	7	7	7	7	7	7
PANEL C: **Log of basic weekly hours worked**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.023***	0.024***	0.025***	0.027***	0.024***	0.019**	0.013*	-0.001	0.005
from work ward	[0.003]	[0.005]	[0.005]	[0.007]	[0.008]	[0.007]	[0.007]	[0.012]	[0.032]
Log of accessibility	-0.014***	0.01	0.01	0.006	0.007	0.002	-0.001	-0.005	-0.048
from home ward	[0.003]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.006]	[0.010]	[0.031]
Log of travel time	0.040***	0.015***	0.015***	0.015***	0.015***	0.015***	0.012***	0.014***	0.012***
between work and home	[0.002]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]	[0.001]	[0.002]	[0.003]
Log of firms							-0.007***	-0.006***	-0.004*
employment							[0.001]	[0.001]	[0.002]
Observations	245,774	220,932	220,932	220,932	220,932	220,932	216,181	122,903	30,194
Kleibergen-Paap F-stat			173,610	275,394	281,027	221,303	232,776	92,211	18,604
Individual no of clusters	81,363	56,521	56,521	56,521	56,521	56,521	55,729	31,973	8,037
Year no of clusters	7	7	7	7	7	7	7	7	7
Distance band	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	20 kms	10 kms
Individual fixed effects		YES	YES	YES	YES	YES	YES	YES	YES
Instrumented			YES	YES	YES	YES	YES	YES	YES
Scheme dummies and trends				YES	YES	YES	YES	YES	YES
Distance to improvement dummies and trends					YES	YES	YES	YES	YES
Ward 2001 characteristic and trends						YES	YES	YES	YES
Personal and job characteristic and trends							YES	YES	YES

2-way clustering (individual and years) All specifications include sector-year dummies. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.2: Individual fixed effects results, basic outcomes – 2002-2008

PANEL A: **Log of total (gross) weekly pay**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.223***	0.069***	0.071***	0.068***	0.063***	0.040***	0.032***	0.016	-0.001
from work ward	[0.006]	[0.008]	[0.008]	[0.010]	[0.010]	[0.011]	[0.011]	[0.013]	[0.032]
Log of accessibility	-0.084***	0.041***	0.043***	0.029***	0.030***	0.025**	0.019**	0.003	-0.041
from home ward	[0.006]	[0.009]	[0.009]	[0.009]	[0.009]	[0.010]	[0.008]	[0.010]	[0.026]
Log of travel time	0.112***	0.027***	0.027***	0.027***	0.027***	0.026***	0.022***	0.022***	0.016***
between work and home	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.005]
Log of firms							0.003***	0.002	0.003
employment							[0.001]	[0.001]	[0.004]
Observations	245,968	221,171	221,171	221,171	221,171	221,171	216,427	123,482	30,404
Kleibergen-Paap F-stat			182,145	278,097	280,805	219,965	230,144	94,319	18,815
Individual no of clusters	81,350	56,553	56,553	56,553	56,553	56,553	55,764	32,096	8,083
Year no of clusters	7	7	7	7	7	7	7	7	7
PANEL B: **Log of total hourly earnings**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.206***	0.043***	0.043***	0.037***	0.033***	0.023**	0.022*	0.029	-0.025
from work ward	[0.006]	[0.007]	[0.007]	[0.009]	[0.010]	[0.011]	[0.012]	[0.027]	[0.026]
Log of accessibility	-0.071***	0.060***	0.063***	0.061***	0.062***	0.062***	0.061**	0.092*	0.000
from home ward	[0.005]	[0.017]	[0.018]	[0.021]	[0.021]	[0.024]	[0.025]	[0.047]	[0.027]
Log of travel time	0.072***	0.010***	0.010***	0.010***	0.010***	0.010***	0.008**	0.006	0.003
between work and home	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]
Log of firms							0.009***	0.007***	0.007***
employment							[0.001]	[0.001]	[0.003]
Observations	244,642	219,842	219,842	219,842	219,842	219,842	215,112	122,770	30,206
Kleibergen-Paap F-stat			181,165	276,945	281,034	218,998	230,035	92,354	18,674
Individual no of clusters	81,084	56,284	56,284	56,284	56,284	56,284	55,495	31,954	8,042
Year no of clusters	7	7	7	7	7	7	7	7	7
PANEL C: **Log of total weekly hours worked**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.019***	0.033***	0.035***	0.032***	0.031***	0.019	0.014	-0.009	0.042
from work ward	[0.003]	[0.009]	[0.009]	[0.011]	[0.011]	[0.014]	[0.015]	[0.030]	[0.036]
Log of accessibility	-0.015***	-0.021	-0.022	-0.04	-0.04	-0.048	-0.053*	-0.088*	-0.04
from home ward	[0.004]	[0.022]	[0.022]	[0.028]	[0.028]	[0.031]	[0.030]	[0.047]	[0.025]
Log of travel time	0.041***	0.018***	0.018***	0.018***	0.018***	0.017***	0.015***	0.021***	0.014***
between work and home	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.005]
Log of firms							-0.006***	-0.004**	-0.005
employment							[0.001]	[0.002]	[0.004]
Observations	246,234	221,388	221,388	221,388	221,388	221,388	216,629	123,183	30,267
Kleibergen-Paap F-stat			179,590	277,381	281,622	218,453	229,866	93,280	18,741
Individual no of clusters	81,466	56,620	56,620	56,620	56,620	56,620	55,826	32,039	8,054
Year no of clusters	7	7	7	7	7	7	7	7	7
Distance band	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	20 kms	10 kms
Individual fixed effects		YES	YES	YES	YES	YES	YES	YES	YES
Instrumented			YES	YES	YES	YES	YES	YES	YES
Scheme dummies and trends				YES	YES	YES	YES	YES	YES
Distance to improvement dummies and trends					YES	YES	YES	YES	YES
Ward 2001 characteristic and trends						YES	YES	YES	YES
Personal and job characteristic and trends							YES	YES	YES

2-way clustering (individual and years) All specifications include sector-year dummies. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.3: Individual fixed effects results, total outcomes – 2002-2008

ant relationship between accessibility from work and outcomes, negative relationships between the outcomes and accessibility from home and positive coefficients on travel time. The positive effect of accessibility from work on wages and hours could be due to agglomeration externalities or to the fact that professionals, who earn more and work full time, are concentrated in work locations where accessibility is higher. At the same time, workers in low paid jobs such as basic services could also be living in these locations and this could explain the negative coefficient of accessibility from home on earnings and hours. Finally, these results suggest that longer commutes are capitalised into higher wages and that workers who work longer hours are those who also commute longer. This explanation is in line with Manning (2003), who suggests that part-time workers do not find it worthwhile to commute for long as the relative time spent in commuting versus working is smaller than for full-time workers.

In column 2 we add individual fixed effects. We control for unobservable time-invariant characteristics of workers which might be correlated with labour market outcomes and accessibility. The coefficient for accessibility from workplace is reduced substantially, and that of accessibility from home becomes positive. This suggest that individuals which have unobservable characteristics correlated negatively with wages and hours are located in wards in which accessibility from home is higher. This is in line with the previous argument that less skilled workers live in denser areas in which they can access more jobs easily, as predicted for example by the spatial mismatch theory. When we introduce the fixed effects, the coefficient of accessibility from home on hours worked becomes very small and insignificant, both for basic and total hours (PANEL C of both tables).

Column 3 shows the results when we instrument A_{rt} using $\hat{A}_{iht}$. The instrument is very strong and the results remain very similar to those of column 2. In column 4 we add scheme dummies and trends, both from work and home. In column 5 we control for the distance to the closest improvement from both home and work, both in levels and in trends. In column 6 we add work-ward and home-ward 2001 characteristics and trends. The results remain very similar across specifications. As an aside, we find some evidence that workers in bigger firms earn more and work fewer hours.

In column 7 we present results from the most demanding specification. It includes fixed effects and all the controls, including personal and job characteristics both in levels and in trends (defined as initial characteristics, as explained above). For weekly and hourly earning we find positive and significant elasticities, both for basic and total outcomes. We find a weak effect of accessibility from work on basic hours worked and no effect on total hours. We find a marginally significant negative effect of accessibility from home on total hours worked. The estimates of the effect of accessibility from work are similar in magnitude to the estimates of the effect of market potential on wages provided for example by Combes et al. (2008a) or Mion & Naticchioni (2009).

In columns 8 and 9 we reduce the sample to the 20 and 10 kilometres bands around the wards of workplace and residence. Most of the effects of accessibility become insignificant, but the positive effect of commuting time on wages and hours remains significant. In fact, when we cluster the errors only at the individual level (results available on request) the coefficients of accessibility from work are significant, so the lack of significance could be due to the arbitrary contemporaneous cross-panel correlation, which inflates the standard errors.

As discussed above, these results could be driven by endogenous relocation of the individuals if they are spatially sorting in order to benefit from the accessibility changes or in order to reduce their commuting time. In the next section we investigate if using the variation of accessibility for individuals while they keep their location fixed has any impact on the estimates of the effects.

4.2.3.2 Individual-work-home fixed effects

In this section we present the results of the estimation of equation (4.2), i.e. including individual-work-home fixed effects μ_{ihw}. The results are displayed in table 4.4 for the basic outcomes and in table 4.5 for the total outcomes. As before, we report the results for weekly pay (PANEL A), hourly earnings (PANEL B) and hours worked (PANEL C). The specifications are the same as in section 4.2.3.1. The rows at the bottom of the tables provide details on the estimated model, from OLS to the model with all the controls.

Column 1 displays the results obtained by OLS which reproduce those of column 1 of tables 4.2 and 4.3. In column 2 we include the individual-work-home fixed-effects, so we only use the variation in accessibility and travel time for individuals across time, for a given location pair. The standard errors are bigger than those of the previous section, as we would expect due to the smaller amount of variation within the panels observed in tables B.15 and B.16. After introducing the fixed-effects, the coefficient of log accessibility from home becomes insignificant and quite imprecisely estimated. The coefficient of log accessibility from work remains positive and significant for basic weekly wages and basic hourly earnings, but becomes insignificant in the rest of the panels. The coefficient of log travel time is only significantly different from zero in PANEL B of table 4.4, and remains significant across specifications.

In column 3 we instrument the accessibility indices with the measures which only use the variation stemming from road improvements. The coefficient of the

PANEL A: **Log of basic weekly pay**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.233***	0.069*	0.314***	0.314***	0.342***	0.317***	0.328***	0.317***	0.151
from work ward	[0.013]	[0.037]	[0.116]	[0.112]	[0.112]	[0.113]	[0.116]	[0.122]	[0.152]
Log of accessibility	-0.087***	-0.007	-0.108	-0.107	-0.107	-0.076	-0.025	-0.064	0.09
from home ward	[0.013]	[0.032]	[0.091]	[0.094]	[0.097]	[0.095]	[0.092]	[0.095]	[0.136]
Log of travel time	0.111***	0.032	0.042	0.041	0.039	0.034	0.051	0.057*	0.079**
between work and home	[0.003]	[0.044]	[0.038]	[0.038]	[0.038]	[0.038]	[0.033]	[0.033]	[0.035]
Log of firms							0.001	-0.002	0.003
employment							[0.002]	[0.003]	[0.005]
Observations	245,883	183,091	183,091	183,091	183,091	183,091	180,105	105,168	26,935
Kleibergen-Paap F-stat			1,028	984	1,230	1,100	1,042	1,513	1,683
Work-ward no of clusters	3,794	3,557	3,557	3,557	3,557	3,557	3,551	2,226	755
Home-ward no of clusters	3,877	3,825	3,825	3,825	3,825	3,825	3,824	2,417	835
PANEL B: **Log of basic hourly earnings**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.212***	0.049*	0.235***	0.233***	0.244***	0.224***	0.228***	0.232***	0.164**
from work ward	[0.012]	[0.030]	[0.059]	[0.058]	[0.058]	[0.060]	[0.056]	[0.058]	[0.073]
Log of accessibility	-0.073***	0.000	-0.058	-0.058	-0.065	-0.052	-0.02	-0.038	0.049
from home ward	[0.011]	[0.023]	[0.064]	[0.064]	[0.065]	[0.064]	[0.063]	[0.065]	[0.080]
Log of travel time	0.072***	0.040*	0.049**	0.049**	0.049**	0.049**	0.061***	0.063***	0.067***
between work and home	[0.002]	[0.021]	[0.023]	[0.022]	[0.022]	[0.023]	[0.019]	[0.019]	[0.016]
Log of firm's							0.005***	0.003***	0.004
employment							[0.001]	[0.001]	[0.003]
Observations	244,166	181,664	181,664	181,664	181,664	181,664	178,691	104,364	26,713
Kleibergen-Paap F-stat			1,035	988	1,236	1,104	1,047	1,526	1,745
Work-ward no of clusters	3,794	3,554	3,554	3,554	3,554	3,554	3,548	2,223	752
Home-ward no of clusters	3,877	3,824	3,824	3,824	3,824	3,554	3,823	2,417	835
PANEL C: **Log of basic weekly hours worked**									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.023***	0.02	0.099	0.098	0.114*	0.111	0.121*	0.104	0.013
from work ward	[0.004]	[0.029]	[0.070]	[0.068]	[0.068]	[0.067]	[0.065]	[0.069]	[0.080]
Log of accessibility	-0.014***	0.01	-0.013	-0.012	-0.005	0.015	0.034	0.014	0.083
from home ward	[0.005]	[0.027]	[0.052]	[0.052]	[0.054]	[0.053]	[0.050]	[0.051]	[0.077]
Log of travel time	0.040***	-0.009	-0.005	-0.005	-0.007	-0.012	-0.008	-0.003	0.014
between work and home	[0.001]	[0.032]	[0.029]	[0.029]	[0.029]	[0.029]	[0.027]	[0.027]	[0.032]
Log of firms							-0.003*	-0.005**	0.000
employment							[0.002]	[0.002]	[0.004]
Observations	245,774	183,020	183,020	183,020	183,020	183,020	180,030	104,716	26,768
Kleibergen-Paap F-stat			1,023	979	1,223	1,092	1,035	1,527	1,749
Work-ward no of clusters	3,795	3,554	3,554	3,554	3,554	3,554	3,548	2,223	752
Home-ward no of clusters	3,878	3,827	3,827	3,827	3,827	3,827	3,826	2,417	835
Distance band	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	20 kms	10 kms
Individual-work-home fixed effects		YES	YES	YES	YES	YES	YES	YES	YES
Instrumented			YES	YES	YES	YES	YES	YES	YES
Scheme trends				YES	YES	YES	YES	YES	YES
Distance to improvement trends					YES	YES	YES	YES	YES
Ward 2001 characteristic trends						YES	YES	YES	YES
Personal and job characteristic trends							YES	YES	YES

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.4: Individual-home-work fixed effects results, total outcomes – 2002-2008

PANEL A: Log of total (gross) weekly pay									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.223***	0.061	0.240**	0.263**	0.285***	0.264**	0.302***	0.300**	0.124
from work ward	[0.013]	[0.039]	[0.110]	[0.105]	[0.103]	[0.106]	[0.111]	[0.120]	[0.144]
Log of accessibility	-0.084***	-0.035	-0.085	-0.059	-0.049	-0.025	-0.007	-0.047	0.054
from home ward	[0.013]	[0.035]	[0.082]	[0.085]	[0.090]	[0.087]	[0.086]	[0.090]	[0.117]
Log of travel time	0.112***	0.021	0.031	0.029	0.027	0.023	0.023	0.025	0.035
between work and home	[0.003]	[0.036]	[0.031]	[0.031]	[0.031]	[0.031]	[0.030]	[0.030]	[0.034]
Log of firms							0.004**	0.006	0.001
employment							[0.002]	[0.005]	[0.003]
Observations	245,968	183,188	183,188	183,188	183,188	183,188	180,200	105,199	26,947
Kleibergen-Paap F-stat			1,031	985	1,229	1,099	1,041	1,513	1,681
Work-ward no of clusters	3,794	3,557	3,557	3,557	3,557	3,557	3,551	2,226	755
Home-ward no of clusters	3,877	3,826	3,826	3,826	3,826	3,826	3,825	2,417	835
PANEL B: Log of total hourly earnings									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.206***	0.018	-0.021	-0.016	-0.014	-0.043	-0.037	-0.062	-0.252
from work ward	[0.012]	[0.047]	[0.139]	[0.137]	[0.138]	[0.139]	[0.143]	[0.153]	[0.214]
Log of accessibility	-0.071***	-0.009	0.154	0.157	0.171	0.19	0.215	0.239	0.445
from home ward	[0.011]	[0.047]	[0.191]	[0.191]	[0.192]	[0.192]	[0.198]	[0.209]	[0.310]
Log of travel time	0.072***	-0.025	-0.009	-0.012	-0.013	-0.015	-0.012	-0.006	0.061***
between work and home	[0.003]	[0.057]	[0.055]	[0.055]	[0.055]	[0.055]	[0.056]	[0.056]	[0.023]
Log of firms							0.005**	0.005	0.001
employment							[0.002]	[0.003]	[0.003]
Observations	244,642	182,082	182,082	182,082	182,082	182,082	179,101	104,593	26,775
Kleibergen-Paap F-stat			1,041	991	1,237	1,105	1,046	1,523	1,732
Work-ward no of clusters	3,794	3,554	3,554	3,554	3,554	3,554	3,548	2,223	752
Home-ward no of clusters	3,877	3,825	3,825	3,825	3,825	3,825	3,824	2,417	835
PANEL C: Log of total weekly hours worked									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.019***	0.043	0.281*	0.314**	0.341**	0.338**	0.375**	0.388**	0.423*
from work ward	[0.005]	[0.054]	[0.148]	[0.147]	[0.147]	[0.149]	[0.156]	[0.166]	[0.242]
Log of accessibility	-0.015***	-0.056	-0.223	-0.194	-0.201	-0.195	-0.204	-0.256	-0.385
from home ward	[0.005]	[0.063]	[0.217]	[0.217]	[0.218]	[0.218]	[0.224]	[0.234]	[0.361]
Log of travel time	0.041***	0.037	0.037	0.036	0.035	0.033	0.028	0.023	-0.04
between work and home	[0.002]	[0.070]	[0.062]	[0.062]	[0.062]	[0.062]	[0.062]	[0.061]	[0.036]
Log of firms							-0.001	0.003	-0.001
employment							[0.002]	[0.004]	[0.003]
Observations	246,234	183,388	183,388	183,388	183,388	183,388	180,392	104,946	26,828
Kleibergen-Paap F-stat			1,028	982	1,225	1,094	1,036	1,525	1,739
Work-ward no of clusters	3,795	3,554	3,554	3,554	3,554	3,554	3,548	2,223	752
Home-ward no of clusters	3,878	3,827	3,827	3,827	3,827	3,827	3,826	2,417	835
Distance band	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	20 kms	10 kms
Individual-work-home fixed effects		YES	YES	YES	YES	YES	YES	YES	YES
Instrumented			YES	YES	YES	YES	YES	YES	YES
Scheme trends				YES	YES	YES	YES	YES	YES
Distance to improvement trends					YES	YES	YES	YES	YES
Ward 2001 characteristic trends						YES	YES	YES	YES
Personal and job characteristic trends							YES	YES	YES

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.5: Individual-home-work fixed effects results, total outcomes – 2002-2008

effect of log accessibility from work on wages and on hourly earnings increases substantially as compared to those of column 2. For accessibility from home the coefficients remain insignificant as in the rest of the specifications.

The difference in size of the coefficients obtained in column 2 and those obtained in column 3 is quite substantial. Specifically, the IV estimates of column 3 are one order of magnitude larger than the FE ones. It is possible that part of the difference is due to attenuation bias caused by measurement error in the accessibility indices. As explained in chapter 2, the calculation of travel times used in the computation of A_{rt} is approximate. This could be introducing some measurement error causing an attenuation bias that the instrument could be helping to reduce. It is nevertheless more likely that most of the difference in the size of both coefficients is due to the existence of unobservable trends which are correlated both with that accessibility from work and labour market outcomes. If employment is concentrating in areas in which wages are growing slower, which would be sensible because is cheaper to hire employees, the fixed-effects estimator would be downward biased.

Even after adding all the trends and controls in column 7, we find positive and highly significant effects of accessibility from work on basic weekly wages. In contrast, we find very weak effects on basic hours worked; column 7 of PANEL C of table 4.4 displays a positive coefficient of the effect of accessibility from work on hours, but it is only significant at 10%. We find positive and significant effects of accessibility from work on basic hourly pay. Given that basic hourly wages are defined as weekly basic pay over weekly basic hours and basic hours worked seem not to be affected by changes in accessibility, the results on basic hourly earnings could be driven by increases in the "numerator" of the ratio. These results is robust when we restrict the sample to individuals located in the 20-kilometre distance bands, and the coefficient remains significant only for basic hourly earnings when we narrow the sample to the 10-kilometre band. Note however that the number of observations used in column 9 is drastically reduced.

In the case of total labour market outcomes, reported in table 4.5, the picture is slightly different. We find positive and significant effects of accessibility from work on weekly total wage. The coefficients are similar to those of PANEL A of table 4.4. This would suggest that most of the effect on total pay can be attributable to the effect on basic pay. However, for total hours worked (PANEL C) we also find significant and positive effects of accessibility from work, while we did not find any effects on basic hours worked. This would suggest that the hours adjustment is working through overtime or through changes from part time to full time status. The estimates are very robust to the inclusion of controls and to the narrowing of the distance bands. As both wages and hours are adjusting, we do not find any effect on hourly earnings (PANEL B).

4.2.3.3 Interpretation of the results

Table 4.6 summarises our main results to facilitate the discussion of the interpretation of the size and sign of the coefficients. It reproduces some of the results of tables 4.2–4.3 and tables 4.4–4.5. For the 30-kilometre distance band table 4.6 displays the coefficients obtained using OLS and using the two sorts of fixed-effects (individual and individual-work-home), with and without controls. Specification in column 7 shows our main findings, the robustness of which is discussed below in section 4.2.3.4.

PANEL A: **Log of basic weekly pay**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of accessibility	0.233***	0.072***	0.074***	0.037***	0.069*	0.314***	0.328***
from work ward	[0.007]	[0.008]	[0.008]	[0.010]	[0.037]	[0.116]	[0.116]
Log of accessibility	-0.087***	0.042***	0.042***	0.021***	-0.007	-0.108	-0.025
from home ward	[0.006]	[0.009]	[0.009]	[0.008]	[0.032]	[0.091]	[0.092]
Observations	245,883	221,066	221,066	216,324	183,091	183,091	180,105
Kleibergen-Paap F-stat			184,524	231,113		1,028	1,042
PANEL B: **Log of total weekly hours worked**							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of accessibility	0.019***	0.033***	0.035***	0.014	0.043	0.281*	0.375**
from work ward	[0.003]	[0.009]	[0.009]	[0.015]	[0.054]	[0.148]	[0.156]
Log of accessibility	-0.015***	-0.021	-0.022	-0.053*	-0.056	-0.223	-0.204
from home ward	[0.004]	[0.022]	[0.022]	[0.030]	[0.063]	[0.217]	[0.224]
Observations	246,234	221,388	221,388	216,629	183,388	183,388	180,392
Kleibergen-Paap F-stat			179,590	229,866		1,028	1,036
Distance band	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms	30 kms
Fixed effects		Indiv	Indiv	Indiv	Ind-w-h	Ind-w-h	Ind-w-h
Instrumented			YES	YES		YES	YES
All controls and trends				YES			YES

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.6: Summary of main results – weekly basic wages and total hours worked

We can draw two main conclusions from the results of sections 4.2.3.1 and 4.2.3.2. The first one is that labour market outcomes are affected by accessibility from work, but only the outcomes which are flexible, i.e. wages and overtime, while basic hours do not adjust. If accessibility from work is capitalising into higher wages, for example because of agglomeration externalities, conditional on commuting costs, workers could find it more worthwhile to work more hours[11]. If the number of basic work hours per week is fixed by contract, one way they can benefit from this wage increase is by working overtime. Another possible explanation is that workers are switching from part time to full time jobs, but we do not directly test this in our data.

The second one is that, for the correct identification and interpretation of the

[11]Unless they have a strong preference for leisure, which our results do not suggest.

effects, it is key to account for workers spatial sorting. Comparison of the coefficients obtained in tables 4.2–4.3 and tables 4.4–4.5 informs us about the effect that controlling for sorting has on the estimates.

In tables 4.2–4.3, when we use the variation of accessibility which stems from workers relocating across wards, we find positive effects of accessibility both from work and from home. However, once we control for sorting, the results of tables 4.4–4.5 suggest that that improving access to jobs from home does not have any impact on labour market outcomes of employed workers whereas accessibility from work has substantial effects on wages and hours. To obtain these results we exploit the variation in accessibility over time for workers who remain in the same work-ward and live in the same home-ward. This finding suggests that the effects of accessibility from home on workers are partially driven by residential sorting. Workers might be moving into places in which accessibility is growing, because better connections makes these residential locations more desirable or because they can afford better housing in places that are further away from job centres and that are now more accessible.

Once individuals have chosen their residential location, access to more jobs from home does not seem to have an effect on their wages or hours, conditional on commuting time. The theoretical channel through which better accessibility from home would have an effect on wages and hours once we are controlling for sorting is unclear. The spatial mismatch theory predicts positive effects of better access to jobs from home on labour market outcomes, especially on the probability of becoming employed or on the length of the unemployment spells. However, from the results obtained using our sample, in which all the individuals are already working, we do not find any evidence in favour of the spatial mismatch hypothesis. This could be due to households having already optimised their residential location in order to have access to better jobs and shorter commutes so further increases in accessibility from home do not have any impact on their wages or hours worked.

The positive effect of accessibility from work on wages and hours, given that in tables 4.4–4.5 we are controlling for commuting travel time, could be working through spatial competition or agglomeration externalities. If firms in which the workers are employed can access a larger pool of workers, then employees might behave more competitively and work harder (specially high skilled workers, as suggested by Rosenthal & Strange, 2008). Also, workers could become more productive in these denser and better connected areas, as has been suggested and empirically verified in the agglomeration economies literature (see Duranton & Puga, 2004, for a review) or in the new economic geography literature (Krugman & Venables, 1995; Redding & Venables, 2004, for example).

In the context of our results, within the different mechanisms which affect industrial concentration and agglomeration[12], labour market pooling is possibly the most relevant channel through which agglomeration could be impacting wages and hours (for example Rosenthal & Strange, 2001, find that this mechanism is the most relevant in determining industrial concentration). Even if we do not explicitly use a standard measure of labour market pooling which captures "the use of similar workers within an industry" (see Rosenthal & Strange, 2001; Overman & Puga, 2010; Jofré-Monseny et al., 2011, for examples), better accessibility to employment could potentially be improving the quality of job-workers matching (see Andersson et al., 2007, for some empirical evidence) or the search behaviour of the individuals (as suggested by Di Addario, 2011).

Concerning the size of our coefficients, our coefficient of the effect of accessibility from work on weekly wages (column 7 of PANEL A–table 4.6) is larger (0.328) than previous estimates of the effect of Harris-type "market potential" measures on nominal wages. For example, the elasticities provided by Mion & Naticchioni (2005); Fingleton (2006); Combes et al. (2008a); Amiti & Cameron (2007) and Graham & Melo (2009) are between 0.02 and 0.2.[13]

First of all, our measure of accessibility is not directly comparable to standard "market access" measures used in the literature. The papers mentioned above use a general market potential definition (Harris, 1954), which is similar in structure to ours but uses geodesic time invariant distance as the measure of proximity between locations. In their case, the identified effect of market access on wages stems from changes in the spatial distribution of employment/income over time, because distances between locations are kept fixed. Instead, we focus on a different channel of variation in the accessibility indices, which is road construction. In the main results of table 4.6, we instrument A_{rt} with $\hat{A}_{rt}$. In this way, we use a different source of variation for identification for the effects of accessibility from the authors above. Proximity between wards changes due to changes in optimal travel times between locations induced by road construction.

Spatial sorting could be explaining part of the difference in size between our estimates and the coefficients found in previous evidence. The identification strategy we follow substantially helps to reduce the bias caused by the spatial sorting of workers. As explained above, we control for sorting at the individual level (both residential and job sorting) by using the individual-home-work fixed effects. This way, for the identification of the effects we exploit the changes in accessibility over

[12] Although the different mechanisms are empirically hard to distinguish due to the "Marshallian equivalence", as noted by Duranton & Puga (2004) and Overman & Puga (2010).

[13] Other authors, like Redding & Venables (2004); H. Hanson (2005); Head & Mayer (2006) or Hering & Poncet (2010) have used a measure of market access derived from a NEG model. Their estimates of the elasticity of nominal wages/GDP per capita with respect of market access are between 0.1–0.3.

time for an individual while staying a given location combination. Moreover, as outlined above, we believe the instrumental variables strategy helps to eliminate the bias induced by spatial relocation of workers across space which might be driven by the changes in accessibility. In other words, we tackle the sorting of workers at the ward level[14].

In fact in our results, the estimates that overlook sorting at the individual level and the endogenous spatial relocation of ward employment (columns 2 and 5 in table 4.6), are similar in magnitude to previous estimates of the literature. Thus, the estimated elasticities of the effect of accessibility from work on weekly basic and total wages in PANEL A of tables 4.2–4.3 are 0.037 and 0.032 (in column 7, our preferred specification). Furthermore, even after controlling for individual sorting, when we do not control for spatial relocation of employment, e.g. in column 5 of table 4.6, the estimated elasticity is 0.069, which is again in the same order of magnitude than previous findings.

If we compare the coefficients in column 5 to those in columns 6 and 7 of table 4.6, once controlling for individuals sorting, when we instrument for spatial relocation of employment the estimated coefficients increase substantially. As explained above in section 4.2.3.2, the downward bias in the fixed-effect estimates of column 5 could be due to employment concentrating around areas in which wages are growing slower. Conditional on other wards characteristics, firms would prefer to employ workers in wards in which the price of labour is growing slower. The instrument could also be reducing attenuation bias caused by measurement error in accessibility[15]. Compared to column 3, attenuation bias could be amplified in column 5 as the number individual-work-home fixed effects is much larger than before (see table B.14). Besides, the precision of the estimates of columns 6 and 7 is lower because the amount of variation in accessibility for individuals that keep their location pair fixed is quite small (see table B.15). Compared to columns 3 and 5, when we control for individual sorting and instrument in columns 6 and 7, the standard errors increase substantially, and it does so the confidence interval of the estimates. As a matter of fact, the confidence intervals of the coefficients of column 5 and 6 overlap and are not statistically different at 5% confidence level.

[14]In their investigation of the determinants of individual wages using a large sample of French workers, Combes et al. (2008a) find that differences in the skill composition of the labour force account for 40 to 50% of aggregate spatial wage disparities. They conclude that workers with better labour market characteristics tend to agglomerate in the larger, denser and more skilled local labour market. The intuition is that sorting of workers across space has an effect on individual wages because more skilled workers sort in specific areas. Our approach is different as we explicitly control for the residential and job sorting of the individuals. Furthermore, we focus on accessibility changes stemming from transport improvements and do not use the variation coming from spatial changes in employment. This should help to reduce sorting issues in the same line as these authors.

[15]Which is possible as annual changes in ward employment calculated using the BSD are likely to be slightly noisy.

Finally, in table 4.7 we provide some interpretation of the size of the estimates within the context of our analysis. We report the mean change of log accessibility $\hat{A}_{rt}$ between 2002 and 2008 (growth rate). We focus on $\hat{A}_{rt}$ because our analysis focuses on changes in accessibility stemming from road construction. In the top panel we display the growth in accessibility for wards situated within 10-20-30 kilometres of any road improvements (All projects), and in wards within the same distance bands from three specific projects undertaken in 2000, 2003 and 2006, the length of which is provided too. In the first place, from this table we see that the changes in accessibility around an improvement undertaken before our period of analysis (2000) is very small but not zero. This illustrates the fact that changes in accessibility, even when using only the variation due to road construction, affect the whole geography. A change in the network impacts travel times between wards depending on the relative position of the wards within the network and with respect to the rest of the road improvements. For the schemes carried out in 2003 and 2006 the changes in accessibility are more substantial.

	DISTANCE BAND		
PANEL A: **Change in log accessibility $\hat{A}_{rt}$ between 2002-2008**	*10 kms*	*20 kms*	*30 kms*
All projects	2.440%	1.120%	0.790%
2000 - M66 Denton - Middleton (15.3 kms)	0.011%	0.011%	0.030%
2003 - A5 Nesscliffe Bypass (21.48 kms)	3.622%	2.026%	1.292%
2006 - A1(M) Ferrybridge to Hook Moor (19.2 kms)	3.131%	2.108%	1.272%
PANEL B: **Effect on basic weekly wages (coefficient=0.328)**	*10 kms*	*20 kms*	*30 kms*
All projects	0.800%	0.367%	0.259%
2000 - M66 Denton - Middleton (15.3 kms)	0.004%	0.003%	0.010%
2003 - A5 Nesscliffe Bypass (21.48 kms)	1.188%	0.665%	0.424%
2006 - A1(M) Ferrybridge to Hook Moor (19.2 kms)	1.027%	0.691%	0.417%
PANEL C: **Effect on total weekly hours (coefficient=0.375)**	*10 kms*	*20 kms*	*30 kms*
All projects	0.915%	0.420%	0.296%
2000 - M66 Denton - Middleton (15.3 kms)	0.004%	0.004%	0.011%
2003 - A5 Nesscliffe Bypass (21.48 kms)	1.358%	0.760%	0.484%
2006 - A1(M) Ferrybridge to Hook Moor (19.2 kms)	1.174%	0.791%	0.477%

Sources: ONS, DfT and authors own calculations.

Table 4.7: Effect of accessibility growth on wages and hours – 2002-2008

Additionally, this table helps us to evaluate the size of the coefficients taking into account the average growth in accessibility within the context of our empirical exercise. We focus on the results obtained for basic weekly pay (which are very similar to the results for gross pay) and total weekly hours worked. Our benchmark estimated elasticities are 0.328 for basic weekly pay and 0.375 for total hours, as presented in column 7 of table 4.6. The average growth in accessibility for Great Britain between 2002 and 2008 (approximated by the change in logs), as displayed in table 2.5 of chapter 2, is 0.3%. The product of the elasticity and the growth in accessibility informs us of the relative growth in weekly wages and hours which can be attributed to the changes in accessibility (due to road construction). For the whole Great Britain, our estimates predict a growth in weekly wages between 2002-2008 due to changes in accessibility of 0.098% , and in total hours worked of 0.113%.

When we focus in wards closer to the improvements we find larger impacts, as the growth in accessibility is bigger. In the bottom panels of table 4.7 we provide the predicted effects on the 10-20-30 kilometres distance bands. For wards within 30 kilometres of improvements, approximately 0.26% of the growth of weekly nominal wages and almost 0.3% of the growth in hours worked in Great Britain during the period can be attributed to the changes in accessibility that stem from road improvements (right column). These impacts increase the closer we get to the improvements (10 and 20 kilometres distance bands). Around specific improvements these effects are larger. For example in within 10 kilometres of the "A5 Nesscliffe Bypass" project, 1.2% of the growth in wages and 1.36% of the growth in total hours are attributable to changes in accessibility.

4.2.3.4 Robustness

In this section we test the robustness of the previous findings. We use the model of column 7 of tables 4.4 and 4.5, which is the most demanding specification. We focus on the results on basic weekly wages (which are very similar to those for total weekly wages) and on total hours worked.

In columns 1 to 3 of table 4.8 we use different measures of the economic size of the wards in the definition of accessibility indices. Column 1 replicates baseline results. Column 2 uses address counts as the measure of economic size[16] and column 3 uses number of plants (local unit counts). The instruments were calculated using 2001 employment, address counts and local unit counts respectively. The results are very similar to those of section 4.2.3.2. In columns 4 to 6 we compute accessibility using employment as economic mass but we use different instruments. Column 4 calculates $\hat{A}_{iht}$ using 1997 employment, and columns 5 and 6 use 2001 address counts and local unit counts as economic mass. The instruments of columns 4 and 5 are weaker than that of column 1, but again, the results are very robust.

In table 4.9 we define the accessibility measure using alternative cost functions. Columns 1 and 4 and 2 and 5 use the "inverse cost weights" function ($a\left(c_{rjt}\right) = c_{rjt}^{-\alpha}$) but with distance decays $\alpha = 0.5$ (flatter) and $\alpha = 1.5$ (steeper). Columns 3 and 6 use the "exponential weights" function ($a\left(c_{rjt}\right) = \exp\left(\alpha c_{rjt}\right)$) with distance decay $\alpha = 0.2$. The elasticities are larger than before but still positive and largely significant.

[16] Ideally we would have used population, but data on population at the ward level on a yearly basis is not available.

PANEL A: **Log of basic weekly pay**						
	(1)	(2)	(3)	(4)	(5)	(6)
Log of accessibility	0.328***	0.298***	0.312***	0.325***	0.321***	0.324***
from work ward	[0.116]	[0.112]	[0.116]	[0.116]	[0.119]	[0.118]
Log of accessibility	-0.025	-0.051	-0.048	-0.028	-0.031	-0.033
from home ward	[0.092]	[0.090]	[0.093]	[0.092]	[0.093]	[0.093]
Log of travel time	0.051	0.043	0.046	0.05	0.049	0.049
between work and home	[0.033]	[0.034]	[0.033]	[0.033]	[0.033]	[0.033]
Observations	180,105	180,105	180,105	180,105	180,105	180,105
Kleibergen-Paap F-stat	1,042	1,290	3,428	1,037	814	690
Work-ward no of clusters	3,551	3,551	3,551	3,551	3,551	3,551
Home-ward no of clusters	3,824	3,824	3,824	3,824	3,824	3,824
PANEL B: **Log of total weekly hours worked**						
	(1)	(2)	(3)	(4)	(5)	(6)
Log of accessibility	0.375**	0.364**	0.386**	0.372**	0.380**	0.396**
from work ward	[0.156]	[0.162]	[0.162]	[0.154]	[0.163]	[0.163]
Log of accessibility	-0.204	-0.232	-0.243	-0.204	-0.22	-0.23
from home ward	[0.224]	[0.231]	[0.240]	[0.221]	[0.239]	[0.240]
Log of travel time	0.028	0.025	0.024	0.028	0.027	0.027
between work and home	[0.062]	[0.063]	[0.061]	[0.062]	[0.061]	[0.061]
Observations	180,392	180,392	180,392	180,392	180,392	180,392
Kleibergen-Paap F-stat	1,036	1,301	3,401	1,032	809	684
Work-ward no of clusters	3,548	3,548	3,548	3,548	3,548	3,548
Home-ward no of clusters	3,826	3,826	3,826	3,826	3,826	3,826
Accessibility	Empl	Addrct	LUsct	Empl	Empl	Empl
Instrument	Empl01	Addrct01	LUsct01	Empl97	Addrct01	LUsct01

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.8: Robustness: different economic sizes and different instruments – 2002-2008

	PANEL A: **Log of basic weekly pay**			PANEL B: **Log of total weekly hours worked**		
	(1)	(2)	(3)	(4)	(5)	(6)
Log of accessibility	0.723***	0.874**	0.874**	0.874**	0.167**	1.685**
from work ward	[0.260]	[0.355]	[0.355]	[0.355]	[0.067]	[0.668]
Log of accessibility	-0.044	-0.433	-0.433	-0.433	-0.01	0.12
from home ward	[0.197]	[0.470]	[0.470]	[0.470]	[0.049]	[0.437]
Log of travel time	0.052	0.032	0.032	0.032	0.02	0.034
between work and home	[0.032]	[0.062]	[0.062]	[0.062]	[0.030]	[0.030]
Observations	180,105	180,392	180,392	180,392	180,200	180,200
Kleibergen-Paap F-stat	366	362	362	362	1,920	58
Work-ward no of clusters	3,551	3,548	3,548	3,548	3,551	3,551
Home-ward no of clusters	3,824	3,826	3,826	3,826	3,825	3,825
Cost function	Inverse	Inverse	Exponent	Inverse	Inverse	Exponent
Decay	0.5	1.5	0.2	0.5	1.5	0.2

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.9: Robustness: different costs functions and distance decays – 2002-2008

In table 4.10 we test the robustness of the results to the inclusion or exclusion of some variables. In column 1 we drop commuting time and in column 2 we introduce travel time in levels instead of in logs. The results remain very similar, the only different is that travel time is now significant and positive in column 2 contrary to insignificant in column 7 of table 4.4. In column 3 we add a dummy which indicates if the individual is changing jobs (that we identify using the enterprise reference number) within the ward of work. The estimated elasticities remain virtually unchanged. This is also the case in column 4 when we introduce a dummy if the worker changes home (identified using the postcode) within the home-ward. Finally, in column 5 we drop those individuals who live and work in the same ward. Once more, we obtain very similar coefficients to those of section 4.2.3.2.

In table 4.11 we test the robustness of the results to the inclusion or exclusion of some specific individual groups. In column 1 we exclude all individuals whose wage setting is subject to collective agreement. The number of observations decreases substantially and the estimates become imprecise, and only weekly significant for total hours worked. In column 2 we include workers from the public sector, which are excluded in the main results. The results are very similar to before but the elasticity of accessibility from work is slightly larger. In column 3, when we obtain the results only using individuals working in the public sector, the coefficient of accessibility from work on wages and hours are both significant (but weaker) and positive, and the size is bigger than in our baseline results.

In columns 4 to 6 we test the robustness of the results to the exclusion of London. In column 4 we exclude individuals working or living in London (inner and outer), in column 5 only those living in London and in column 6 only working in the capital. The results are very similar to the main estimates. Finally, in column 7 we restrict the estimation to those individuals which do not change home-ward location while observed in the panel, i.e. those for which μ_i is equal to μ_{ihw}. The coefficient of the effect of accessibility from work on wages is very similar to the baseline, but the effect on hours is larger. These individuals might not be moving work or home for some unobserved reasons, for example they are home owners or social renters, they have restricted mobility due to family reasons, etc. Therefore, if they cannot relocate to take advantage of the changes in accessibility, they might adjust by working longer hours or switching to full time employment. This could explain the larger estimated effect on hours worked.

PANEL A: **Log of basic weekly pay**					
	(1)	(2)	(3)	(4)	(5)
Log of accessibility	0.286***	0.316***	0.327***	0.328***	0.326***
from work ward	[0.110]	[0.116]	[0.116]	[0.116]	[0.119]
Log of accessibility	-0.04	-0.011	-0.024	-0.026	-0.019
from home ward	[0.097]	[0.094]	[0.092]	[0.092]	[0.093]
Log of travel time			0.05	0.051	0.076**
between work and home			[0.033]	[0.032]	[0.033]
Travel time		0.236**			
between work and home		[0.102]			
Individual changes			0.002		
job within work ward			[0.003]		
Individual changes				0.006*	
house within home ward				[0.003]	
Observations	185,818	180,105	180,105	180,105	158,609
Kleibergen-Paap F-stat	1,104	1,076	1,042	1,042	787
Work-ward no of clusters	3,556	3,551	3,551	3,551	3,353
Home-ward no of clusters	3,830	3,824	3,824	3,824	3,785
PANEL B: **Log of total weekly hours worked**					
	(1)	(2)	(3)	(4)	(5)
Log of accessibility	0.355**	0.373**	0.375**	0.375**	0.350**
from work ward	[0.140]	[0.153]	[0.156]	[0.156]	[0.147]
Log of accessibility	-0.209	-0.167	-0.204	-0.203	-0.21
from home ward	[0.229]	[0.228]	[0.224]	[0.224]	[0.233]
Log of travel time			0.028	0.028	0.072
between work and home			[0.062]	[0.062]	[0.064]
Travel time		0.284			
between work and home		[0.230]			
Individual changes			-0.001		
job within work-ward			[0.005]		
Individual changes				-0.001	
house within home-ward				[0.006]	
Observations	186,209	180,392	180,392	180,392	158,998
Kleibergen-Paap F-stat	1,099	1,070	1,036	1,035	778
Work-ward no of clusters	3,553	3,548	3,548	3,548	3,351
Home-ward no of clusters	3,834	3,826	3,826	3,826	3,790
Specification	No ltrvt	Level trvt	Chg jobs	Chg houses	Diff HW

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.10: Robustness: changes in the specification – 2002-2008

PANEL A: Log of basic weekly pay

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of accessibility	0.256	0.337***	0.501*	0.301***	0.306***	0.308***	0.341***
from work ward	[0.211]	[0.110]	[0.297]	[0.108]	[0.108]	[0.108]	[0.117]
Log of accessibility	0.034	-0.051	-0.135	-0.059	-0.06	-0.058	0.068
from home ward	[0.143]	[0.083]	[0.150]	[0.082]	[0.082]	[0.081]	[0.105]
Log of travel time	0.111*	0.022	-0.042	0.022	0.022	0.023	0.035
between work and home	[0.058]	[0.021]	[0.028]	[0.021]	[0.021]	[0.021]	[0.031]
Observations	96,119	254,294	73,488	205,991	208,091	213,822	114,919
Kleibergen-Paap F-stat	1,132	952	352	1,259	1,253	1,271	987
Work-ward no of clusters	3,315	3,710	2,703	3,192	3,194	3,549	3,350
Home-ward no of clusters	3,745	3,847	3,511	3,333	3,629	3,337	3,724

PANEL B: Log of total weekly hours worked

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log of accessibility	0.261*	0.407***	0.645**	0.387***	0.389***	0.385***	0.698***
from work ward	[0.142]	[0.136]	[0.254]	[0.135]	[0.135]	[0.136]	[0.218]
Log of accessibility	0.064	-0.17	-0.029	-0.177	-0.177	-0.181	-0.237
from home ward	[0.119]	[0.179]	[0.138]	[0.177]	[0.177]	[0.176]	[0.282]
Log of travel time	0.019	0.008	-0.038	0.005	0.005	0.005	0.033
between work and home	[0.049]	[0.046]	[0.041]	[0.046]	[0.046]	[0.046]	[0.074]
Observations	96,485	254,619	73,537	205,655	207,768	213,739	115,172
Kleibergen-Paap F-stat	1,123	949	348	1,259	1,254	1,272	990
Work-ward no of clusters	3,310	3,707	2,696	3,190	3,192	3,546	3,346
Home-ward no of clusters	3,748	3,848	3,510	3,332	3,627	3,336	3,728
Specification	No coll	Priv-Publ	Public	No London	No hLond	No wLond	No move

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.11: Robustness: different groups– 2002-2008

4.2.3.5 Heterogeneous effects

In this section we estimate the effects of accessibility from work and home for different sub-groups to both check the robustness of the results for specific groups of workers and also to be able to identify any heterogeneous treatment effects[17].

PANEL A: Log of basic weekly pay					
	(1)	(2)	(3)	(4)	(5)
Log of accessibility	0.419***	0.385***	0.266	0.387**	0.286***
from work ward	[0.106]	[0.112]	[0.169]	[0.162]	[0.109]
Log of accessibility	0.015	0.015	-0.11	0.007	-0.082
from home ward	[0.116]	[0.081]	[0.197]	[0.160]	[0.096]
Log of travel time	0.054	0.089***	-0.004	0.111**	0.025
between work and home	[0.034]	[0.026]	[0.052]	[0.048]	[0.039]
Observations	130,511	99,509	80,596	64,520	112,729
Kleibergen-Paap F-stat	747	853	1,033	1,491	533
Work-ward no of clusters	3,382	3,169	3,107	3,220	2,884
Home-ward no of clusters	3,759	3,710	3,645	3,618	3,701
PANEL B: Log of total weekly hours worked					
	(1)	(2)	(3)	(4)	(5)
Log of accessibility	0.289**	0.378	0.355***	0.253	0.524**
from work ward	[0.134]	[0.249]	[0.123]	[0.172]	[0.235]
Log of accessibility	-0.041	-0.218	-0.166	-0.042	-0.298
from home ward	[0.107]	[0.340]	[0.150]	[0.140]	[0.365]
Log of travel time	0.045	0.103	-0.054	0.136	-0.009
between work and home	[0.083]	[0.106]	[0.047]	[0.134]	[0.042]
Observations	130,730	99,897	80,495	64,389	113,129
Kleibergen-Paap F-stat	744	836	1,043	1,505	527
Work-ward no of clusters	3,382	3,163	3,104	3,216	2,884
Home-ward no of clusters	3,762	3,717	3,644	3,622	3,709
Group	Aged 20-50	Males	Females	Firms0-150	Firms+150

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.12: Robustness: by age, gender and firm size– 2002-2008

In table 4.12 we estimate the effects by age, by gender and by firm size. When we focus on the "primary working-age" workers, i.e. those aged 20 to 50, we find stronger effects of accessibility from work on wages and very similar effects on total hours worked. In columns 2 and 3 we split the results by gender. These results suggest that the effects of accessibility on wages is only applicable to men, while women are responding to increases in accessibility from workplace by increasing the number of hours worked (maybe switching from part time to full time jobs).

In columns 4 and 5 of table 4.12 we run the regressions for individuals working in firms with 150 employees or fewer and firms with more than 150 employees. The employment figure corresponds to the firm in which the worker is employed, not to the specific plant in which he is located. The effect of accessibility from work on weekly wages remains significant for both groups, and it is slightly higher for

[17] The results in this section are preliminary and need further work.

smaller firms. In bigger plants we also see an adjustment via hours worked, and the coefficient is larger than the baseline results. This could be due to bigger firms being more flexible with respect to working hours or workers being able to switch to full time jobs by remaining in the same firm in bigger enterprises.

PANEL A: **Log of basic weekly pay**						
	(1)	(2)	(3)	(4)	(5)	(6)
Log of accessibility	0.269*	0.350**	0.308*	0.238	0.103	0.411***
from work ward	[0.138]	[0.156]	[0.168]	[0.203]	[0.598]	[0.116]
Log of accessibility	0.048	-0.056	-0.061	0.158	-0.333	-0.007
from home ward	[0.133]	[0.111]	[0.192]	[0.122]	[0.344]	[0.098]
Log of travel time	0.026	0.093**	0.007	0.022	-0.03	0.133***
between work and home	[0.049]	[0.043]	[0.061]	[0.040]	[0.130]	[0.045]
Observations	101,473	78,632	57,139	44,334	23,457	55,175
Kleibergen-Paap F-stat	1,103	656	1,216	697	317	692
Work-ward no of clusters	3,273	2,994	2,907	2,533	1,995	2,696
Home-ward no of clusters	3,673	3,614	3,458	3,188	2,721	3,437
PANEL B: **Log of total weekly hours worked**						
	(1)	(2)	(3)	(4)	(5)	(6)
Log of accessibility	0.227*	0.527*	0.303**	0.121	0.579**	0.488
from work ward	[0.124]	[0.300]	[0.138]	[0.204]	[0.249]	[0.374]
Log of accessibility	0.01	-0.36	-0.157	0.199	-0.268	-0.365
from home ward	[0.124]	[0.418]	[0.163]	[0.238]	[0.319]	[0.488]
Log of travel time	-0.066	0.199	-0.073	-0.092*	0.008	0.256
between work and home	[0.043]	[0.133]	[0.052]	[0.050]	[0.097]	[0.167]
Observations	100,992	79,400	57,072	43,920	23,423	55,977
Kleibergen-Paap F-stat	1,105	646	1,224	692	317	676
Work-ward no of clusters	3,267	2,997	2,904	2,521	1,989	2,698
Home-ward no of clusters	3,671	3,621	3,457	3,178	2,715	3,452
Skill/gender	H Skills	L Skills	HS female	HS male	LS female	LS male

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.13: Robustness: by skill level (high/low) and gender – 2002-2008

In table 4.13 we divide the workers into two skills groups, high and low, based on their initial occupation category. This classification is based on the Standard Occupation Classification 2000 manual (Volume 1, page 6)[18]. Higher levels relate to more skilled occupations (for example corporate manager or science and technology professionals), while lower levels relate to less skilled or more mechanical occupations (for example, elementary administration and services occupations). Columns 1-2 display the results for high and low skilled workers, and columns 4 to 6 splits the sample by skills and gender. We find positive and significant effects of accessibility from work on weekly wages in both groups, but the effect is larger and more significant for high skilled workers. We find the same pattern for the effects on hours, but the coefficients are significant only at the 10% level. By gender, as in table 4.12

[18]More information can be found in `http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/standard-occupational-classification-2000/index.html`.

we find that accessibility has an impact on female working hours[19].

PANEL A: **Log of basic weekly pay**						
	(1)	(2)	(3)	(4)	(5)	(6)
Log of accessibility	0.325***	0.302	0.29	0.354***	0.125	1.495*
from work ward	[0.124]	[0.320]	[0.244]	[0.104]	[0.316]	[0.906]
Log of accessibility	-0.023	-0.059	-0.244	0.065	0.169	-2.659**
from home ward	[0.079]	[0.380]	[0.223]	[0.061]	[0.342]	[1.329]
Log of travel time	0.061**	0.034	0.015	0.077***	0.017	1.061*
between work and home	[0.029]	[0.081]	[0.070]	[0.023]	[0.074]	[0.547]
Observations	142,930	37,175	50,020	92,910	30,576	6,599
Kleibergen-Paap F-stat	991	834	835	867	872	153
Work-ward no of clusters	3,357	2,678	2,603	3,081	2,499	1,198
Home-ward no of clusters	3,780	3,269	3,371	3,679	3,082	1,545
PANEL B: **Log of total weekly hours worked**						
	(1)	(2)	(3)	(4)	(5)	(6)
Log of accessibility	0.101	2.007	0.175	0.066	0.588*	11.707*
from work ward	[0.114]	[1.241]	[0.122]	[0.143]	[0.338]	[6.980]
Log of accessibility	0.035	-1.798	-0.278*	0.153*	0.059	-19.675*
from home ward	[0.071]	[1.640]	[0.144]	[0.083]	[0.313]	[10.395]
Log of travel time	0.000	0.120	0.044	-0.021	-0.077	11.944
between work and home	[0.030]	[0.200]	[0.059]	[0.031]	[0.061]	[8.843]
Observations	143,323	37,069	49,998	93,325	30,497	6,572
Kleibergen-Paap F-stat	979	848	838	849	882	153
Work-ward no of clusters	3,351	2,674	2,601	3,074	2,496	1,197
Home-ward no of clusters	3,784	3,269	3,370	3,687	3,081	1,540
Status/gender	Full-time	Part-time	FT female	FT male	PT female	PT male

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table 4.14: Robustness: status (full-time/part-time) and gender – 2002-2008

In table 4.14 we split the sample depending on the initial full-time or part-time status. Columns 1-2 show the results for all genders, and columns 4 to 6 divide the sample by gender and status. We find that the effect of work accessibility on wages only remains significant for full-time workers, and the size is similar to the main results. When splitting by gender, males seem to be driving all the impact of accessibility on wages. We do not find any specific pattern on the effect on hours.

The appendix reports some extra results. In table B.17 we obtain the results for the five wide industrial sectors. The number of observations used for each estimation is smaller than before, so the estimates are less precise. We find positive and significant effects of accessibility from workplace on weekly wages for manufacturing and construction (columns 1 and 2). In these sectors firms could be benefiting from increased input-output sharing or access to larger markets and capitalising any

[19] In table B.18 in the appendix we define 4 wide skill categories based on the occupation codes, defined the first time the individual is observed in the panel. We find similar effects of accessibility from workplace on wages and total hours for workers in category skill2, which is the one that has the largest number of observations. We find weak evidence for workers in category skill3. However, the standard errors in columns 1, 3 and 4 are large so the estimates are quite imprecise.

productivity increase due to this into higher wages paid to their workers. We do not find any evidence of effects on accessibility from work on wages for the other three industrial sector categories (consumer services, producer services or other sectors). We find no effect on accessibility on hours worked. The estimates are quite imprecise because the sample size is substantially reduced when we split the sample by sector.

Finally, in tables B.19 and B.20 we test the robustness of the results to different initial characteristics of the ward in which they are located. We divided the ward in two groups based on the median level of accessibility in 2001, median level of address density in 2001 (table B.19), median level of unemployment rate in 2001 and median level of proportion of commuters by motor vehicles in 2001 (table B.20).

The effect of accessibility from work on weekly wages is quite robust to the initial level of accessibility (columns 1 to 4 of table B.19), but the effect on total hours works only seems to remain positive and significant for ward in which the initial level of accessibility was high. For initial density, the results are similar for total hours worked, i.e. accessibility from workplace only has an effect on hours in ward in which initial residential density was high. For wages, the evidence mixed.

In the case of initial levels of unemployment or motor commuters, the effects of accessibility from work on weekly wages are very robust across groups, although the estimates are less precise than in the main results. The coefficients are also very similar to the baseline estimates. However, accessibility from work only seems to have an effect on total hours work for individuals living in ward with high unemployment rate and working in ward with low unemployment rate (columns 1 and 4 of table B.20). Effects on hours depending on the proportion of motor commuters are weaker, and do not seem to follow any specific pattern.

4.3 Effects on employment status and travel times

The results in section 4.2 inform us about the effect of accessibility on labour market outcomes of workers which are already employed. But changes in accessibility from work and home wards could also affect other outcomes. In this section we provide evidence on the effect of changes in accessibility on the probability of being employed versus being unemployed or out of work. We also investigate if the effects of accessibility on labour market outcomes are working through the channel of reducing commuting costs, measured by travel time to workplace in our data.

4.3.1 Empirical specification and identification strategy

For the analysis of the effect of accessibility on other labour market outcomes we use data from the Labour Force Survey (more details in section 4.3.2). Due to the setup of this data, we use a slightly different identification strategy. We test the effect of accessibility on two outcomes: on employment status (a dummy variable which takes the value 1 if the individual is employed and 0 if he is unemployed or out of work) and on commuting time (which is only available in some quarters).

We have quarterly observations for individuals for which we have information on their personal characteristics (gender, age, ethnicity, marital status, household composition, tenancy status, level of education, etc.) and their job characteristics (industrial sector, occupation, full time/part time status, public/private sector, main/secondary job, hours worked, income, etc.). We have information on the location of households at the ward level. However, we have limited information on the location of jobs. When the individual is working, we know if the individual works and lives within the same local authority. For these reasons, we can only test the effect of accessibility from home. We also have information on their travel time to workplace[20].

Even if the individual data has quarterly frequency, the accessibility measures only change on a yearly basis. Due to the design of the survey, we cannot define individual fixed effects[21]. Instead we can define so called "pseudo-fixed effects" ($\tilde{\mu}_i$) and estimate a pseudo-panel. This strategy is commonly used in survey data which has the same structure as the LFS, i.e. repeated cross-sections (see for example Nickell et al., 2002; Warunsiri & McNown, 2010; Dearden et al., 2011). We allocate a different fixed-effect to individuals with a combination of specific set of characteristics. We exploit the variation of different individuals around the mean of a "representative" individual, defined by the set of characteristics. Depending on how many different characteristics we use to define these dummies, the estimation strategy becomes more demanding[22]. In our case we define the pseudo fixed effects using gender, 10-year age group and ethnicity (the base category is white, male, aged 16-25).

We estimate the following specification:

$$y_{iht} = \beta_0 + \beta_1 A_{iht} + \theta X_{it} + \delta Z_{jt} + \lambda W_{ht} + \tilde{\mu}_i + \xi_t + \xi_q + \varepsilon_{iht} \tag{4.5}$$

[20] And on the transport mode used to go to work, but only for very few observations.

[21] Even in the best case scenario (if the same individual is surveyed twice in the same address in quarters 1 and 5) we can only observe the same individual with one year gap.

[22] We could for example define an individual-home ward pseudo fixed-effect, $\tilde{\mu}_{ih}$ which would exploit the variation for the individuals defined by the gender-age-ethnicity within a given location. For this we need enough observations within each ward for each category defined by the pseudo-fixed effect. Given the small spatial unit of analysis we are using this is not feasible.

where y_{iht} is a given labour market outcome for individual i, living in ward h at time t; A_{iht} is accessibility to employment from home, X_{it} is a matrix of individual characteristics (not included in the pseudo fixed-effects), Z_{jt} is a matrix of job characteristics (when employed), W_{ht} is a matrix of home-ward characteristics, $\tilde{\mu}_i$ denotes the pseudo fixed-effects as defined above, ξ_t and ξ_q are year and quarter dummies which control for common annual shocks and quarter seasonality and finally ε_{iht} denotes the idiosyncratic error term.

As for the results on wages and hours worked, we address the identification of coefficient β by estimating a reduced-form specification which uses $\hat{A}_{iht}$ (1997 employment). To tackle the endogeneity of the transport improvements we only use information on individuals living within 20 kilometres from a improvement opened between 1998 and 2007[23].

4.3.2 Data

We use data from the Labour Force Survey (LFS), which is a quarterly sample survey of households living at private addresses in Great Britain. Its purpose is to provide information on the UK labour market that can then be used to develop, manage, evaluate and report on labour market policies. The survey seeks information on respondents' personal circumstances and their labour market status during a specific reference period, normally a period of one week or four weeks (depending on the question) immediately prior to the interview. Data is available quarterly from 1992. The survey is based on household residing at a given address, not on specific individuals. It contains interviews of all members of the household in five consecutive quarters. The same address can host different individuals across the five quarters of interviews. Each quarter a new wave enters the survey and an old wave leaves. Table 4.15 displays the number of observations by year and quarter.

The smallest geographical unit available for the location of the household is ward (as defined in 1998). The location of workplace is only reported at the regional level (for around 20 regions). However the LFS indicates if the worker works and lives in the same Local Authority District (of which there are 354 in Great Britain). The survey provides information on personal and household characteristics, level of education, sector, occupation, hours worked, travel to work and employment status, although information on earnings is not as good and reliable as that contained in ASHE[24]. From this LFS we can then recover information on unemployment or inactivity status. It also has information on the household composition and detailed

[23] The sample size of the LFS is sufficiently large to be able to focus on the 20-kilometre distance band, as we did in chapter 3.

[24] The response rate for the earning and hours questions is quite low and the data on earnings and hours is self-reported, and thus less reliable than employer provided data.

YEAR	QUARTER 1	2	3	4	**Total**
1998	32,042	31,974	31,520	31,375	126,911
1999	30,851	31,762	31,344	31,449	125,406
2000	31,225	31,010	30,415	30,235	122,885
2001	30,124	29,762	29,894	28,966	118,746
2002	27,362	25,280	22,153	19,504	94,299
2003	22,233	23,609	25,123	24,288	95,253
2004	22,071	21,283	20,465	18,568	82,387
2005	23,314	26,330	26,254	26,159	102,057
2006	26,300	26,385	26,244	26,153	105,082
2007	26,406	26,403	26,489	26,403	105,701
2008	26,453	26,049	25,520	26,062	104,084
Total	298,381	299,847	295,421	289,162	1,182,811

Table 4.15: LFS number of observations

information on personal characteristics. The main disadvantage of the survey is that it is not a panel. However, as discussed above, we can construct a panel of repeated cross sections for a given quarter and use the abundant individual information to construct a pseudo-panel and to control for characteristics like age and gender, occupation group, industry, etc.

4.3.3 Results

4.3.3.1 Travel time to workplace

Table 4.16 shows the results for the effect of log accessibility from home on travel time to workplace. We use data between 1998 and 2008. In columns 1 to 8 we estimate a reduced-form specification in which we use $\hat{A}_{iht}$ and in column 9 we use A_{iht}. The number of observations is much smaller than later because the information on travel to work is only available in the LFS 4^{th} quarter (Autumn).

We estimate the results on a pool of individuals and include year, monthly dummies and pseudo-fixed effects. The standard errors are robust to heteroskedasticity and clustered at the ward level. In column 1 we use OLS. We find a positive association between travel time and accessibility. This could be just because in denser areas, where accessibility is higher, workers travel longer for jobs because of traffic congestion. In column 2 we introduce personal characteristics: student status, skill level, housing tenure, marital status, household composition, number of children (table B.21 shows the details of the coefficients of the personal and job characteristics)[25]. In column 3 we add the job characteristics: occupation (1 digit), full time dummy, second job dummy, work from home dummy, public sector worker dummy

[25]The baseline category is 1st quarter 1998, white-male-16-25, without qualifications, household owns the house, single or living with other people, 1 adult with no children, managers and senior officials, working in sector 01:agriculture,hunting,etc.

and industrial sector dummies (2 digits). Columns 2 and 3 also show positive and significant effects of accessibility on travel time. However in these specifications we do not take into account that individuals might be changing jobs when accessibility from home is increased because in relative terms it becomes worthwhile traveling further.

Log of travel time to workplace									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility	0.0585***	0.053***	0.0115*	-0.016***	-0.004	-0.0125**	0.122	0.0845	
(1997 employment)	0.0066	0.0060	0.0046	0.0039	0.0039	0.0040	0.1223	0.1261	
Log of accessibility									-0.057
(annual employment)									0.041
Live & work in the			-0.677***	-0.722***	-0.721***	-0.723***	-0.747***	-0.747***	-0.747***
same Local Authority			0.0092	0.0083	0.0082	0.0083	0.0087	0.0087	0.0087
Observations	110,148	110,148	110,148	110,148	110,148	110,148	110,148	110,148	110,148
Adjusted R^2	0.027	0.057	0.269	0.295	0.297	0.299	0.272	0.272	0.272
Personal characteristics		YES	YES	YES	YES	YES	YES	YES	YES
Job characteristics			YES	YES	YES	YES	YES	YES	YES
District dummies				YES	YES	YES			
Ward attributes/trends					YES	YES		YES	YES
Scheme dummies/trends						YES		YES	YES
Ward fixed-effects							YES	YES	YES

Clustered (ward) s.e. All specifications include year, monthly dummies & pseudo-fe (age, gender, white). * p<0.05, ** p<0.01, *** p<0.001. Distance band 20 kms. Personal characteristics include student status, skill level, housing tenure, marital status, household composition, number of children. Job characteristics include occupation (1 digit), full time dummy, second job dummy, work from home dummy, public sector worker dummy and industrial sector dummies (2 digits). Ward attributes include distance to closest transport improvement, mean age of the population, proportion of household living in social housing, proportion of household with higher qualifications, average distance traveled to place of work, proportion of households with cars or vans, proportion of employees going to work by motor vehicles, area. Source: ONS, DfT, CASWEB and authors own calculations.

Table 4.16: Travel time to workplace results

In column 4 we control unobservables at the district level and the coefficient becomes negative and significant. Districts correspond to Local Authorities and there are around 350 in Great Britain. The district dummies control by time-invariant unobservable characteristics of the districts which might be correlated with accessibility and travel time at the same time. These could include aspects related to road infrastructure like the district initial endowment of roads, its quality or its level of congestion. In column 5 and 6 we add ward 2001 attributes (in levels and in trends) and scheme dummies and trends. Column 6 controls for district fixed-effects, ward attributes and scheme dummies: it provides evidence of a negative relationship between accessibility and commuting time. In columns 7 and 8, when we introduce the ward fixed-effects, the estimates become very imprecise, possibly due to the lack of variation across pseudo-fixed effects within each ward. In column 9 we use the accessibility measure with time-varying employment, but the estimates are still insignificant.

4.3.3.2 Employment status

In this sub-section we provide the estimates of the effect of accessibility on the probability of being employed versus being out of work (table 4.16) and versus being unemployed (table 4.17). As before, columns 1 to 7 estimate a reduced-form specification and column 8 uses the measure of accessibility with time-varying employment. We estimate a linear probability model (LPM). Alternatively we could have estimated a Probit model, but given the large number of fixed-effects it becomes technically complex. Moreover, we believe the LPM estimates provide a sensible baseline.

PANEL A: **Employment status (1 employed, 0 out of work)**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of accessibility	-0.012***	-0.0008	0.0026*	0.0007	0.0001	0.023	0.019	
(1997 employment)	0.0018	0.0011	0.0012	0.0012	0.0013	0.0342	0.037	
Log of accessibility								0.006
(annual employment)								0.0118
Observations	1,182,811	1,182,811	1,182,811	1,182,811	1,182,811	1,182,811	1,182,811	1,182,811
Adjusted R^2	0.087	0.234	0.237	0.238	0.239	0.205	0.206	0.206
PANEL B: **Employment status (1 employed, 0 unemployed)**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of accessibility	-0.0046***	-0.0002	0.0006	0.0005	0.0005	-0.018	-0.003	
(1997 employment)	0.0007	0.0005	0.0005	0.0006	0.0006	0.0163	0.0181	
Log of accessibility								-0.01
(annual employment)								0.0067
Observations	920,644	920,644	920,644	920,644	920,644	920,644	920,644	920,644
Adjusted R^2	0.036	0.085	0.087	0.088	0.088	0.070	0.070	0.070
Personal characteristics		YES	YES	YES	YES	YES	YES	YES
District dummies			YES	YES	YES			
Ward attributes/trends				YES	YES		YES	YES
Scheme dummies/trends					YES		YES	YES
Ward fixed-effects						YES	YES	YES

Clustered (ward) s.e. All specifications include year, monthly dummies & pseudo-fe (age, gender, white). * p<0.05, ** p<0.01, *** p<0.001. Distance band 20 kms. Personal characteristics include student status, skill level, housing tenure, marital status, household composition, number of children. Ward attributes include distance to closest transport improvement, mean age of the population, proportion of household living in social housing, proportion of household with higher qualifications, average distance traveled to place of work, proportion of households with cars or vans, proportion of employees going to work by motor vehicles, area. Source: ONS, DfT, CASWEB and authors own calculations.

Table 4.17: Employment status results

Column 1 shows a negative significant association between accessibility and the probability of being employed. This could be because individuals with specific characteristics which make them "less employable" locate in places where accessibility is higher. In column 2 we control for personal characteristics. The effect becomes weaker and insignificant. In column 3 we add district dummies. The coefficient of accessibility in the top panel (for employment status when 0 is out of work) becomes positive but it is only weakly significant. In column 4 we add ward attributes which are obtained from the 2001 census[26]. In column 5 we add scheme trends and the

[26]It does not exist time-varying data for the period of analysis available at such small spatial scale.

coefficients decrease and become insignificant again. From column 6 we add ward fixed-effects to control for unobservable factors at the ward level which might be correlated with employment and accessibility. The coefficient becomes very imprecise and the standard errors increase considerably.

In column 7 we add scheme trends and ward attributes trends, and the coefficients become even less precise. As for the results of the effect on travel time, it is very likely that the big standard errors of columns 6 and 7 are due to the lack of variation of the data once we only exploit the within ward variation, because there are not sufficient observations in each ward in each year and in each pseudo-fixed effect category to be able to identify the parameters. In column 8 we use the accessibility measure with time-varying yearly employment A_{iht}. The results are more precise than in column 7 as expected, because accessibility changes much more from year to year when we use yearly employment. Nevertheless the coefficient of accessibility is still quite imprecise and in all cases insignificant.

To summarise, the results in sections 4.3.3.1 and 4.3.3.2 provide some evidence that accessibility to employment from home reduces travel time to workplace but it does not have any effect on the probability of being employed. For these results we we are ignoring any effect that accessibility might have on relocation because we do not observe individuals over time. These results are consistent with those found is section 4.2, in which we found that accessibility from home does not have an effect on labour market outcomes once we only use the variation for a given location.

4.4 Conclusions

In this chapter we have investigated the effect of changes in accessibility, induced by road construction, on individual labour market outcomes. We make use of rich individual datasets and small geographical scale to be able to infer causality on the estimates.

We provide evidence of the effect of accessibility both from workplace and residence on individual wages and hours worked. Controlling for commuting time allows us to learn about the potential theoretical channels through which accessibility might be impacting on labour market outcomes. After controlling for sorting, we find positive effects of accessibility from work on earnings and total hours. These results could be driven by agglomeration externalities which are capitalised into higher nominal wages and by increased spatial competition which might make employees work longer hours. Workers could also be responding to higher nominal wages by increasing overtime or switching from part time to full time jobs. These results are very robust to different specifications and across different groups of workers.

Two main conclusion can be drawn from these results. First, changes in accessibility impact workers by affecting "flexible" outcomes, i.e. wages and overtime. These results are consistent with what we would expect if basic hours are fixed by contract. If higher accessibility from workplace is capitalised into higher wages, workers could be increasing the number of hours work to benefit from the wage increase, but the margin in which they can adjust this is through over time. Secondly, our identification strategy allows us to separate the effect of spatial sorting from other channels through which accessibility might be impacting labour market outcomes. As discussed in the text, once we focus on the variation of accessibility for a given work-home location combination, there is no effect of accessibility from home on hours or wages. But the effect of accessibility from workplace is significant even once we control for sorting, and the size of the coefficient increases substantially. This result stresses the importance of accounting for sorting to be able to correctly attribute the changes in earnings and hours to changes in accessibility.

We also investigate the effect that changes in accessibility might have on employment status and travel times. We find no effect of accessibility from home on the probability of being employed. We find some negative correlations between accessibility from home and travel time to workplace, but these become imprecise when we become more demanding in terms of the identification.

The findings in chapter 4 provide new evidence on the effect of agglomeration externalities and proximity to markets on wages and hours worked using a novel strategy that carefully tackles multiple endogeneity issues. We use transport improvements stemming from road construction as the source of changes in market access. Transport policy is a substantial part of economic policy and the estimates of chapters 3 and 4 help to shed light on the economic impacts that transport infrastructure investments can have on workers and firms.

The results for the effect of accessibility on labour market outcomes complement the results provided in the previous chapter. The fact that in chapter 3 we did not find effects of accessibility of wages paid by firms can be somehow puzzling. However, measure used to approximate firm average wage per worker is very crude (ARD firm wage bill over the firm employment) and only available for a subset of firms. Therefore, the results on wages obtained using ASHE data are likely to be much more reliable.

Appendix A

Spatial impacts of immigration: evidence from the Spanish housing market

A.1 Further issues regarding the interpretation of β

There are still some other issues we would need to take into account in order to interpret β correctly. As native and immigrants are not perfect substitutes in the labour market (Peri & Sparber, 2009), it is possible that the consumption of housing services is also different for foreign-born and natives. Immigrants are likely to consume less housing services than natives (Borjas, 2002), in terms of density (they live in more crowded houses) and in terms of participation in the formal housing markets. Data analysis from the National Immigration Survey (2007) reveals that a significant fraction of immigrants (around 20%) live with relatives/friends or in other informal situations, especially shortly after their arrival. It could be the case that there exist higher credit restrictions for immigrants (Díaz-Serrano & Raya, 2011), that immigrants select into different type of housing than natives or that they live in denser properties (Martori et al., 2006). The β parameter captures the average effect of immigration on house prices, but it does not allow us to say anything about the existence of these phenomena.

This paper studies the impact of immigrants on two types of prices of housing services: the purchase price (generally referred a house price) and the rental price (rents). These two prices are related and influence each other. Immigrants, if renting, would affect rents directly, and purchase prices indirectly. Natives, or other immigrants, would find it profitable to invest in a property in order to rent out it to immigrants, therefore pushing up demand of housing and prices. Depending on the relative price of renting versus buying, immigrants would favour one type of housing tenancy over the other.

In fact, data from the National Immigration Survey (2007) shows that immigrants mainly rent properties. Of the immigrants residing in Spain at the beginning of 2007, that had arrived after 1997, only 21% own the property where they reside, while around 60% rent. In fact, the housing tenancy choices vary a lot between nationality groups. According to National Immigration Survey (2007) immigrants from EU15 countries tend to own the property they reside in (around 70% do), while immigrants from South America, East Europe and Sub-Saharan African predominantly rent (between 70 and 85% do). During the period 1998-2010, the composition of the foreign-born population stocks and inflows has change substantially. Analysis of table A.3 reveals that the weight, both in terms of stocks and of inflows, of foreign-born from the EU15 countries has decreased considerably during the period, the share over the total of immigrants coming from the enlargement countries (especially Romania) and from Latin America has increase largely while the importance of other large immigration groups, like Sub-saharan, has remained relatively stable.

Moreover, most immigrants from East European, Latin American and Sub-Saharan nationalities migrate to Spain "in order to find a/improve job", according to the National Immigration Survey (2007). We would expect most immigrants to favour rental tenancy, as the nature of their migration movement could be considered, at least initially, temporary. If natives invest in properties expecting to get a rent from immigrants, then immigrants would affect purchase prices directly (by purchasing) and indirectly (by motivating others to purchase them). Given that most immigrants rent, the effect on purchase prices would mainly be through the indirect channel, plus any induced effect on purchases from re-located natives. We could expect the effect on purchase prices to be higher than the direct effect on rents, because it would be the sum of the "direct" purchases by immigrants and the purchases by natives in order to rent the properties. Additionally, the estimated effect would depend on the relative demands from both population groups, in case immigrants and natives have different preferences over housing tenure. However, my methodology does not allow me to disentangle the direct and the indirect effects or to control for different preferences, so I can only estimate the average effect.

A final potential concern for the interpretation of β is the choice of the spatial unit of analysis. When immigrants locate in different areas they affect equilibrium prices within "housing markets", which may not necessarily correspond to an administratively defined region. To capture more accurately housing markets, ideally we would use travel-to-work or metropolitan areas, which can be constructed from municipality data using commuter flows. Yet, there exist data limitations and there is no consensus on which is the appropriate definition of a housing market to use[1]. I

[1] Royuela & Vargas (2007) define the housing markets for Catalonia using commuting and migra-

use province (NUTS3) because of data limitations, which, given their size, I consider would capture relatively well the housing markets.

A.2 Further details on the construction of the instrument

A.2.1 Gravity estimations

In order to predict total inflows by nationality in a given year, I use a gravity-type model that only contains push-factors from origin to predict the total inflow from nationality n to Spain in a given year t. The estimated equation is:

$$\ln\left(FBinflow_{from\_n\_to\_Spain,t}\right) = \rho' \ln\left(ECON_{n,t-1}\right) + \omega' \ln\left(GEO_n\right) + \gamma_g + \lambda_t + \xi_{n,t} \tag{A.1}$$

where $ECON_{n,t-1}$ is a matrix of time-varying economic conditions of the sending country (gross domestic output in real terms, total population, percentage of urban population, percentage of internet users, an index of globalisation and dummy of belonging to the EU27). GEO_n is a matrix of time-invariant geographic characteristics of the sending country like (log of) distance to Spain, (log of) area, number of cities, latitude and longitude and dummies for common language, common border and common colonial past with Spain. I include year dummies λ_t and country-group dummies γ_g (the groups appearing in table A.4). I can alternatively include country dummies, which drops the time-invariant variables. I also estimate a similar model using foreign-born stocks on the left hand side (the economic variables are lagged two terms because population is measure on the 1^{st} of January). Data is available for 109 of the 119 countries of table A.4, which represent more than 99% of the inflows into Spain for the period. Results for different specifications are showed in table A.5 for the total national inflows and in table A.6 for the national foreign-born stocks. The specifications include country and country-group dummies alternatively, and the two first columns include year dummies while the last two do not include them. All the models have high predictive power.

From the results in tables A.5 and A.6 I recover the predicted inflows to and predicted stocks of foreign-born in Spain from nationality n in a for every year 2001-2010. I use the prediction from estimates from column (1) for the construction of the instrument, and I use the rest of the specifications estimates for the robustness check. These are combined it with the share by province in base year in a similar

tion flows. On the other hand, Boix & Veneri (2009) define the "metropolitan areas" in Spain. There also exists a definition of "urban areas" done by the Department of Housing. More information can be found in "Atlas Estadstico de las reas Urbanas (2006)".

manner as in (1.5). The imputed predicted foreign-born inflow for each nationality becomes:

$$imp\_pred\_FBinflow^n_{i,t-1} = share^n_{i,base} * \left(\sum_{j \neq i}^{J} pred\_FBinflow^n_{j,t-1}\right) \tag{A.2}$$

The total imputed predicted inflow to each province i at time t is therefore defined as the sum of (A.2) across nationalities:

$$imp\_pred\_FBinflow_{i,t-1} = \sum_n^N \left(imp\_pred\_FBinflow^n_{i,t-1}\right) \tag{A.3}$$

Data from the World Bank World Development Indicators (for the economic variables) and from the *Centre d'Études Prospectives et d'Informations Internationales - CEPII* (for the geographical variables) is used for the estimation of the gravity model of equation (A.1).

A.2.2 Prediction for native location

I use past census data to predict the numbers of natives residing in province i in year t. Total natives in a province are the sum of those born and residing there and those who were born somewhere else in Spain and have moved there. A person born in a given province b (similar to nationalities in the foreign-born case) can either stay where she was born (stayers) or to move and reside in a different province i (movers). R is the total number of provinces in Spain in which natives can locate. For consistence, I use native location patterns from census 1991 as base year (results with 1981 and 2001 censuses are very similar and are used in the robustness checks).

For a given period in the past (base year), we define the share of stayers in province r as the proportion of natives born and staying in a province over all the natives born in the province:

$$share^b_{b,base} = \frac{natives^b_{b,base}}{\sum_i^R natives^b_{r,base}} \tag{A.4}$$

And the share of movers as the proportion of natives born in b but residing in r over all the natives born in b but residing somewhere else:

$$share^b_{r,base} = \frac{natives^b_{r,base}}{\sum_{i \neq b}^R natives^b_{r,base}} \tag{A.5}$$

Share (A.4) is multiplied by the total natives born and staying in the same province in year t to predict the number of stayers in a given year. Share (A.5) is multiplied by the total number of natives living outside the province they were born in year

t (subtracting the natives living in the province for which we want to calculate the prediction, similarly to the case of the foreign-born prediction) to predict the number of natives born in b living in province i at the different points in time. We sum these prediction for each province i across each province of residence r to obtain $imp\_natives_{i,t-1}$.

The correlation between the prediction and the actual number is very high (over 0.90 in all cases), probably because mobility in Spain is low (Decressin & Fatas, 1995) so most natives stay in the province they were born and this is persistent over time. The results are very similar to using actual native numbers in the $imp\_population_{i,t-1}$ in instrument (1.8).

A.3 Instrumental variables estimation comments

The first issue to address in the instrumental variables estimation is that, given that the standard errors are clustered at the province level (50 clusters), when region or province dummies are included the number of clusters is smaller than the number of exogenous regressors plus excluded instruments. This would cause the covariance matrix of orthogonality conditions not to be of full-rank, so we cannot invert it in order to calculate the standard errors. To solve this problem we can partial-out some exogenous regressors (for example the time and region or province dummies) in order to allow the covariance matrix of orthogonality conditions to be invertible. According to the Frisch-Waugh-Lovell (FWL) theorem (Frisch & Waugh, 1933), the coefficients estimated for a regression in which some exogenous regressors are partialled out from the dependent variable, the endogenous regressors, the other exogenous regressors and the excluded instruments are be the same as the coefficients estimated for the original model for certain estimators (see Baum et al., 2007, pg. 484-5).

Secondly, we can be worried about the loss of precision if our instrument is weak (excluded instruments only weakly correlated with included endogenous regressors). The general "rule-of-thumb" is that the F-statistic of the included instrument in the 1^{st} stage must be greater than 10. This is true in all our specifications. The tables which present results obtained using instrumental variables display several diagnosis tests which allow us to critically evaluate the strength of instruments.

We can construct a Lagrange Multipliers-LM statistic to test whether the excluded instruments are "relevant", meaning correlated with the endogenous regressors (under-identification tests). The test is essentially the test of the rank of a matrix: under the null hypothesis that the equation is under-identified, the matrix of reduced form coefficients on the excluded instruments has rank is K1-1, where K1 is number of endogenous regressors. Under the null, the statistic is distributed as chisquared with degrees of freedom (L1-K1+1). A rejection of the null indicates that the matrix is full column rank, i.e., the model is identified (Baum et al., 2010). Kleibergen & Paap (2006) develop a robust version of a test for the rank of a matrix (test for under-identification) which allows to compute the statistic under non-i.i.d errors (cluster-robust in our case). The tables show the value of the "LM test statistic of under-identification (Kleibergen-Paap)" and of the "p-value of under-identification LM statistic". If the p-value is below a given significance level, for example 5%, we can reject the null, the model is be identified, and we can be quite certain that our instruments are relevant. The advantage of using this test is that it can be used with non i.i.d. errors and does not require the computation of additional critical values (Baum et al., 2007).

Cragg-Donald F statistic is a Wald test which also tests for under-identification (which in the case of one endogenous variable is equivalent to the F statistic of included instruments in the first-stage regression). It is also reported in the tables (F-stat weak identification (Cragg-Donald)). Stock & Yogo (2002) compiled critical values for the Cragg-Donald F statistic. Their null hypothesis is that the estimator is weakly identified in the sense that it is subject to bias that is unacceptably large. The compute different critical values depending what we consider "too large". Nevertheless, if the errors are not i.i.d., the Cragg-Donald statistics are not valid anymore. A robust version of this test, developed by Kleibergen-Paap, is also displayed in the tables, but comparison with the Stock-Yogo values should be done with caution, as the critical values are not adjusted for the specific type of cluster-robust variance-covariance matrix. In any case, if anything the F-statistic of the included instruments of the first stage goes significatively down when clustering the standard errors (because these increase substantially, so the t-statistic of the first stage regression of instrument on the endogenous regressor goes down) and we can be quite certain that if the Kleibergen-Paap F-statistic is above the critical Stock-Yogo values our instrument would be sufficiently strong. In all cases the values of the Kleibergen-Paap F-statistic is above the 10% maximal IV size Stock-Yogo weak identification test critical value. This fact, jointly to the low p-values of the LM under-identification test, suggests that our instruments are sufficiently strong.

A.4 Additional robustness checks

Tables A.15 to A.18 show the results changing the different components of the instrument (1.9). There are two components in the construction of the predicted foreign-born stock by province in each year: the base year, which is used to locate spatially the immigrants based on past location patterns by nationality, the national inflow, which is used to have time-variation in the instrument. The instrument also changes depending on how we defined the population of the denominator (if we use predictions for foreign-born, for native or for both). In the main results the instrument is constructed using as 1991 as the base year for the immigration and native location patterns; using the national inflows and stocks by nationality predicted from the gravity estimates of column 1 of tables A.5 and A.6 -which includes country and year dummies-; and using the predicted stocks of immigration and natives to construct the denominator of the instrument. I test the robustness of the results to different definitions of the denominator, different base years for the location patterns and different predictions of the gravity modes.

Table A.15 changes the denominator of instrument (1.9). The table displays the results for rental (top) and purchase prices (bottom). Columns 1 to 4 use actual population in the denominator, as for example in Gonzalez & Ortega (2009) or Saiz (2007). Columns 5 to 8 use the prediction of foreign-born but the actual native stock (as in the definition of the instrument 1.8). Columns 9 to 12 use both the foreign-born and the native stock predictions (instrument 1.9), but using 1981 as the base year. We can expect the results from columns 1 to 4 to be failing to control for the endogenous location of foreign-born. This seems to be the case: for rental prices the coefficients in columns are smaller than those of the main specifications and for purchase prices they are larger compared to the main instrumental variable results. This would suggest that immigrants are locating in provinces where rents are growing faster and purchase prices are growing slower, so it is important to instrument for the endogenous location decisions of immigrants. The estimates of columns 5 to 8 would fail to control for the endogenous location decisions of natives: the coefficients are again smaller for rental prices and bigger for purchase prices, compared to The results obtained by using the immigrant foreign-born stock prediction and the native stock prediction but using 1981 as the base year (NAT81) are very similar to the main results.

Table A.16 show the results using different computations of the national inflow used in equation (1.5), i.e. term $\sum_{j\neq i}^{J} FBstock_{j,t}^{n}$. Tables show the results for the immigration ratio and for the population growth, for rental (top) and purchase prices (bottom panel). They use the instrument constructed either with the actual national inflow, not the prediction (”Actual inflow”) either with the estimates of different predictions of the gravity models of tables A.5 and A.6. ”Column 2” results includes country dummies but not year dummies. The exclusion of year dummies would be justified if there existed yearly common shocks to all countries that make Spain a more attractive country to locate (for example if the country relaxed restrictions to immigration in a given year, like for example the massive immigrant regularization carried on in 2005). ”Column 3” and ”Column 4” use the specification that includes country-group dummies and the time-invariant variables, like distance. The tables display the results from the models with region or province dummies and with and without regional trends, always including controls. When we use the actual inflow the coefficients are slightly bigger than those of table 1.8 but very similar. When we use the model of column 2, the results also remain very similar to the baseline ones. The explanatory power of the gravity model which includes country-fixed effects is very high. When we use the model with country-group fixed effects the results are weaker, specially for purchase prices.

Tables A.17 and A.18 show the results changing the base year used as a predictor of the spatial location patterns of the immigrants. I focus on the results from the model with region dummies and with or without regional trends and controls, i.e. they compare to columns 1 and 2 of table 1.8. The results using province dummies are too imprecise. The instrument is recalculated using as base years 2001, 1985, 1930 and 1940. The tables present the coefficients of the effect of immigration ratio and population growth. The instrument using 2001 and 1985 as the base year are strong but the estimated elasticities but only significant for the specifications which do not include regional trends. The results using 1940 and 1930 are much more imprecise and all insignificant.

A.5 Effects of immigration on the construction sector

In table A.22 I estimate the effect of changes of the immigration ratio and of population growth on the growth of private dwellings stock. The estimated equation is similar to equation (1.1) but using the change in the log of private dwellings stock as dependent variable. Instrumental variables are used to obtain these results. I include the supply attributes that might explain regional trends on the growth of housing stocks. Columns 1 to 6 present the results for the immigration ratio and columns 7 to 12 for the population growth. Column 1 includes regional dummies and province attributes. Column 2 adds the province time-invariant supply related characteristics. Developable land and altitude are significant a have the expected signs. The coefficient for ruggedness on the other side has a positive sign, contrary to expected, even if the coefficient is very small. The coefficient on the share of rented properties is positive, indicated that more houses are built in regions where renting is more common. Column 3 adds the time-varying controls and column 4 the regional trends. Columns 5 and 6 show the first-differences fixed-effect estimates, with and without regional trends. Columns 7 to 12 have the same structure.

The coefficients range between 1.15 and 1.6 for the immigration ratio and around 0.75 to 0.95 for the population growth, and are always significantly different from zero (except in columns 6 and 12, the most demanding specifications). The positive effect is similar to that of Gonzalez & Ortega (2009), although the estimates are not directly comparable because they use a different specification.

These estimates suggest that dwellings stock grew almost proportionally to immigration induced population growth. If all these new houses had been sold to or occupied by immigrants, we would have expected a small effect of immigration on house prices. However, the effect of immigration on prices is robustly positive and of size bigger than 1. It could be the case that the positive effect on construction is just due to the fact that immigrants are more than proportionally concentrated in the construction sector (as already pointed out in section 1.2.2), so their effect in construction would be not through demand but through supplying abundant (non-skilled) labour supply and therefore activating labour-intense sectors like construction. Nevertheless, it does not exist a causal relationship between immigration and growth in the employment sector. In table A.23 I use the same specifications and estimation strategy as in table A.22 to test the effect of immigration and population growth on employment in the construction sector. The standard errors are very high and make the estimates highly imprecise, which would suggest that most of the variation in employment in the construction sector is captured in the time and regional and province dummies.

A.6 Additional tables and figures

Year	**Population**	**Housing stock**	**Stock over population**
2001	40,972,359	20,988,378	0.512
2002	41,692,558	21,504,402	0.516
2003	42,573,670	22,010,730	0.517
2004	43,055,014	22,573,867	0.524
2005	43,967,766	23,160,019	0.527
2006	44,566,232	23,808,108	0.534
2007	45,054,694	24,443,903	0.543
2008	46,008,985	25,076,820	0.545
2009	46,593,673	25,504,442	0.547
2010	46,864,418	25,783,555	0.550

Source: Department of Housing

Table A.1: Residential density in Spain 2001-2010

	1998		2001		2008		2010	
	Total	%Const	Total	%Const	Total	%Const	Total	%Const
Total	13904.2	9.96%	16146.3	11.62%	20257.6	12.11%	18456.5	8.94%
Spanish	13638.4	9.98%	15402.1	11.47%	17122.8	10.66%	15660.5	8.32%
Foreigners	221.5	10.00%	682.8	15.26%	2929.6	20.65%	2549.5	13.05%
from EU	88.7	6.63%	159.6	9.46%	871.3	20.72%	814.1	15.02%
from rest of Europe	11.6	10.54%	84.8	24.10%	112.6	22.02%	104.8	16.60%
from Latin America	43.2	4.28%	262.5	12.50%	1407.2	19.94%	1182.6	12.29%
from rest of the world	78.0	16.90%	176.0	20.39%	538.5	22.12%	447.9	10.65%
Percentage foreigners	*1.59%*		*4.23%*		*14.46%*		*13.81%*	

Source: Labour Force Survey, National Institute of Statistics

Table A.2: Employment in construction for natives and foreigners - selected years

Country group	United Kingdom, France and Germany	Rest of EU15, Norway and Switzerland	Romania, Bulgaria Poland and Hungary	Rest of EU27	Balkans, USSR and Turkey
1998 stock	26.99%	18.99%	1.52%	0.23%	1.12%
2010 stock	12.31%	9.36%	19.06%	0.80%	3.28%
total inflow	10.48%	8.16%	21.25%	0.88%	3.55%
Country group	Rest of Europe	Sub-Saharan Africa	Rest of Africa	United States and Canada	Latin and Central America
1998 stock	0.19%	18.59%	4.63%	2.24%	18.50%
2010 stock	0.05%	14.38%	4.05%	0.50%	30.62%
total inflow	0.03%	13.86%	3.98%	0.28%	32.13%
Country group	Philippines, China and Indo Continent	Rest of Asia	Oceania	Rest of Countries	TOTAL
1998 stock	4.58%	2.13%	0.25%	0.04%	637,090
2010 stock	5.01%	0.51%	0.04%	0.01%	5,747,734
Total inflow	5.07%	0.31%	0.02%	0.01%	5,110,644

Source: Population Registers, National Institute of Statistics

Table A.3: Foreign-born by nationality groups

List of countries/nationality groups			
France	United Kingdom, France & Germany	Cote d'Ivoire	Rest of Africa
United Kingdom	United Kingdom, France & Germany	Egypt	Rest of Africa
Germany	United Kingdom, France & Germany	Ethiopia	Rest of Africa
Austria	Rest of EU15, Norway & Switzerland	Guinea-Bissau	Rest of Africa
Belgium	Rest of EU15, Norway & Switzerland	Equatorial Guinea	Rest of Africa
Denmark	Rest of EU15, Norway & Switzerland	Kenya	Rest of Africa
Finland	Rest of EU15, Norway & Switzerland	Liberia	Rest of Africa
Greece	Rest of EU15, Norway & Switzerland	South Africa	Rest of Africa
Ireland	Rest of EU15, Norway & Switzerland	Sierra Leone	Rest of Africa
Italy	Rest of EU15, Norway & Switzerland	Togo	Rest of Africa
Luxembourg	Rest of EU15, Norway & Switzerland	Zaire	Rest of Africa
Norway	Rest of EU15, Norway & Switzerland	Africa other	Rest of Africa
Netherlands	Rest of EU15, Norway & Switzerland	Canada	United States & Canada
Portugal	Rest of EU15, Norway & Switzerland	United States of America	United States & Canada
Sweden	Rest of EU15, Norway & Switzerland	Mexico	Latin & Central America
Switzerland	Rest of EU15, Norway & Switzerland	Costa Rica	Latin & Central America
Bulgaria	Rumania, Bulgaria, Pol& & Hungary	Cuba	Latin & Central America
Hungary	Rumania, Bulgaria, Pol& & Hungary	Dominica	Latin & Central America
Poland	Rumania, Bulgaria, Pol& & Hungary	El Salvador	Latin & Central America
Romania	Rumania, Bulgaria, Pol& & Hungary	Guatemala	Latin & Central America
Cyprus	Rest of EU27	Honduras	Latin & Central America
Malta	Rest of EU27	Nicaragua	Latin & Central America
Latvia	Rest of EU27	Panama	Latin & Central America
Estonia	Rest of EU27	Dominican Republic	Latin & Central America
Lithuania	Rest of EU27	Argentina	Latin & Central America
Czech Republic	Rest of EU27	Bolivia	Latin & Central America
Slovakia	Rest of EU27	Brazil	Latin & Central America
Slovenia	Rest of EU27	Colombia	Latin & Central America
Iceland	Rest of Europe	Chile	Latin & Central America
Liechtenstein	Rest of Europe	Ecuador	Latin & Central America
Andorra	Rest of Europe	Paraguay	Latin & Central America
Europe other	Rest of Europe	Peru	Latin & Central America
Albania	Balkans, USSR & Turkey	Uruguay	Latin & Central America
Ukraine	Balkans, USSR & Turkey	Venezuela	Latin & Central America
Moldova	Balkans, USSR & Turkey	America other	Latin & Central America
Belarus	Balkans, USSR & Turkey	Bangladesh	Philippines, China & Indo-continent
Georgia	Balkans, USSR & Turkey	China	Philippines, China & Indo-continent
Bosnia Herzegovina	Balkans, USSR & Turkey	Philippines	Philippines, China & Indo-continent
Croatia	Balkans, USSR & Turkey	India	Philippines, China & Indo-continent
Armenia	Balkans, USSR & Turkey	Pakistan	Philippines, China & Indo-continent
Russia	Balkans, USSR & Turkey	Saudi Arabia	Rest of Asia
Serbia & Montenegro	Balkans, USSR & Turkey	Indonesia	Rest of Asia
Macedonia	Balkans, USSR & Turkey	Iraq	Rest of Asia
Turkey	Balkans, USSR & Turkey	Iran	Rest of Asia
Gambia	Sub-Saharan Africa	Israel	Rest of Asia
Ghana	Sub-Saharan Africa	Japan	Rest of Asia
Guinea	Sub-Saharan Africa	Jordan	Rest of Asia
Mali	Sub-Saharan Africa	Lebanon	Rest of Asia
Nigeria	Sub-Saharan Africa	Nepal	Rest of Asia
Senegal	Sub-Saharan Africa	South Korea	Rest of Asia
Algeria	North Africa	Syria	Rest of Asia
Morocco	North Africa	Thailand	Rest of Asia
Mauritania	North Africa	Vietnam	Rest of Asia
Tunisia	North Africa	Kazakhstan	Rest of Asia
Burkina Faso	Rest of Africa	Asia other	Rest of Asia
Angola	Rest of Africa	Australia	Oceania
Benin	Rest of Africa	New Zealand	Oceania
Cape Verde	Rest of Africa	Oceania other	Oceania
Cameroon	Rest of Africa	Stateless	Stateless
Congo	Rest of Africa		

Table A.4: List of countries and nationality groups

Depvar: inflow of immigrants from country n to Spain	(1)	(2)	(3)	(4)
Log of GDP in billions of const dollars in t-1	-1.386***	-0.520***	-1.093**	-0.601***
	[0.467]	[0.185]	[0.434]	[0.182]
Log of total population in 1000s in t-1	-1.603	0.890***	1.987	0.960***
	[1.441]	[0.229]	[1.231]	[0.226]
Percentage of urban population in t-1	0.876	3.355***	3.521	3.141***
	[4.429]	[0.824]	[4.194]	[0.784]
Percentage of internet users in t-1	-1.934***	-0.061	-0.576	0.568
	[0.431]	[0.453]	[0.452]	[0.395]
Globalisation index in t-2	0.015	-0.011	0.083***	0.002
	[0.017]	[0.014]	[0.017]	[0.013]
Dummy if country belongs to the EU	1.044***	0.464*	0.935***	0.470*
	[0.176]	[0.260]	[0.210]	[0.265]
Log of distance between country and Spain		-1.794***		-1.806***
		[0.436]		[0.423]
Log of area in sq kms		0.311***		0.303***
		[0.104]		[0.099]
Number of cities in the country in Henderson data		-0.308***		-0.313***
		[0.050]		[0.053]
Latitude in degrees		0.002		0.000
		[0.007]		[0.007]
Longitude in degrees		0.026***		0.025***
		[0.008]		[0.008]
Dummy if country official language is Spanish		2.246***		2.283***
		[0.619]		[0.668]
Dummy if country is contiguous to Spain		-0.413		-0.379
		[0.544]		[0.543]
Dummy if country was a colony of Spanish Empire		-0.285		-0.285
		[0.543]		[0.556]
Constant	23.327*	15.189***	-15.664	15.602***
	[13.297]	[3.484]	[10.529]	[3.567]
Year 1999	0.717***	0.503***		
	[0.113]	[0.117]		
Year 2000	1.445***	1.145***		
	[0.154]	[0.151]		
Year 2001	1.860***	1.457***		
	[0.172]	[0.158]		
Year 2002	2.101***	1.650***		
	[0.198]	[0.175]		
Year 2003	1.572***	1.031***		
	[0.250]	[0.227]		
Year 2004	2.499***	1.824***		
	[0.256]	[0.200]		
Year 2005	2.094***	1.281***		
	[0.313]	[0.240]		
Year 2006	2.075***	1.239***		
	[0.386]	[0.276]		
Year 2007	2.746***	1.736***		
	[0.370]	[0.254]		
Year 2008	2.579***	1.416***		
	[0.392]	[0.266]		
Year 2009	2.019***	0.698**		
	[0.448]	[0.286]		
Fixed-effects	Country	Group	Country	Group
Observations	1142	1142	1142	1142
Adjusted R^2	0.872	0.648	0.818	0.597

Clustered (country) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.5: Gravity equations inflows estimates

Depvar: stock of immigrants from country n in Spain	(1)	(2)	(3)	(4)
Log of GDP in billions of const dollars in t-2	-0.275	-0.207	0.864**	-0.388**
	[0.252]	[0.146]	[0.433]	[0.155]
Log of total population in 1000s in t-2	-4.644***	0.675***	0.816	0.847***
	[1.203]	[0.191]	[1.299]	[0.203]
Percentage of urban population in t-2	-3.328	3.596***	6.141	3.122***
	[3.315]	[0.725]	[4.155]	[0.728]
Percentage of internet users in t-1	-2.021***	-0.377	0.112	1.623**
	[0.554]	[0.316]	[0.288]	[0.674]
Globalisation index in t-3	0.018	-0.014	0.097***	0.019
	[0.013]	[0.012]	[0.019]	[0.013]
Dummy if country belongs to the EU	0.464*	0.263	0.693**	0.557*
	[0.235]	[0.268]	[0.269]	[0.294]
Log of distance between country and Spain		-1.631***		-1.732***
		[0.392]		[0.384]
Log of area in sq kms		0.190*		0.188*
		[0.108]		[0.102]
Number of cities in the country in Henderson data		-0.333***		-0.355***
		[0.047]		[0.058]
Latitude in degrees		-0.001		-0.004
		[0.007]		[0.007]
Longitude in degrees		0.020***		0.019***
		[0.007]		[0.007]
Dummy if country official language is Spanish		1.905***		2.108***
		[0.646]		[0.782]
Dummy if country is contiguous to Spain		-0.148		-0.081
		[0.548]		[0.546]
Dummy if country was a colony of Spanish Empire		-0.133		-0.160
		[0.585]		[0.659]
Constant	52.244***	18.895***	-11.945	18.605***
	[11.495]	[3.241]	[10.885]	[3.516]
Year 1999	0.121***	0.051		
	[0.042]	[0.033]		
Year 2000	0.480***	0.347***		
	[0.076]	[0.053]		
Year 2001	0.997***	0.781***		
	[0.127]	[0.089]		
Year 2002	1.427***	1.124***		
	[0.167]	[0.113]		
Year 2003	1.782***	1.403***		
	[0.204]	[0.133]		
Year 2004	1.981***	1.472***		
	[0.243]	[0.144]		
Year 2005	2.332***	1.704***		
	[0.272]	[0.154]		
Year 2006	2.560***	1.824***		
	[0.305]	[0.167]		
Year 2007	2.739***	1.902***		
	[0.341]	[0.181]		
Year 2008	3.017***	2.063***		
	[0.376]	[0.191]		
Year 2009	3.235***	2.163***		
	[0.412]	[0.201]		
Fixed-effects	Country	Group	Country	Group
Observations	1308	1308	1308	1308
Adjusted R^2	0.951	0.745	0.922	0.667

Clustered (country) standard errors in brackets. t=2002/2010. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.

Table A.6: Gravity equations stocks estimates

Variables	*Time period*	*Mean*	*Std. Dev.*	*Min*	*Max*
Change of log rent prices (per sq m)	2002/2010	0.0333	0.0152	-0.0053	0.0831
Change of log house prices (per sq m) - IVIE	2002/2010	0.0715	0.0849	-0.1437	0.2765
Change of log house prices (per sq m) - Housing Dpt average	2002/2010	0.0638	0.0798	-0.1437	0.2542
Change of log house prices (per sq m) - Housing Dpt 2nd quarter	2002/2010	0.0654	0.0845	-0.1682	0.3192
Inflow of population in t-1 over population end t-2	2001/2009	0.0125	0.0127	-0.0100	0.0612
Inflow of immigrants in t-1 over population end t-2	2001/2009	0.0105	0.0083	-0.0054	0.0456
Inflow of natives in t-1 over population end t-2	2001/2009	0.0021	0.0065	-0.0168	0.0313
Inflow of population in t-1 over population end t-2 (working-age)	2001/2009	0.0091	0.0100	-0.0083	0.0500
Inflow of immigrants in t-1 over population end t-2 (working-age)	2001/2009	0.0080	0.0068	-0.0062	0.0398
Inflow of natives in t-1 over population end t-2 (working-age)	2001/2009	0.0011	0.0049	-0.0091	0.0244
Dummy if region only has one province	Time invariant	0.1400	0.3472	0	1
Dummy if province is an island	Time invariant	0.0600	0.2377	0	1
Log of the surface of natural parks (in sq kms)	Time invariant	11.0181	1.1307	8.5007	12.6216
Coast dummy	Time invariant	0.4400	0.4967	0	1
Length of coastline (in kms)	Time invariant	156.8200	281.9197	0	1428
Log of hours of average temperature (January)	Time invariant	1.9595	0.4622	1.0784	2.9025
Log of mm of rain precipitation (January)	Time invariant	3.7124	0.5833	2.7770	5.3642
Log of number of retails shops	2000	9.4060	0.7777	7.7450	11.5112
Log of number of restaurants and bars	2000	8.1272	0.9006	5.6971	10.3467
Importance of tourism sector - comparative index	2000	19.9704	32.0277	1.2700	163.2900
Log of the number of doctors	2000	7.5119	1.0929	3.3322	10.2324
Change of log of GDP	2000/2008	0.0685	0.0231	-0.0087	0.1365
Change of log of unemployment rate	2000/2008	-0.0049	0.0244	-0.1343	0.0954
Change of log of number of credit establishments	2000/2008	0.0110	0.0277	-0.0728	0.0972
Change of percentage of savings banks	2000/2008	0.0090	0.0104	-0.0465	0.0494
Share of developable land in 2000 (Corine)	2000	0.8556	0.0732	0.4652	0.9609
Relative index of altitude	Time invariant	95.1959	39.7765	20.3942	165.2263
Relative index of ruggedness	Time invariant	110.3139	48.2588	31.0299	247.6143
Percentage of rented properties in 2001	2001	0.1027	0.0363	0.0578	0.2116
Percentage of empty homes in 2001	2001	0.1485	0.0243	0.0846	0.1913
Log of change of stock of private dwellings	2000/2008	0.0261	0.0121	0.0067	0.0935

Table A.7: Summary statistics

PROVINCES	*Change of log of rent prices*		*Change of log of purchase prices*		*Lagged population ratio*		*Lagged immigration ratio*		*Lagged natives ratio*	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Alava	0.0382	0.0666	0.0309	0.0129	0.0105	0.0031	0.0090	0.0023	0.0015	0.0014
Albacete	0.0887	0.1056	0.0301	0.0116	0.0100	0.0039	0.0081	0.0043	0.0019	0.0012
Alicante	0.0614	0.1025	0.0345	0.0135	0.0290	0.0158	0.0224	0.0137	0.0066	0.0023
Almeria	0.0899	0.1027	0.0373	0.0202	0.0300	0.0121	0.0219	0.0116	0.0082	0.0019
Avila	0.0693	0.0968	0.0218	0.0099	0.0053	0.0057	0.0076	0.0051	-0.0022	0.0023
Badajoz	0.0835	0.0687	0.0400	0.0221	0.0046	0.0044	0.0034	0.0019	0.0012	0.0033
Baleares (Illes)	0.0600	0.0816	0.0382	0.0125	0.0260	0.0126	0.0201	0.0107	0.0059	0.0032
Barcelona	0.0599	0.0719	0.0398	0.0114	0.0154	0.0082	0.0145	0.0071	0.0009	0.0019
Burgos	0.0512	0.0629	0.0360	0.0114	0.0077	0.0064	0.0096	0.0061	-0.0019	0.0013
Caceres	0.0733	0.0644	0.0193	0.0105	0.0016	0.0020	0.0022	0.0017	-0.0006	0.0014
Cadiz	0.1017	0.1071	0.0398	0.0137	0.0099	0.0029	0.0036	0.0015	0.0063	0.0017
Castellon	0.0790	0.1014	0.0319	0.0148	0.0247	0.0115	0.0187	0.0098	0.0060	0.0022
Ciudad Real	0.0793	0.0872	0.0254	0.0138	0.0113	0.0059	0.0092	0.0050	0.0021	0.0021
Cordoba	0.0942	0.1061	0.0438	0.0189	0.0050	0.0022	0.0033	0.0017	0.0018	0.0011
Corua (La)	0.0683	0.0749	0.0308	0.0096	0.0038	0.0025	0.0042	0.0015	-0.0004	0.0013
Cuenca	0.0887	0.0920	0.0294	0.0151	0.0086	0.0065	0.0131	0.0070	-0.0045	0.0032
Girona	0.0915	0.0946	0.0422	0.0174	0.0296	0.0107	0.0208	0.0090	0.0087	0.0022
Granada	0.0751	0.0868	0.0364	0.0112	0.0137	0.0056	0.0074	0.0031	0.0063	0.0041
Guadalajara	0.0814	0.1206	0.0334	0.0177	0.0435	0.0112	0.0196	0.0089	0.0239	0.0057
Guipuzcoa	0.0578	0.0725	0.0319	0.0138	0.0044	0.0022	0.0063	0.0021	-0.0019	0.0013
Huelva	0.0834	0.0946	0.0230	0.0111	0.0129	0.0045	0.0082	0.0041	0.0046	0.0025
Huesca	0.0836	0.1015	0.0409	0.0146	0.0117	0.0061	0.0126	0.0056	-0.0009	0.0014
Jaen	0.0849	0.0795	0.0312	0.0089	0.0042	0.0022	0.0030	0.0014	0.0012	0.0020
Leon	0.0519	0.0564	0.0289	0.0113	0.0000	0.0047	0.0045	0.0023	-0.0046	0.0031
Lleida	0.0890	0.0863	0.0286	0.0178	0.0209	0.0087	0.0191	0.0083	0.0019	0.0016
Rioja (La)	0.0650	0.0728	0.0312	0.0149	0.0198	0.0115	0.0147	0.0085	0.0051	0.0041
Lugo	0.0636	0.0810	0.0118	0.0074	-0.0033	0.0023	0.0037	0.0020	-0.0070	0.0008
Madrid	0.0650	0.1061	0.0415	0.0137	0.0207	0.0100	0.0165	0.0087	0.0042	0.0031
Malaga	0.0883	0.1369	0.0399	0.0179	0.0239	0.0093	0.0153	0.0067	0.0086	0.0029
Murcia	0.0780	0.1076	0.0378	0.0170	0.0231	0.0084	0.0154	0.0076	0.0077	0.0012
Navarra	0.0463	0.0664	0.0275	0.0156	0.0152	0.0058	0.0117	0.0058	0.0035	0.0010
Orense	0.0485	0.0434	0.0304	0.0121	-0.0031	0.0016	0.0039	0.0014	-0.0069	0.0009
Asturias	0.0721	0.0760	0.0319	0.0076	0.0009	0.0026	0.0047	0.0019	-0.0038	0.0012
Palencia	0.0582	0.0727	0.0377	0.0142	-0.0031	0.0030	0.0041	0.0017	-0.0072	0.0021
Palmas(Las)	0.0420	0.0690	0.0282	0.0131	0.0186	0.0093	0.0128	0.0072	0.0058	0.0023
Pontevedra	0.0744	0.0745	0.0373	0.0083	0.0055	0.0019	0.0045	0.0018	0.0010	0.0012
Salamanca	0.0642	0.0821	0.0185	0.0076	0.0011	0.0055	0.0046	0.0030	-0.0036	0.0038
SC de Tenerife	0.0484	0.0707	0.0263	0.0102	0.0205	0.0107	0.0145	0.0069	0.0060	0.0049
Cantabria	0.0664	0.0819	0.0376	0.0137	0.0108	0.0034	0.0073	0.0025	0.0036	0.0018
Segovia	0.0748	0.0773	0.0493	0.0215	0.0124	0.0092	0.0133	0.0093	-0.0008	0.0022
Sevilla	0.0988	0.0884	0.0376	0.0151	0.0104	0.0033	0.0042	0.0017	0.0061	0.0019
Soria	0.0629	0.0790	0.0227	0.0117	0.0047	0.0057	0.0106	0.0035	-0.0058	0.0049
Tarragona	0.0896	0.1163	0.0394	0.0140	0.0315	0.0121	0.0199	0.0095	0.0115	0.0031
Teruel	0.0634	0.0693	0.0306	0.0160	0.0072	0.0077	0.0124	0.0080	-0.0052	0.0015
Toledo	0.0882	0.1302	0.0425	0.0131	0.0298	0.0104	0.0138	0.0070	0.0160	0.0047
Valencia	0.0722	0.0926	0.0361	0.0106	0.0165	0.0073	0.0129	0.0061	0.0037	0.0018
Valladolid	0.0610	0.0783	0.0267	0.0102	0.0077	0.0034	0.0068	0.0033	0.0009	0.0016
Vizcaya	0.0591	0.0806	0.0457	0.0163	0.0021	0.0020	0.0059	0.0016	-0.0038	0.0012
Zamora	0.0593	0.0806	0.0270	0.0135	-0.0045	0.0030	0.0039	0.0029	-0.0084	0.0012
Zaragoza	0.0827	0.1029	0.0428	0.0135	0.0142	0.0067	0.0133	0.0066	0.0009	0.0026
Total	0.0715	0.0849	0.0333	0.0152	0.0125	0.0127	0.0105	0.0083	0.0021	0.0065

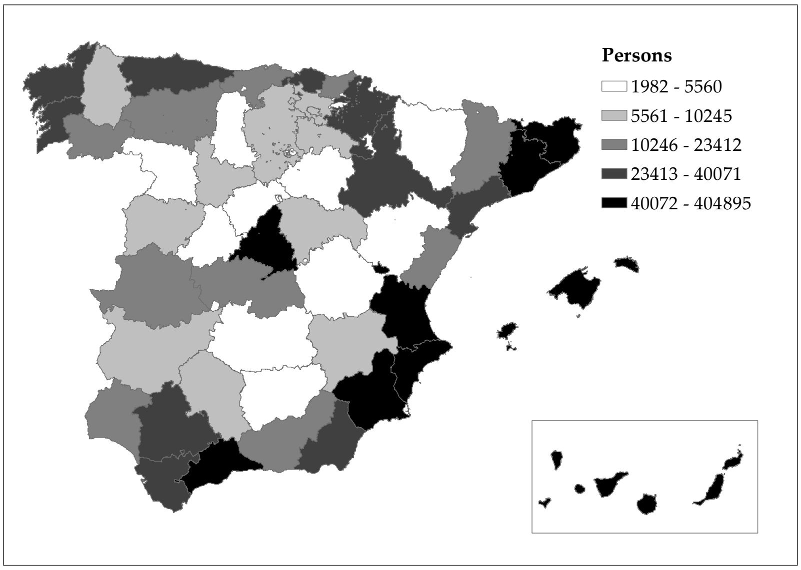
Table A.8: Descriptive statistics by province

	Sum of the squares: provinces (groups) over 9 years				
	Change of log of rent prices	*Change of log of purchase prices*	*Lagged population ratio*	*Lagged immigration ratio*	*Lagged natives ratio*
Between groups	0.0258	0.1055	0.0511	0.0156	0.0164
Within groups	0.0774	3.1280	0.0213	0.0150	0.0026
Total	0.1032	3.2335	0.0724	0.0306	0.0190

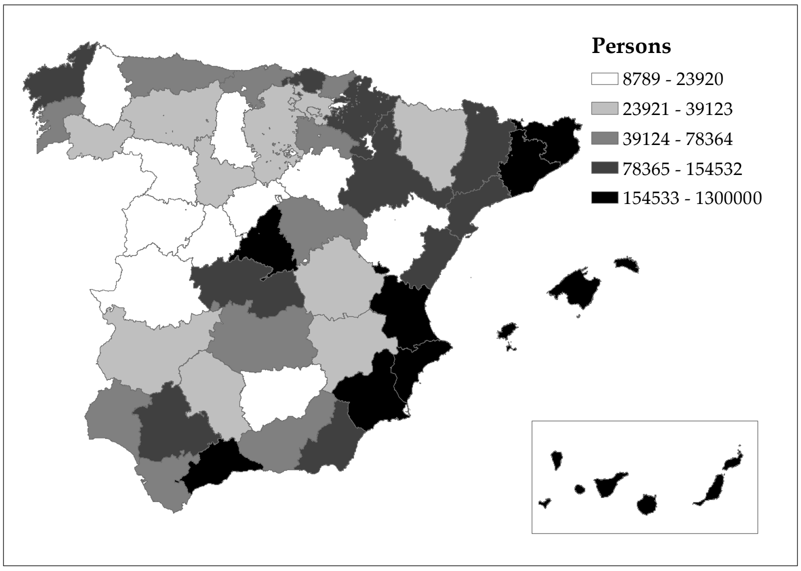
Table A.9: Sum of squares between and within provinces

YEARS	*Change of log of rent prices*		*Change of log of purchase prices*		*Lagged population ratio*		*Lagged immigration ratio*		*Lagged natives ratio*	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2002	0.0396	0.0135	0.1188	0.0355	0.0134	0.0147	0.0123	0.0087	0.0011	0.0073
2003	0.0380	0.0130	0.1290	0.0451	0.0147	0.0143	0.0131	0.0091	0.0015	0.0075
2004	0.0338	0.0134	0.1454	0.0584	0.0099	0.0104	0.0080	0.0052	0.0019	0.0067
2005	0.0376	0.0163	0.1494	0.0389	0.0180	0.0148	0.0143	0.0099	0.0037	0.0071
2006	0.0385	0.0136	0.1132	0.0266	0.0124	0.0117	0.0094	0.0065	0.0030	0.0068
2007	0.0357	0.0115	0.0766	0.0273	0.0108	0.0114	0.0091	0.0065	0.0017	0.0063
2008	0.0388	0.0116	0.0092	0.0212	0.0193	0.0137	0.0171	0.0090	0.0022	0.0065
2009	0.0272	0.0099	-0.0742	0.0296	0.0104	0.0083	0.0083	0.0044	0.0021	0.0054
2010	0.0103	0.0059	-0.0235	0.0305	0.0040	0.0060	0.0025	0.0025	0.0015	0.0046
Total	0.0715	0.0849	0.0333	0.0152	0.0125	0.0127	0.0105	0.0083	0.0021	0.0065

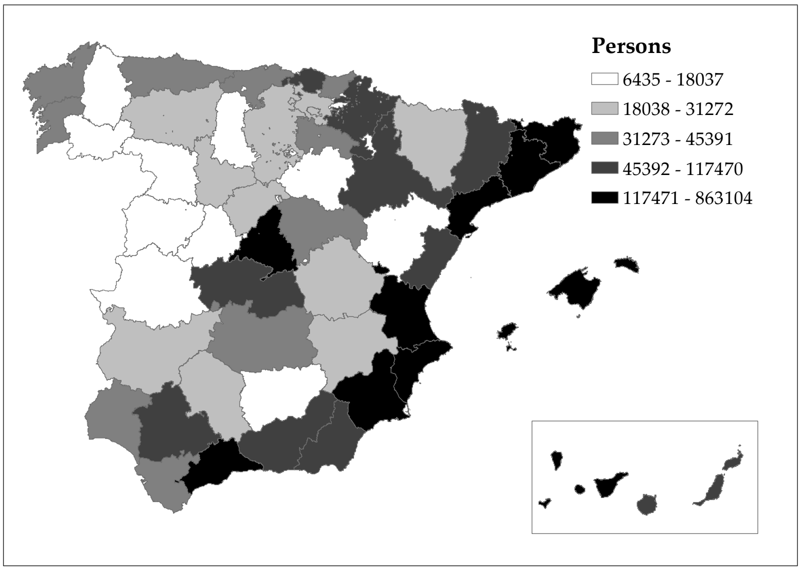
Table A.10: Descriptive statistics by year



A.1.1 Foreign-born stock 2001

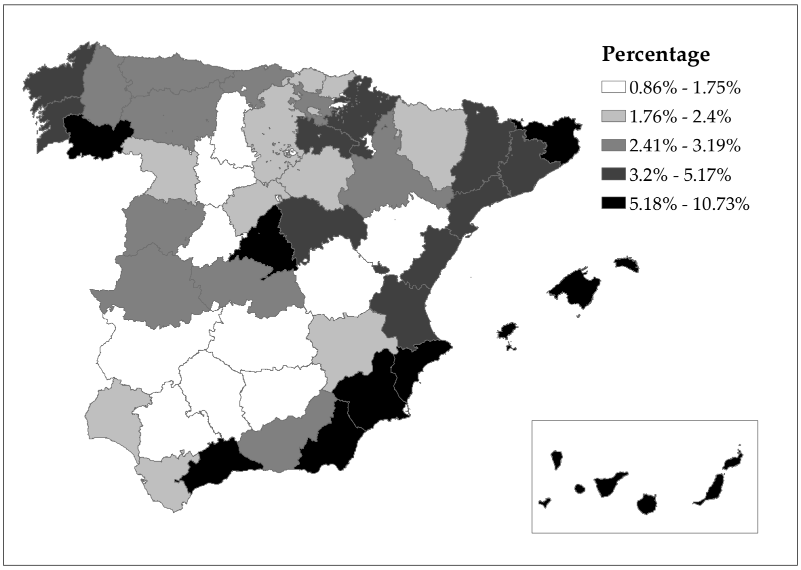


A.1.2 Foreign-born stock 2010

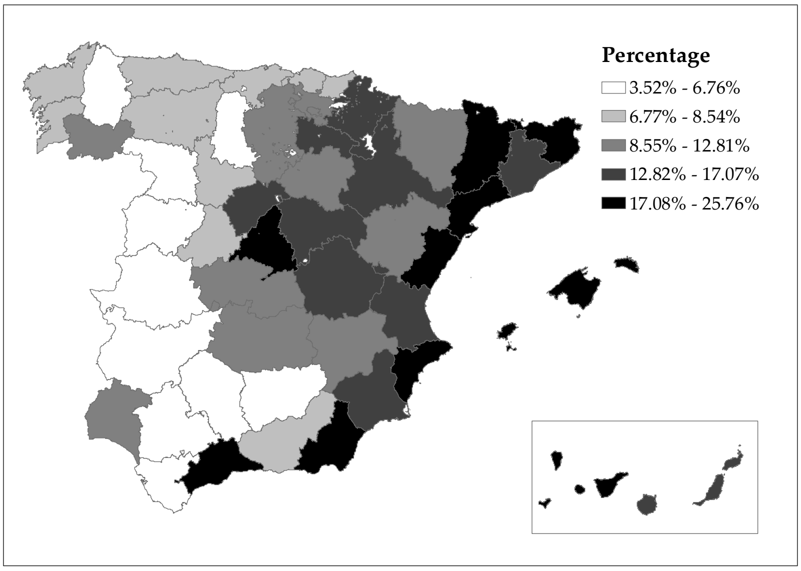


A.1.3 Change in foreign-born stock 2001-2010

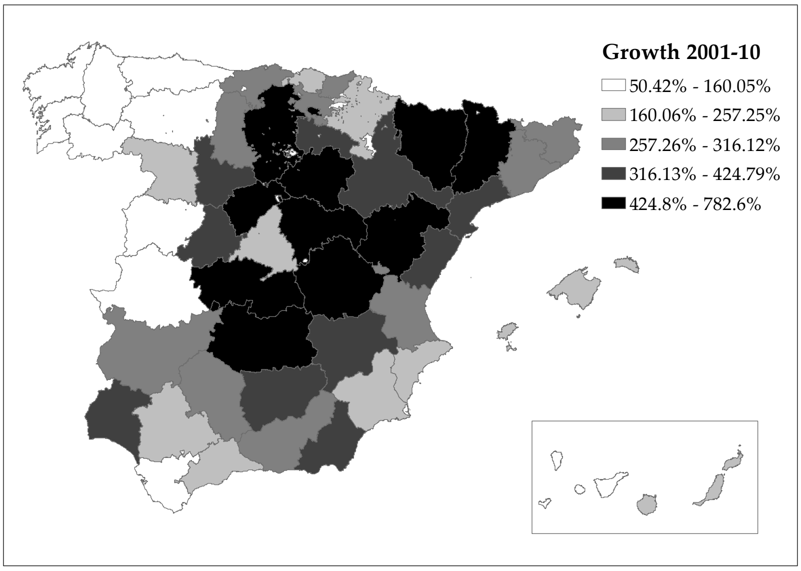
Figure A.1: Spatial distribution of foreign-born stocks



A.2.1 Share foreign-born over population 2001

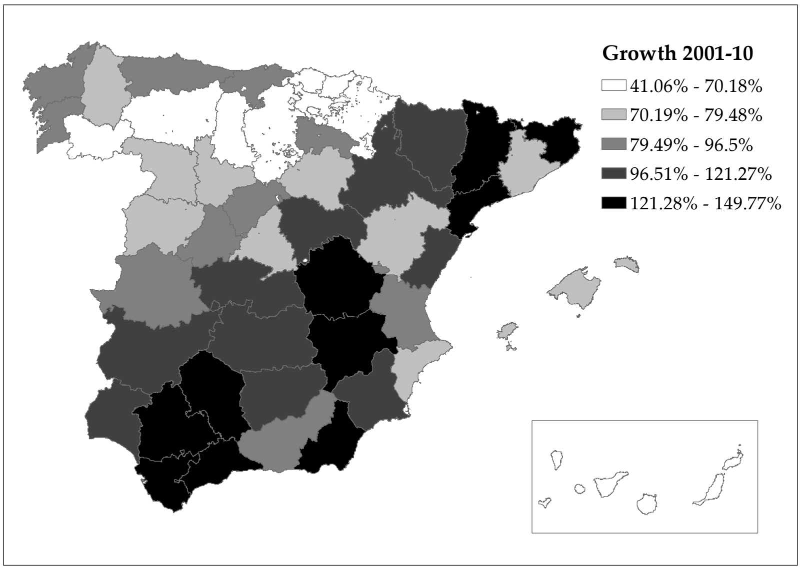


A.2.2 Share foreign-born over population 2010

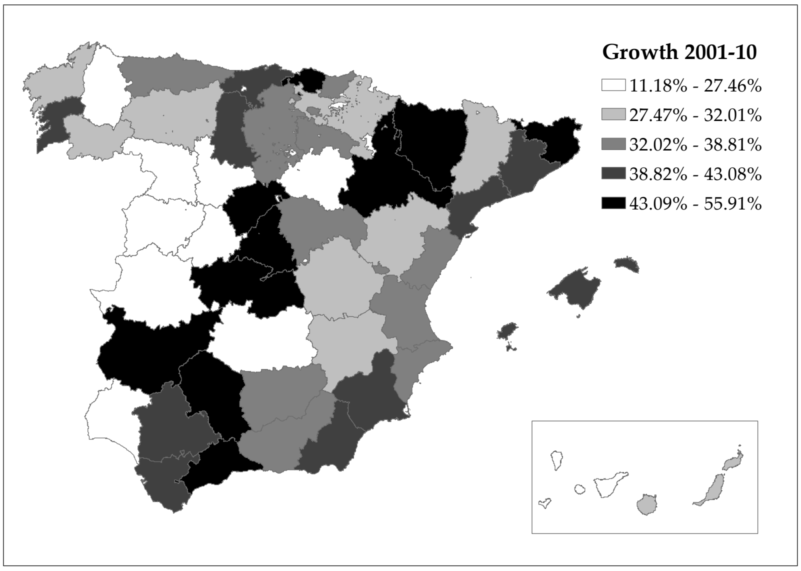


A.2.3 Growth foreign-born stocks 2001-2010

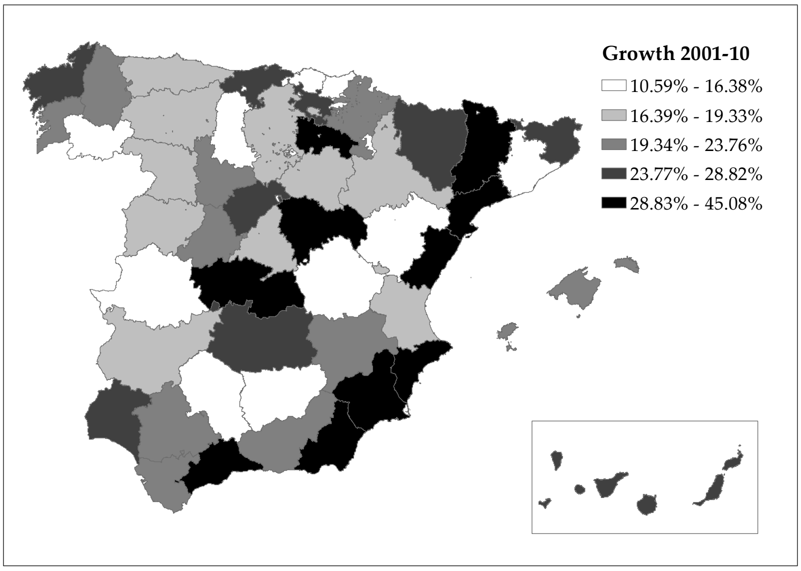
Figure A.2: Spatial distribution of share and growth of foreign-born



A.3.1 Growth of purchase prices 2001-2010



A.3.2 Growth of rental prices 2001-2010



A.3.3 Growth of housing stock 2001-2010

Figure A.3: Spatial distribution of growth in prices and construction

Inflow of natives in t over population end t-1	Region dummies			Province dummies		Region dummies			Province dummies	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of immigrants in t-1 over population end t-2	0.274***	0.239***	0.248***	0.010	-0.021	0.494**	0.449**	0.442**	0.010	-0.033
	[0.082]	[0.073]	[0.088]	[0.030]	[0.030]	[0.209]	[0.198]	[0.205]	[0.054]	[0.058]
Coast dummy	0.001	0.000	0.000			0.000	-0.000	-0.000		
	[0.002]	[0.002]	[0.002]			[0.002]	[0.002]	[0.002]		
Length of coastline	0.000***	0.000**	0.000**			0.000**	0.000**	0.000**		
	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]		
Log of hours of average temperature (January)	-0.005	-0.005	-0.005			-0.006	-0.006	-0.005		
	[0.006]	[0.006]	[0.006]			[0.006]	[0.006]	[0.005]		
Log of mm of rain precipitation (January)	0.000	0.001	0.001			0.001	0.001	0.001		
	[0.003]	[0.003]	[0.003]			[0.002]	[0.002]	[0.002]		
Log of number of retails shops in 2000	0.000	0.000	0.000			0.002	0.002	0.002		
	[0.002]	[0.002]	[0.002]			[0.003]	[0.002]	[0.003]		
Log of number of restaurants and bars in 2000	0.001	0.001	0.001			0.001	0.000	0.000		
	[0.001]	[0.001]	[0.001]			[0.001]	[0.001]	[0.001]		
Importance of tourism sector - comparative index 2000	-0.000	-0.000	-0.000			-0.000	-0.000	-0.000		
	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]		
Log of the number of doctors - 2000	0.001*	0.001*	0.001			0.001	0.001	0.001		
	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]		
Log of the surface of natural parks	-0.000	-0.000	-0.000			0.000	0.000	0.000		
	[0.001]	[0.001]	[0.001]			[0.001]	[0.001]	[0.001]		
Change of log of GDP in t-2		0.034**	0.033**	0.005	0.002		0.025**	0.027**	0.005	0.002
		[0.013]	[0.016]	[0.006]	[0.007]		[0.011]	[0.013]	[0.006]	[0.006]
Change of log of unemployment rate in t-2		0.003	0.004	0.005	0.005		-0.001	0.000	0.005	0.005
		[0.009]	[0.009]	[0.006]	[0.007]		[0.010]	[0.010]	[0.006]	[0.006]
Change of log of number of credit establishments in t-2		0.022*	0.032**	-0.019**	-0.009		0.015	0.024*	-0.019**	-0.009
		[0.013]	[0.015]	[0.008]	[0.008]		[0.012]	[0.014]	[0.008]	[0.007]
Change of percentage of savings banks in t-2		0.022	0.017	-0.009	-0.019		0.021	0.015	-0.009	-0.019
		[0.033]	[0.036]	[0.014]	[0.013]		[0.030]	[0.032]	[0.014]	[0.013]
Time invariant controls	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Time-varying controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Region trends	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Instrumented	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Observations	400	400	400	400	400	400	400	400	400	400
Adjusted R^2	0.55	0.56	0.55	0.11	0.17	0.07	0.10	0.08	-0.15	-0.23
LM test stat under-identification (K-P)						13.87	13.91	14.30	7.84	7.77
P-value of under-identification LM statistic						0.00	0.00	0.00	0.01	0.01
F-stat weak identification (K-P)						28.05	24.43	25.91	16.55	17.41

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * p<0.05, ** p<0.01, *** p<0.001. All specifications include year dummies.

Table A.11: Native displacement - lagged inflows

Inflow of natives in t over population end t-1	Region dummies			Province dummies		Region dummies			Province dummies	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of immigrants in t over population end t-1	0.256***	0.227***	0.236***	0.142***	0.124***	0.657***	0.627***	0.648***	0.398***	0.360***
	[0.063]	[0.056]	[0.070]	[0.033]	[0.033]	[0.202]	[0.190]	[0.223]	[0.099]	[0.110]
Observations	450	450	450	450	450	450	450	450	450	450
Adjusted R^2	0.64	0.66	0.66	0.46	0.52					
LM test stat under-identification (K-P)						12.27	12.05	11.93	7.71	7.51
P-value of under-identification LM statistic						0.00	0.00	0.00	0.01	0.01
F-stat weak identification (K-P)						20.58	17.73	19.81	14.72	16.67
Inflow of natives in t over population end t-1	Region dummies			Province dummies		Region dummies			Province dummies	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Inflow of immigrants in t-1 over population end t-2	0.185***	0.163***	0.181***	0.013	0.006	0.545***	0.525**	0.539**	0.190**	0.185**
	[0.060]	[0.054]	[0.065]	[0.024]	[0.027]	[0.209]	[0.208]	[0.221]	[0.081]	[0.085]
Observations	400	400	400	400	400	400	400	400	400	400
Adjusted R^2	0.63	0.64	0.64	0.50	0.54					
LM test stat under-identification (K-P)						13.76	13.82	14.09	8.42	8.30
P-value of under-identification LM statistic						0.00	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)						23.56	20.37	21.07	17.60	18.72
Time invariant controls	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Time-varying controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Region trends	No	No	Yes	No	Yes	No	No	Yes	No	Yes
Instrumented	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * p<0.05, ** p<0.01, *** p<0.001. All specifications include year dummies.

Table A.12: Native displacement - aged 16-64

Change in the log of rental prices in t	PROVINCE DUMMIES					
	(1)	(2)	(3)	(4)	(5)	(6)
Inflow of immigrants in t-1 over population in January t-1	0.126		0.283			
	[0.579]		[0.661]			
Inflow of natives in t-1 over population in January t-1		-0.253	-0.364			
		[0.311]	[0.358]			
Inflow of population in t-1 over population in January t-1				-0.125	0.088	-0.181
				[0.233]	[0.403]	[0.226]
Change in the log of purchase prices in t	PROVINCE DUMMIES					
	(1)	(2)	(3)	(4)	(5)	(6)
Inflow of immigrants in t-1 over population in January t-1	2.739**		2.773*			
	[1.233]		[1.495]			
Inflow of natives in t-1 over population in January t-1		1.013	-0.080			
		[0.722]	[1.094]			
Inflow of population in t-1 over population in January t-1				0.973**	1.912**	0.726
				[0.453]	[0.841]	[0.531]
Observations	450	450	450	450	450	450
LM test stat under-identification (Kleibergen-Paap)	7.37	26.12	7.34	25.33	8.21	25.13
P-value of under-identification LM statistic	0.01	0.00	0.01	0.00	0.00	0.00
F-stat weak identification (Kleibergen-Paap)	14.32	515.98	5.33	51.60	27.59	93.88
P-value Hansen J statistic				0.18		
A-P F-test of excluded instruments (IMM)			11.01			
A-P F-test of excluded instruments (NAT)			488.19			
Instrument(s)	IMM	NAT	BOTH	BOTH	IMM	NAT

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. N=450
All specifications include year dummies, region dummies and trends and controls.

Table A.13: Instrumental variables results rental/purchase prices - different population groups

Change in the log of purchase prices in t	REGION DUMMIES		PROVINCE DUMMIES		REGION DUMMIES		PROVINCE DUMMIES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IVIE 2002-2007				*IVIE 2000-2010*			
Inflow of immigrants in t-1 over population end t-2	1.888**	1.332	1.762*	0.582	1.871**	1.660**	1.416	1.066
	[0.930]	[0.884]	[1.006]	[1.165]	[0.864]	[0.736]	[0.940]	[0.898]
Inflow of population in t-1 over population end t-2	1.148**	0.819	1.217*	0.438	1.158**	1.035**	1.099	0.850
	[0.523]	[0.512]	[0.670]	[0.893]	[0.490]	[0.413]	[0.694]	[0.691]
Observations	300	300	300	300	550	550	550	550
	Housing Dept Annual Mean 2002-2010				*Housing Dept 2^{nd} quarter 2002-2010*			
Inflow of immigrants in t-1 over population end t-2	2.375***	2.305***	2.129**	2.038**	2.898***	2.841***	2.708**	2.681**
	[0.798]	[0.762]	[0.875]	[0.913]	[0.911]	[0.906]	[1.112]	[1.248]
Inflow of population in t-1 over population end t-2	1.438***	1.384***	1.464**	1.423**	1.755***	1.706***	1.862**	1.872**
	[0.511]	[0.488]	[0.592]	[0.633]	[0.586]	[0.586]	[0.744]	[0.854]
Observations	450	450	450	450	450	450	450	450
Region trends	No	Yes	No	Yes	No	Yes	No	Yes

Clustered (province) standard errors in brackets. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies and controls. N=450.

Table A.14: Instrumental variables results purchase prices - different prices

Change in the log of rental prices in t	REGION DUMMIES		PROVINCE DUMMIES		REGION DUMMIES		PROVINCE DUMMIES		REGION DUMMIES		PROVINCE DUMMIES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Actual population				*Only FB stock*				*Base year 1981*			
Inflow of immigrants in t-1	0.816*	0.945*	0.207	0.198	0.761*	0.858*	0.198	0.189	0.931**	1.023**	0.193	0.175
over population end t-2	[0.450]	[0.564]	[0.477]	[0.628]	[0.426]	[0.518]	[0.465]	[0.587]	[0.403]	[0.473]	[0.466]	[0.565]
Inflow of population in t-1	0.496*	0.566*	0.138	0.133	0.469*	0.526*	0.134	0.130	0.563**	0.613**	0.134	0.123
over population end t-2	[0.263]	[0.326]	[0.315]	[0.418]	[0.252]	[0.307]	[0.312]	[0.401]	[0.234]	[0.277]	[0.320]	[0.395]
Change in the log of purchase prices in t	REGION DUMMIES		PROVINCE DUMMIES		REGION DUMMIES		PROVINCE DUMMIES		REGION DUMMIES		PROVINCE DUMMIES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Actual population				*Only FB stock*				*Base year 1981*			
Inflow of immigrants in t-1	2.990***	2.840***	3.514***	3.463**	2.705***	2.556***	3.278***	3.201**	2.418***	2.115***	3.009***	2.616**
over population end t-2	[0.904]	[1.007]	[1.136]	[1.635]	[0.857]	[0.876]	[1.075]	[1.414]	[0.795]	[0.720]	[0.965]	[1.142]
Inflow of population in t-1	1.820***	1.702***	2.344***	2.319**	1.669***	1.568***	2.220***	2.203**	1.461***	1.269***	2.083***	1.841**
over population end t-2	[0.467]	[0.513]	[0.655]	[1.047]	[0.461]	[0.463]	[0.634]	[0.934]	[0.429]	[0.377]	[0.583]	[0.789]
Region trends	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Clustered (province) standard errors in brackets. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies and controls. N=450/

Table A.15: Instrumental variables results rental/purchase prices - different population prediction

Change in the log of rental prices in t	Region dummies (1)	Region dummies (2)	Province dummies (3)	Province dummies (4)	Region dummies (5)	Region dummies (6)	Province dummies (7)	Province dummies (8)
	Actual inflow				*Column 2*			
Inflow of immigrants in t-1 over population end t-2	1.345***	1.405**	0.563	0.452	1.377*	1.417**	-1.379	-0.718
	[0.471]	[0.553]	[0.521]	[0.646]	[0.726]	[0.700]	[1.561]	[0.866]
Inflow of population in t-1 over population end t-2	0.803***	0.846**	0.417	0.361	0.720*	0.774**	-0.790	-0.459
	[0.274]	[0.331]	[0.372]	[0.502]	[0.373]	[0.376]	[0.796]	[0.538]
	Column 3				*Column 4*			
Inflow of immigrants in t-1 over population end t-2	0.852**	0.944*	0.133	0.051	1.133*	1.252*	-0.740	-0.958
	[0.429]	[0.534]	[0.413]	[0.574]	[0.592]	[0.674]	[0.764]	[0.956]
Inflow of population in t-1 over population end t-2	0.575**	0.650*	0.095	0.039	0.709**	0.804**	-0.477	-0.665
	[0.263]	[0.336]	[0.294]	[0.433]	[0.341]	[0.395]	[0.474]	[0.649]
Change in the log of purchase prices in t	Region dummies (1)	Region dummies (2)	Province dummies (3)	Province dummies (4)	Region dummies (5)	Region dummies (6)	Province dummies (7)	Province dummies (8)
	Actual inflow				*Column 2*			
Inflow of immigrants in t-1 over population end t-2	2.166***	2.092***	2.640**	2.653*	1.593	1.718*	1.913	2.204
	[0.781]	[0.769]	[1.212]	[1.575]	[1.153]	[0.913]	[2.174]	[2.005]
Inflow of population in t-1 over population end t-2	1.293***	1.259***	1.956**	2.121*	0.833	0.939**	1.095	1.410
	[0.440]	[0.416]	[0.859]	[1.274]	[0.562]	[0.452]	[1.258]	[1.247]
	Column 3				*Column 4*			
Inflow of immigrants in t-1 over population end t-2	0.997	1.284*	1.500*	2.102*	0.359	0.606	0.487	1.122
	[0.728]	[0.673]	[0.870]	[1.267]	[0.924]	[0.856]	[2.816]	[3.338]
Inflow of population in t-1 over population end t-2	0.673	0.884*	1.075*	1.588	0.225	0.389	0.314	0.779
	[0.489]	[0.453]	[0.622]	[0.977]	[0.570]	[0.531]	[1.823]	[2.308]
Region trends	No	Yes	No	Yes	No	Yes	No	Yes

Clustered (province) standard errors in brackets. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies and controls. N=450.

Table A.16: Instrumental variables results rental/purchase prices - different gravity prediction

Change in the log of rental prices in t	Region dummies							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2001		*1985*		*1930*		*1940*	
Inflow of immigrants in t-1 over population end t-2	0.468**	0.470	0.170	0.802	5.936	6.063	1.677	0.421
	[0.144]	[0.241]	[0.207]	[0.471]	[1.743]	[9.001]	[0.787]	[1.573]
LM test stat under-identification (K-P)	19.27	17.32	13.18	13.06	0.03	0.01	2.94	3.28
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.86	0.90	0.09	0.07
F-stat weak identification (K-P)	106.28	119.99	26.66	25.35	0.03	0.01	6.51	7.99
Inflow of population in t-1 over population end t-2	0.324**	0.324	0.109	0.514	1.046	-1.395	1.063	0.283
	[0.220]	[0.356]	[0.322]	[0.767]	[33.029]	[70.405]	[1.206]	[2.382]
LM test stat under-identification (K-P)	19.52	17.40	13.02	12.89	0.43	0.11	2.90	2.72
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.51	0.73	0.09	0.10
F-stat weak identification (K-P)	56.87	52.30	33.32	27.41	0.44	0.10	4.63	4.10
Change in the log of rental prices in t	Region dummies							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2001		*1985*		*1930*		*1940*	
Inflow of immigrants in t-1 over population end t-2	-0.001	0.824	-0.022	1.468*	-0.003	6.216	3.152	3.368
	[0.198]	[0.502]	[0.233]	[0.636]	[0.707]	[9.216]	[1.552]	[4.081]
LM test stat under-identification (K-P)	20.97	21.30	9.49	9.29	0.38	0.02	0.52	0.86
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.54	0.89	0.47	0.35
F-stat weak identification (K-P)	52.67	83.26	22.88	19.58	0.46	0.02	0.59	1.01
Inflow of population in t-1 over population end t-2	-0.001	0.737	-0.015	1.055*	-0.001	0.964	1.425	2.101
	[0.231]	[0.557]	[0.333]	[0.880]	[1.901]	[76.671]	[4.695]	[7.162]
LM test stat under-identification (K-P)	20.66	20.08	9.75	10.05	1.38	0.59	1.28	1.05
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.24	0.44	0.26	0.31
F-stat weak identification (K-P)	52.56	85.07	44.90	30.55	2.19	0.66	1.66	1.30
Region trends	No	Yes	No	Yes	No	Yes	No	Yes

Clustered (province) standard errors in brackets. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies and controls. N=450.

Table A.17: Instrumental variables results rental prices - different base years

Change in the log of purchase prices in t	REGION DUMMIES							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2001		*1985*		*1930*		*1940*	
Inflow of immigrants in t-1 over population end t-2	0.876**	0.470	1.236*	0.802	-3.074	6.063	-0.025	0.421
	[0.403]	[0.356]	[0.741]	[0.767]	[25.340]	[70.405]	[2.732]	[2.382]
LM test stat under-identification (K-P)	19.27	17.32	13.18	13.06	0.03	0.01	2.94	3.28
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.86	0.90	0.09	0.07
F-stat weak identification (K-P)	106.28	119.99	26.66	25.35	0.03	0.01	6.51	7.99
Inflow of population in t-1 over population end t-2	0.606**	0.324	0.792*	0.514	-0.542	-1.395	-0.016	0.283
	[0.281]	[0.241]	[0.445]	[0.471]	[4.421]	[9.001]	[1.732]	[1.573]
LM test stat under-identification (K-P)	19.52	17.40	13.02	12.89	0.43	0.11	2.90	2.72
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.51	0.73	0.09	0.10
F-stat weak identification (K-P)	56.87	52.30	33.32	27.41	0.44	0.10	4.63	4.10

Change in the log of purchase prices in t	REGION DUMMIES							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2001		*1985*		*1930*		*1940*	
Inflow of immigrants in t-1 over population end t-2	1.514**	0.824	1.849**	1.468*	2.620	6.216	0.338	3.368
	[0.644]	[0.557]	[0.765]	[0.880]	[12.761]	[76.671]	[10.073]	[7.162]
LM test stat under-identification (K-P)	20.97	21.30	9.49	9.29	0.38	0.02	0.52	0.86
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.54	0.89	0.47	0.35
F-stat weak identification (K-P)	52.67	83.26	22.88	19.58	0.46	0.02	0.59	1.01
Inflow of population in t-1 over population end t-2	1.299**	0.737	1.296**	1.055*	0.974	0.964	0.153	2.101
	[0.553]	[0.502]	[0.538]	[0.636]	[4.502]	[9.216]	[4.562]	[4.081]
LM test stat under-identification (K-P)	20.66	20.08	9.75	10.05	1.38	0.59	1.28	1.05
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.24	0.44	0.26	0.31
F-stat weak identification (K-P)	52.56	85.07	44.90	30.55	2.19	0.66	1.66	1.30
Region trends	No	Yes	No	Yes	No	Yes	No	Yes

Clustered (province) standard errors in brackets. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies and controls. N=450.

Table A.18: Instrumental variables results purchase prices - different base years

Change in the log of rental prices in t	REGION DUMMIES		PROVINCE DUMMIES		REGION DUMMIES		PROVINCE DUMMIES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inflow of immigrants in t-1 over population end t-2	0.845**	0.989**	-0.423	-0.303				
	[0.385]	[0.463]	[0.512]	[0.538]				
Inflow of population in t-1 over population end t-2					0.523**	0.583**	-0.282	-0.207
					[0.235]	[0.272]	[0.339]	[0.361]
Observations	400	400	400	400	400	400	400	400
LM test stat under-identification (K-P)	14.31	14.30	10.93	9.05	14.25	14.60	11.18	9.59
P-value of under-identification LM statistic	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
F-stat weak identification (K-P)	13.35	13.52	9.59	8.93	16.38	17.30	17.64	18.94
P-value Hansen J statistic	0.72	0.97	0.22	0.09	0.90	0.79	0.20	0.09
Change in the log of purchase prices in t	REGION DUMMIES		PROVINCE DUMMIES		REGION DUMMIES		PROVINCE DUMMIES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Inflow of immigrants in t-1 over population end t-2	2.252***	1.846**	2.556**	1.988				
	[0.839]	[0.746]	[1.244]	[1.287]				
Inflow of population in t-1 over population end t-2					1.298***	1.053***	1.707**	1.331
					[0.440]	[0.377]	[0.789]	[0.857]
Observations	400	400	400	400	400	400	400	400
LM test stat under-identification (K-P)	14.31	14.30	10.93	9.05	14.25	14.60	11.18	9.59
P-value of under-identification LM statistic	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01
F-stat weak identification (K-P)	13.35	13.52	9.59	8.93	16.38	17.30	17.64	18.94
P-value Hansen J statistic	0.56	0.61	0.98	0.70	0.42	0.50	0.98	0.68
Region trends	No	Yes	No	Yes	No	Yes	No	Yes

Clustered (province) standard errors in brackets. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All specifications include year dummies and controls. N=400.

Table A.19: Instrumental variables results rental prices - 2 instruments

Change in the log of rental prices in t	Region dummies (1)	(2)	(3)	(4)	Province dummies (5)	(6)	(7)	Region dummies (8)	(9)	(10)	Province dummies (11)	(12)
Inflow of immigrants in t-1 over population end t-2	0.974**	0.711*	0.666	0.769	0.072	0.029						
	[0.393]	[0.417]	[0.461]	[0.607]	[0.409]	[0.517]						
Inflow of population in t-1 over population end t-2							0.570***	0.463*	0.473	0.548	0.050	0.020
							[0.218]	[0.272]	[0.334]	[0.445]	[0.282]	[0.360]
Coast dummy	-0.004	-0.004	-0.004	-0.004			-0.004	-0.004	-0.004	-0.004		
	[0.004]	[0.004]	[0.004]	[0.004]			[0.004]	[0.003]	[0.003]	[0.003]		
Length of coastline	0.000	0.000	-0.000	-0.000			0.000	0.000	-0.000	-0.000		
	[0.000]	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]	[0.000]		
Log of hours of average temperature (January)	0.003	0.016*	0.015*	0.015*			0.007	0.018**	0.018**	0.018**		
	[0.006]	[0.009]	[0.009]	[0.009]			[0.006]	[0.008]	[0.008]	[0.008]		
Log of mm of rain precipitation (January)	0.004	0.003	0.004	0.004			0.003	0.004	0.005	0.005		
	[0.004]	[0.005]	[0.006]	[0.006]			[0.004]	[0.005]	[0.006]	[0.006]		
Log of number of retails shops in 2000	0.008**	0.006*	0.007*	0.007*			0.006*	0.006*	0.006*	0.007*		
	[0.004]	[0.004]	[0.004]	[0.004]			[0.003]	[0.004]	[0.004]	[0.004]		
Log of number of restaurants and bars in 2000	0.002	0.002	0.002	0.002			0.002	0.002	0.001	0.001		
	[0.001]	[0.001]	[0.001]	[0.001]			[0.002]	[0.002]	[0.002]	[0.002]		
Importance of tourism sector - comparative index 2000	-0.000	-0.000**	-0.000*	-0.000*			-0.000	-0.000	-0.000	-0.000		
	[0.000]	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]	[0.000]		
Log of the number of doctors - 2000	-0.002***	-0.002**	-0.002***	-0.002***			-0.003***	-0.003**	-0.003**	-0.003**		
	[0.001]	[0.001]	[0.001]	[0.001]			[0.001]	[0.001]	[0.001]	[0.001]		
Log of the surface of natural parks - 2000	-0.001	-0.001	-0.000	-0.000			-0.001	-0.001	-0.000	-0.000		
	[0.001]	[0.002]	[0.002]	[0.002]			[0.001]	[0.002]	[0.002]	[0.002]		
Share of developable land in 2000 (Corine)		0.012	0.010	0.013				0.004	0.004	0.006		
		[0.014]	[0.017]	[0.019]				[0.013]	[0.015]	[0.016]		
Relative index of altitude		0.000	0.000	0.000				0.000**	0.000*	0.000*		
		[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]		
Relative index of ruggedness		-0.000	-0.000	-0.000				-0.000	-0.000	-0.000		
		[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]		
Percentage of rented properties in 2001		0.107*	0.099	0.101				0.084	0.081	0.081		
		[0.061]	[0.064]	[0.066]				[0.057]	[0.059]	[0.059]		
Percentage of empty homes in 2001		-0.129	-0.116	-0.111				-0.083	-0.075	-0.064		
		[0.080]	[0.073]	[0.078]				[0.103]	[0.098]	[0.113]		
Change of log of GDP in t-2			-0.016	-0.015	-0.000	-0.006			-0.028	-0.030	-0.001	-0.006
			[0.028]	[0.027]	[0.025]	[0.025]			[0.033]	[0.032]	[0.026]	[0.025]
Change of log of unemployment rate in t-2			0.048	0.045	0.050*	0.049*			0.051*	0.048	0.051*	0.049*
			[0.030]	[0.032]	[0.027]	[0.029]			[0.029]	[0.031]	[0.026]	[0.029]
Change of log of number of credit establishments in t-2			0.054	0.044	0.082*	0.084**			0.051	0.038	0.082**	0.084**
			[0.048]	[0.051]	[0.042]	[0.041]			[0.050]	[0.057]	[0.041]	[0.041]
Change of percentage of savings banks in t-2			-0.066	-0.021	-0.124*	-0.094			-0.075	-0.037	-0.124*	-0.094
			[0.071]	[0.077]	[0.066]	[0.071]			[0.072]	[0.078]	[0.067]	[0.071]
Log of change of stock of private dwellings in t-2			0.042	0.021	0.104	0.134			-0.002	-0.030	0.103	0.134
			[0.075]	[0.103]	[0.097]	[0.091]			[0.094]	[0.137]	[0.096]	[0.091]
Time invariant controls	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No
Supply time invariant controls	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No
Time-varying controls	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Region trends	No	No	No	Yes	No	Yes	No	No	No	Yes	No	Yes
Observations	450	450	450	450	450	450	450	450	450	450	450	450
LM test stat under-identification (K-P)	12.66	9.04	9.34	9.38	8.06	7.80	12.56	9.90	10.62	10.00	9.63	8.76
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)	24.50	18.80	15.67	14.82	16.76	15.63	30.12	21.04	16.52	14.30	30.29	30.28

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * p<0.05, ** p<0.01, *** p<0.001. N=450

Table A.20: Instrumental variables results rental prices - adding supply controls

Change in the log of rental prices in t	Region dummies (1)	Region dummies (2)	Province dummies (3)	Province dummies (4)	Region dummies (5)	Region dummies (6)	Province dummies (7)	Province dummies (8)
Inflow of immigrants in t-1 over population end t-2	0.734	0.834	0.049	-0.022				
	[0.520]	[0.701]	[0.401]	[0.503]				
Inflow of population in t-1 over population end t-2					0.468	0.520	0.034	-0.015
					[0.332]	[0.442]	[0.276]	[0.349]
Coast dummy	-0.004	-0.004			-0.004	-0.004		
	[0.004]	[0.004]			[0.003]	[0.003]		
Length of coastline	-0.000	-0.000			-0.000	-0.000		
	[0.000]	[0.000]			[0.000]	[0.000]		
Log of hours of average temperature (January)	0.013	0.012			0.017*	0.016*		
	[0.011]	[0.012]			[0.009]	[0.009]		
Log of mm of rain precipitation (January)	0.005	0.005			0.005	0.006		
	[0.006]	[0.006]			[0.006]	[0.006]		
Log of number of retails shops in 2000	0.007*	0.008*			0.007*	0.007*		
	[0.004]	[0.005]			[0.004]	[0.004]		
Log of number of restaurants and bars in 2000	0.002	0.002			0.001	0.001		
	[0.001]	[0.001]			[0.002]	[0.002]		
Importance of tourism sector - comparative index 2000	-0.000*	-0.000*			-0.000	-0.000		
	[0.000]	[0.000]			[0.000]	[0.000]		
Log of the number of doctors - 2000	-0.002**	-0.002**			-0.003**	-0.003**		
	[0.001]	[0.001]			[0.001]	[0.001]		
Log of the surface of natural parks - 2000	-0.001	-0.000			-0.000	-0.000		
	[0.002]	[0.002]			[0.002]	[0.002]		
Relative index of altitude	0.000	0.000			0.000*	0.000*		
	[0.000]	[0.000]			[0.000]	[0.000]		
Relative index of ruggedness	-0.000	-0.000			-0.000	-0.000		
	[0.000]	[0.000]			[0.000]	[0.000]		
Percentage of rented properties in 2001	0.091	0.089			0.076	0.072		
	[0.058]	[0.058]			[0.054]	[0.053]		
Percentage of empty homes in 2001	-0.106	-0.097			-0.072	-0.061		
	[0.081]	[0.089]			[0.102]	[0.118]		
Change of log of GDP in t-2	-0.017	-0.018	-0.000	-0.007	-0.029	-0.032	-0.000	-0.007
	[0.029]	[0.027]	[0.025]	[0.025]	[0.034]	[0.033]	[0.026]	[0.025]
Change of log of unemployment rate in t-2	0.046	0.042	0.050*	0.050*	0.050*	0.047	0.050*	0.050*
	[0.031]	[0.034]	[0.027]	[0.029]	[0.029]	[0.032]	[0.026]	[0.029]
Change of log of number of credit establishments in t-2	0.052	0.040	0.082*	0.085**	0.050	0.035	0.082**	0.085**
	[0.050]	[0.055]	[0.042]	[0.041]	[0.052]	[0.060]	[0.042]	[0.041]
Change of percentage of savings banks in t-2	-0.073	-0.030	-0.124*	-0.094	-0.077	-0.040	-0.124*	-0.094
	[0.068]	[0.072]	[0.066]	[0.070]	[0.070]	[0.075]	[0.067]	[0.070]
Log of change of stock of private dwellings in t-2	0.047	0.058	0.133	0.205**	0.025	0.041	0.133	0.204**
	[0.080]	[0.097]	[0.105]	[0.087]	[0.094]	[0.116]	[0.105]	[0.087]
Time invariant controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region trends	No	Yes	No	Yes	No	Yes	No	Yes
Observations	450	450	450	450	450	450	450	450
LM test stat under-identification (K-P)	7.72	7.20	8.01	7.76	10.03	9.16	9.55	8.68
P-value of under-identification LM statistic	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)	5.71	4.81	7.88	7.44	8.32	6.86	14.26	14.22
A-P F-test of excluded instruments (immigration)	14.82	12.85	15.70	14.92	19.74	16.00	28.30	28.61
A-P F-test of excluded instruments (supply)	63.08	58.93	51.22	47.79	59.08	55.34	51.03	48.69

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * p<0.05, ** p<0.01, *** p<0.001.
All specifications include year dummies, controls and region trends. N=450

Table A.21: Instrumental variables results rental prices - instrumenting supply

Change in the log of private housing stock in t	Region dummies				Province dummies			Region dummies			Province dummies	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Inflow of immigrants in t-1 over population end t-2	1.626**	1.241*	1.214*	1.141*	1.327*	1.232						
	[0.795]	[0.685]	[0.704]	[0.692]	[0.784]	[0.757]						
Inflow of population in t-1 over population end t-2							0.953**	0.808*	0.809*	0.763*	0.912*	0.860*
							[0.427]	[0.419]	[0.444]	[0.442]	[0.494]	[0.490]
Coast dummy	-0.001	-0.000	-0.000	-0.000			-0.001	-0.001	-0.001	-0.001		
	[0.003]	[0.002]	[0.002]	[0.002]			[0.003]	[0.003]	[0.003]	[0.003]		
Length of coastline	0.000	0.000	0.000	0.000			-0.000	0.000	0.000	0.000		
	[0.000]	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]	[0.000]		
Log of hours of average temperature (January)	0.002	0.004	0.004	0.004			0.008	0.008	0.008	0.008		
	[0.005]	[0.006]	[0.006]	[0.006]			[0.005]	[0.005]	[0.005]	[0.005]		
Log of mm of rain precipitation (January)	-0.002	-0.008***	-0.008***	-0.008***			-0.004	-0.006**	-0.006***	-0.007***		
	[0.003]	[0.003]	[0.003]	[0.003]			[0.002]	[0.003]	[0.002]	[0.002]		
Log of number of retails shops in 2000	0.003	-0.004	-0.004	-0.005			0.000	-0.004	-0.004	-0.004		
	[0.005]	[0.004]	[0.004]	[0.004]			[0.003]	[0.003]	[0.003]	[0.003]		
Log of number of restaurants and bars in 2000	0.000	0.001	0.001	0.001			-0.000	0.000	0.000	0.000		
	[0.001]	[0.001]	[0.001]	[0.001]			[0.001]	[0.001]	[0.001]	[0.001]		
Importance of tourism sector - comparative index 2000	-0.000*	-0.000**	-0.000**	-0.000**			-0.000	-0.000	-0.000	-0.000		
	[0.000]	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]	[0.000]		
Log of the number of doctors - 2000	0.000	0.002*	0.002**	0.002**			0.000	0.001	0.001	0.001		
	[0.001]	[0.001]	[0.001]	[0.001]			[0.001]	[0.001]	[0.001]	[0.001]		
Log of the surface of natural parks - 2000	0.001	0.001	0.001	0.001			0.001	0.001	0.001	0.001		
	[0.001]	[0.001]	[0.001]	[0.001]			[0.001]	[0.001]	[0.001]	[0.001]		
Share of developable land in 2000 (Corine)		0.036*	0.035*	0.034*				0.021*	0.020*	0.020		
		[0.018]	[0.018]	[0.018]				[0.012]	[0.012]	[0.012]		
Relative index of altitude		-0.000***	-0.000***	-0.000***				-0.000*	-0.000*	-0.000*		
		[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]		
Relative index of ruggedness		0.000**	0.000**	0.000**				0.000	0.000	0.000		
		[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]		
Percentage of rented properties in 2001		0.090*	0.089*	0.089*				0.050	0.050	0.053		
		[0.048]	[0.048]	[0.050]				[0.049]	[0.048]	[0.048]		
Percentage of empty homes in 2001		-0.004	0.002	-0.007				0.076	0.076	0.062		
		[0.070]	[0.062]	[0.059]				[0.102]	[0.093]	[0.088]		
Change of log of GDP in t-2			0.011	0.014	-0.015	-0.012			-0.011	-0.009	-0.023	-0.020
			[0.030]	[0.025]	[0.027]	[0.023]			[0.036]	[0.033]	[0.027]	[0.023]
Change of log of unemployment rate in t-2			-0.054*	-0.051**	-0.065**	-0.062***			-0.050*	-0.046**	-0.054**	-0.052**
			[0.030]	[0.022]	[0.027]	[0.021]			[0.029]	[0.022]	[0.025]	[0.022]
Change of log of number of credit establishments in t-2			0.011	0.005	-0.020	-0.035			0.003	-0.007	-0.014	-0.032
			[0.040]	[0.047]	[0.040]	[0.046]			[0.042]	[0.051]	[0.037]	[0.043]
Change of percentage of savings banks in t-2			0.018	0.026	-0.003	-0.001			0.002	0.002	0.001	-0.003
			[0.049]	[0.049]	[0.056]	[0.058]			[0.046]	[0.048]	[0.050]	[0.053]
Time invariant controls	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No
Supply time invariant controls	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No
Time-varying controls	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Region trends	No	No	Yes	Yes	No	Yes	No	No	Yes	Yes	No	Yes
Observations	450	450	450	450	450	450	450	450	450	450	450	450
LM test stat under-identification (K-P)	12.66	9.04	9.06	8.78	7.56	7.37	12.56	9.90	10.15	9.08	9.15	8.21
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)	24.50	18.80	17.92	18.17	14.70	14.32	30.12	21.04	20.64	18.11	26.75	27.59

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.
All specifications include year dummies, controls and region trends. N=450

Table A.22: Instrumental variables results growth housing stock

Change in the log of employment in construction in t	Region dummies				Province dummies		Region dummies				Province dummies	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Inflow of immigrants in t-1 over population end t-2	-1.615	-1.074	-0.191	-2.927	1.964	-1.522						
	[2.665]	[3.212]	[3.462]	[3.753]	[4.318]	[4.909]						
Inflow of population in t-1 over population end t-2							-0.946	-0.699	-0.127	-1.958	1.350	-1.063
							[1.556]	[2.082]	[2.307]	[2.522]	[2.955]	[3.443]
Coast dummy	-0.002	0.002	0.003	0.002			-0.002	0.003	0.003	0.003		
	[0.014]	[0.012]	[0.013]	[0.014]			[0.013]	[0.011]	[0.013]	[0.013]		
Length of coastline	0.000	0.000	0.000	0.000*			0.000	0.000	0.000	0.000*		
	[0.000]	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]	[0.000]		
Log of hours of average temperature (January)	-0.016	-0.025	-0.030	-0.016			-0.022	-0.029	-0.031	-0.026		
	[0.018]	[0.028]	[0.031]	[0.036]			[0.018]	[0.023]	[0.025]	[0.035]		
Log of mm of rain precipitation (January)	-0.005	-0.006	-0.015	-0.016			-0.003	-0.007	-0.015	-0.020		
	[0.012]	[0.012]	[0.011]	[0.013]			[0.012]	[0.015]	[0.013]	[0.017]		
Log of number of retails shops in 2000	-0.020	-0.013	-0.012	-0.027			-0.017	-0.013	-0.012	-0.027		
	[0.021]	[0.021]	[0.022]	[0.024]			[0.016]	[0.022]	[0.023]	[0.026]		
Log of number of restaurants and bars in 2000	0.008	0.008	0.011**	0.014**			0.008	0.009	0.011*	0.016**		
	[0.005]	[0.006]	[0.005]	[0.006]			[0.005]	[0.007]	[0.007]	[0.008]		
Importance of tourism sector - comparative index 2000	0.001*	0.000	0.000	0.000			0.000**	0.000	0.000	0.000		
	[0.000]	[0.000]	[0.000]	[0.000]			[0.000]	[0.000]	[0.000]	[0.000]		
Log of the number of doctors - 2000	0.001	0.001	0.004	0.007			0.002	0.002	0.004	0.009		
	[0.004]	[0.005]	[0.005]	[0.005]			[0.005]	[0.008]	[0.007]	[0.009]		
Log of the surface of natural parks - 2000	-0.006	-0.006	-0.006	-0.010			-0.005	-0.006	-0.007	-0.010		
	[0.004]	[0.006]	[0.006]	[0.007]			[0.004]	[0.006]	[0.006]	[0.007]		
Share of developable land in 2000 (Corine)		0.019	0.014	-0.025				0.032	0.017	0.010		
		[0.054]	[0.064]	[0.073]				[0.041]	[0.046]	[0.062]		
Relative index of altitude		-0.000	-0.000	-0.000				-0.000	-0.000	-0.000		
		[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]		
Relative index of ruggedness		-0.000	0.000	0.000				0.000	0.000	0.000		
		[0.000]	[0.000]	[0.000]				[0.000]	[0.000]	[0.000]		
Percentage of rented properties in 2001		-0.290*	-0.211	-0.227				-0.256	-0.205	-0.135		
		[0.158]	[0.185]	[0.242]				[0.182]	[0.203]	[0.296]		
Percentage of empty homes in 2001		-0.197	-0.365	-0.596				-0.266	-0.377	-0.772		
		[0.389]	[0.369]	[0.407]				[0.576]	[0.541]	[0.612]		
Change of log of GDP in t-2			-0.290	-0.464	-0.392	-0.568*			-0.286	-0.406	-0.404	-0.558*
			[0.305]	[0.303]	[0.307]	[0.297]			[0.341]	[0.346]	[0.314]	[0.309]
Change of log of unemployment rate in t-2			-0.449	-0.040	-0.503	-0.113			-0.450	-0.051	-0.485	- 0.126
			[0.401]	[0.376]	[0.401]	[0.373]			[0.399]	[0.377]	[0.401]	[0.373]
Change of log of number of credit establishments in t-2			-0.578*	-0.267	-0.639*	-0.366			-0.577*	-0.236	-0.630*	-0.369
			[0.329]	[0.342]	[0.342]	[0.358]			[0.341]	[0.354]	[0.336]	[0.351]
Change of percentage of savings banks in t-2			-0.634	-1.064*	-0.775	-1.341**			-0.631	-1.002*	-0.769	-1.339**
			[0.555]	[0.574]	[0.600]	[0.668]			[0.550]	[0.560]	[0.608]	[0.667]
Time invariant controls	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No
Supply time invariant controls	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No
Time-varying controls	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Region trends	No	No	Yes	Yes	No	Yes	No	No	Yes	Yes	No	Yes
Observations	450	450	450	450	450	450	450	450	450	450	450	450
LM test stat under-identification (K-P)	12.66	9.04	9.06	8.78	7.56	7.37	12.56	9.90	10.15	9.08	9.15	8.21
P-value of under-identification LM statistic	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
F-stat weak identification (K-P)	24.50	18.80	17.92	18.17	14.70	14.32	30.12	21.04	20.64	18.11	26.75	27.59

Clustered (province) standard errors in brackets. t=2002/2010. Significance levels: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.
All specifications include year dummies, controls and region trends. N=450

Table A.23: Instrumental variables results growth employment in construction

Appendix B

Economic Impacts of Transport Policy

B.1 Methodology

City size	**Distance/Travel mode**				
	Avg. distance	Tube/Train	Motor	Other	*Commuters*
More than 1,000,000	10.64	20.34%	70.33%	12.77%	*5,688,280*
1,000,000 to 500,000	11.78	3.76%	88.36%	15.72%	*1,289,221*
500,000 to 100,000	12.50	3.56%	87.40%	18.51%	*4,525,166*
100,000 to 10,000	14.19	4.09%	86.16%	18.76%	*4,536,818*
Below 10,000	16.42	2.96%	85.65%	14.22%	*2,154,226*
All people	12.76	8.88%	81.62%	16.07%	*18,193,711*

Percentages with respect to commuters. Source: *Census 2001 - CASWEB.* Other includes walking & cycling. Distance is in kms.

Table B.1: Average commuting distance & mode by city size – aged 16-74

	1998/99	**2001/02**	**2004/05**	**2007/08**
Road	3234.5	3687.8	4056.3	4766.8
£ per person	56.8	64.1	69.6	80.2
Rail	2644.3	3654.1	4279.2	5569.2
£ per person	46.4	63.5	73.4	93.7

Source: Department for Transport & Eurostat. Outturn prices: millions of £

Table B.2: Investment in infrastructure in Great Britain – 1998-2008

Year	A roads trunk	Motorway trunk	Class 1 or A	Motorway principal	Class 2 or B	Class 3 or C	Unclassified	*Major roads*	*Minor roads*	*All roads*
1997	11,798	3,333	34,558	45	30,213	84,277	223,668	49,734	338,159	387,893
1998	11,682	3,376	34,714	44	30,209	84,392	224,225	49,816	338,825	388,641
1999	11,698	3,404	34,871	45	30,205	84,509	224,783	50,018	339,497	389,515
2000	11,701	3,422	34,906	45	30,200	84,624	225,339	50,074	340,163	390,237
2001	11,369	3,431	35,285	45	30,196	84,742	225,901	50,130	340,839	390,969
2002	10,679	3,433	35,995	45	30,192	84,858	226,462	50,152	341,511	391,663
2003	9,615	3,432	37,037	46	30,188	84,976	227,048	50,130	342,212	392,342
2004	9,147	3,478	37,521	46	30,178	84,223	223,082	50,192	337,482	387,674
2005	8,682	3,466	37,974	54	30,189	84,459	223,184	50,176	337,832	388,008
2006	8,723	3,503	38,032	53	30,018	84,469	229,605	50,311	344,092	394,403
2007	8,683	3,518	38,060	41	30,265	84,423	229,889	50,302	344,577	394,879
2008	8,634	3,518	38,057	41	30,161	84,574	229,482	50,250	344,217	394,467
2009	8,596	3,519	38,173	41	30,141	84,813	229,145	50,329	344,099	394,428

Major roads include motorways and A roads. Minor road include the rest. Units: kilometres. Source: Transport Statistics Great Britain

Table B.3: Length of roads by type – 1997-2009

Commuting time	**Travel mode**				
	Rail/tube	Motor	Other	*Commuters*	*% times*
Less than 15 minutes	0.27%	79.04%	20.70%	*41,846*	*49.45%*
15-30 minutes	2.39%	88.61%	9.00%	*25,731*	*30.41%*
30-45 minutes	8.61%	87.82%	3.57%	*8,858*	*10.47%*
45-60 minutes	23.01%	75.02%	1.97%	*5,440*	*6.43%*
60-90 minutes	34.53%	64.07%	1.40%	*2,068*	*2.44%*
90-120 minutes	28.96%	67.71%	3.33%	*480*	*0.57%*
More than 120 minutes	23.27%	65.84%	10.89%	*202*	*0.24%*
Total times	4.30%	82.15%	13.55%	*84,625*	*100.00%*

Percentages with respect to commuters. Source: British Household Panel Survey. Other includes walking & cycling.

Table B.4: Commuting times by travel mode – 1997-2008

	Wards	Mean	Standard deviation	90^{th} percentile	Maximum value
1997 employment and time-varying travel times					
All	115,940	14.770	1.180	16.080	19.400
10 kms	15,279	15.450	0.880	16.490	18.590
20 kms	36,850	15.410	0.830	16.340	18.590
30 kms	53,383	15.350	0.850	16.280	18.590
Time-varying employment and 1998 travel times					
All	115,940	14.760	1.170	16.060	19.500
10 kms	15,279	15.420	0.860	16.430	18.590
20 kms	36,850	15.400	0.820	16.320	18.590
30 kms	53,383	15.330	0.840	16.260	18.590
Time-varying employment and time-varying travel times					
All	115,940	14.760	1.170	16.060	19.500
10 kms	15,279	15.430	0.860	16.430	18.590
20 kms	36,850	15.400	0.820	16.320	18.590
30 kms	53,383	15.340	0.840	16.260	18.590

Sources: Department of Transport, BSD and own authors' calculations

Table B.5: Summary statistics annual log of accessibility – 1998-2008

	Wards	Mean	Standard deviation	90^{th} percentile	Maximum value
2001 employment and time-varying travel times					
All	73,780	14.710	1.170	19.320	19.320
10 kms	6,034	15.050	0.720	17.620	17.620
20 kms	17,171	15.190	0.710	18.170	18.170
30 kms	27,216	15.260	0.790	18.190	18.190
Time-varying employment and 2001 travel times					
All	73,780	14.770	1.170	16.060	19.500
10 kms	6,034	15.090	0.730	15.810	17.800
20 kms	17,171	15.240	0.720	16.000	18.300
30 kms	27,216	15.310	0.790	16.220	18.500
Time-varying employment and time-varying travel times					
All	73,780	14.760	1.170	16.060	19.500
10 kms	6,034	15.100	0.730	15.810	17.800
20 kms	17,171	15.240	0.720	16.000	18.300
30 kms	27,216	15.310	0.790	16.220	18.590

Sources: Department of Transport, BSD and own authors' calculations

Table B.6: Summary statistics annual log of accessibility – 2002-2008

	Wards	Mean	Standard deviation	90^{th} percentile	Maximum value	Proportion of zeroes
1997 employment and time-varying travel times						
All	115,940	0.00000016%	0.800%	0.100%	46.460%	32.330%
10 kms	15,279	-0.00000030%	2.040%	0.930%	46.460%	5.390%
20 kms	36,850	0.00000062%	1.400%	0.510%	46.460%	6.160%
30 kms	53,383	0.00000038%	1.180%	0.380%	46.460%	6.060%
Time-varying employment and 1998 travel times						
All	115,940	-0.00000057%	6.000%	7.330%	70.010%	0.000%
10 kms	15,279	-0.00000100%	5.110%	6.540%	41.490%	0.000%
20 kms	36,850	-0.00000115%	5.190%	6.480%	41.490%	0.000%
30 kms	53,383	-0.00000065%	5.140%	6.480%	41.650%	0.000%
Time-varying employment and time-varying travel times						
All	115,940	-0.00000005%	6.130%	7.570%	70.010%	0.000%
10 kms	15,279	-0.00000184%	5.890%	7.250%	56.340%	0.000%
20 kms	36,850	-0.00000023%	5.600%	6.880%	56.340%	0.000%
30 kms	53,383	-0.00000010%	5.450%	6.820%	56.340%	0.000%

Sources: Department of Transport, BSD and own authors' calculations

Table B.7: Summary statistics average within changes log of accessibility – 1998-2008

	Wards	Mean	Standard deviation	90^{th} percentile	Maximum value	Proportion of zeroes
2001 employment and time-varying travel times						
All	73,780	0.00000016%	0.660%	0.060%	44.890%	43.410%
10 kms	6,034	-0.00000030%	2.250%	1.120%	44.890%	9.860%
20 kms	17,171	0.00000062%	1.360%	0.430%	44.890%	11.400%
30 kms	27,216	0.00000038%	1.090%	0.310%	44.890%	10.850%
Time-varying employment and 2002 travel times						
All	73,780	-0.00000057%	5.750%	6.620%	56.920%	0.000%
10 kms	6,034	-0.00000100%	4.880%	5.840%	24.640%	0.000%
20 kms	17,171	-0.00000115%	5.060%	5.940%	35.570%	0.000%
30 kms	27,216	-0.00000065%	5.010%	5.890%	35.570%	0.000%
Time-varying employment and time-varying travel times						
All	73,780	-0.00000005%	5.860%	6.750%	56.920%	0.000%
10 kms	6,034	-0.00000184%	5.970%	6.640%	51.320%	0.000%
20 kms	17,171	-0.00000023%	5.530%	6.400%	51.320%	0.000%
30 kms	27,216	-0.00000010%	5.340%	6.210%	51.320%	0.000%

Sources: Department of Transport, BSD and own authors' calculations

Table B.8: Summary statistics average within changes log of accessibility – 2002-2008

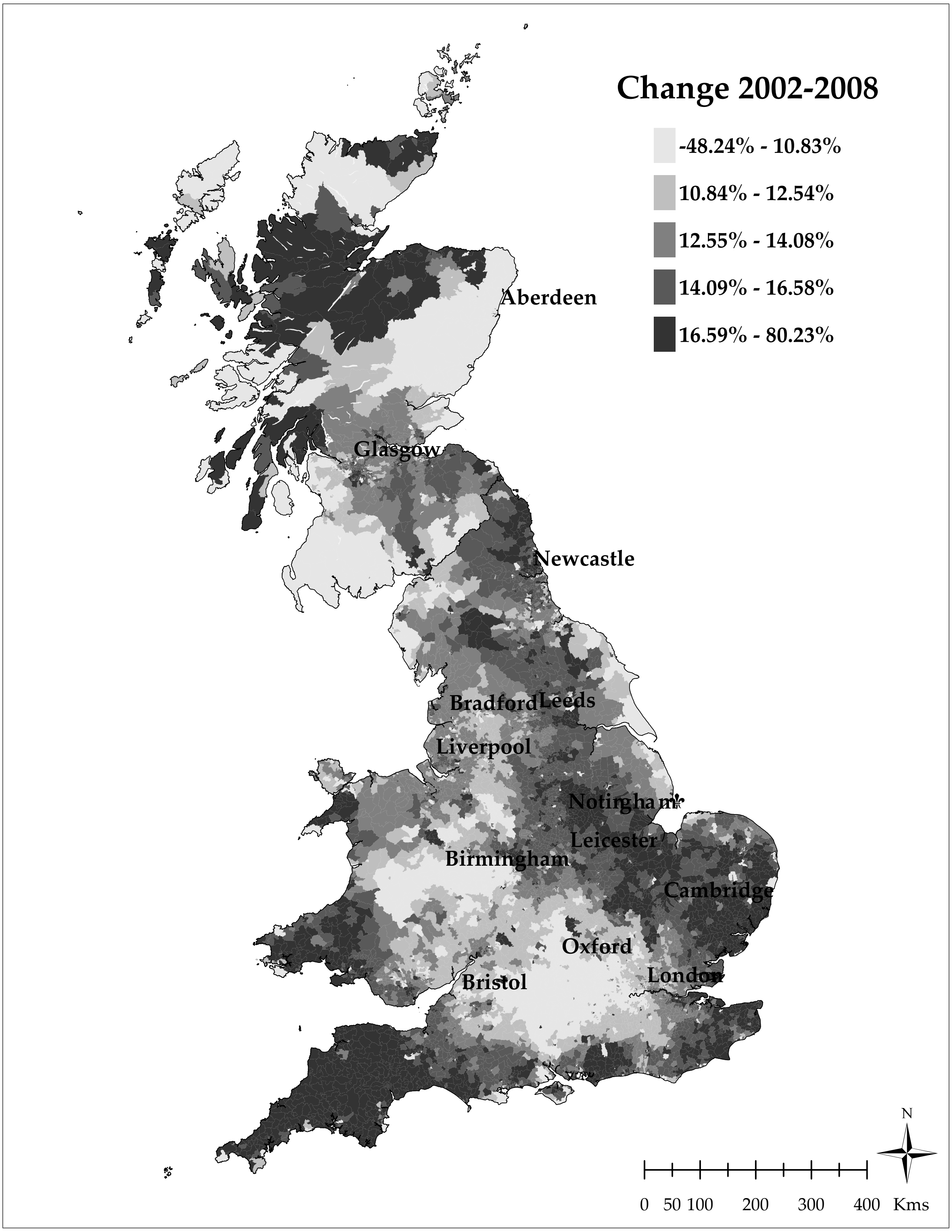


Figure B.1: Changes in log accessibility (annual employment) – 2002-2008

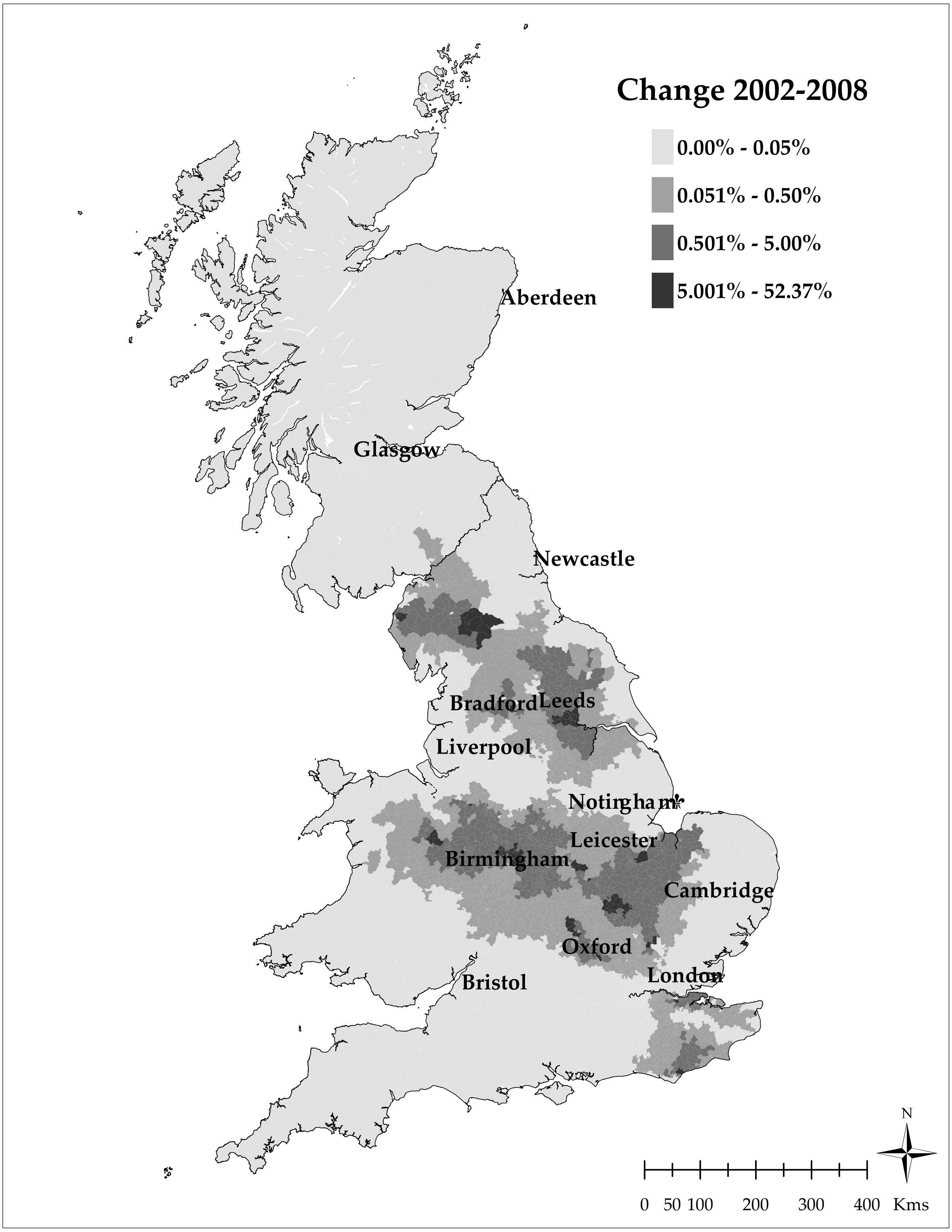


Figure B.2: Changes in log accessibility (2001 employment) – 2002-2008

Figure B.3: Changes in log accessibility (2001 employment) – 2002-2008

B.2 Aggregate and firm outcomes

DISTANCE BAND	10 kilometres		20 kilometres		30 kilometres	
Log of accessibility	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
1997 employment	15.443	0.858	15.413	0.818	15.346	0.836
Time-varying employment	15.454	0.873	15.423	0.826	15.353	0.842
Employment	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ALL SECTORS	3,730.06	9,639.28	3,155.12	7,266.17	2,878.74	6,324.49
MANUFACTURING	472.21	942.05	435.18	897.29	408.73	852.58
CONSTRUCTION	163.29	298.02	150.93	306.18	143.25	287.25
CONSUMER SERVICES	796.05	2,102.82	710.96	1,615.40	656.79	1,430.97
PRODUCER SERVICES	1,034.15	4,613.95	767.92	3,294.65	674.21	2,811.93
OTHER	1,264.36	2,975.75	1,090.14	2,354.89	995.76	2,076.73
Local units	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
ALL SECTORS	313.86	635.59	285.01	459.20	267.28	401.94
MANUFACTURING	24.95	38.02	22.33	33.72	20.71	30.12
CONSTRUCTION	23.76	17.81	23.70	17.27	23.35	17.56
CONSUMER SERVICES	84.72	142.16	76.99	111.97	71.87	100.08
PRODUCER SERVICES	104.31	352.54	91.39	244.38	84.29	215.09
OTHER	76.13	128.94	70.61	93.57	67.06	81.56

Sources: BSD-ONS via SDS and own author calculations.

Table B.9: Summary statistics ward employment and number of plants – 1998-2008

	(1)	(2)	(3)	(4)
ALL SECTORS				
Coefficient	0.263***	0.024	-0.06	-0.086
Std. Error	0.009	0.023	0.181	0.226
No. observations	101960	101960	100255	99935
Kleibergen-Paap F-stat			321	257
MANUFACTURING				
Coefficient	0.062***	0.035	-0.297	-0.45
Std. Error	0.014	0.042	0.302	0.319
No. observations	33183	33183	32908	32839
Kleibergen-Paap F-stat			220	243
CONSTRUCTION				
Coefficient	0.162***	0.027	0.582	0.695
Std. Error	0.027	0.064	0.749	1.168
No. observations	5351	5351	5294	5263
Kleibergen-Paap F-stat			30	26.8
CONSUMER SERVICES				
Coefficient	0.268***	0.03	-0.052	0.174
Std. Error	0.016	0.046	0.341	0.47
No. observations	26848	26848	26331	26229
Kleibergen-Paap F-stat			93.8	83
PRODUCER SERVICES				
Coefficient	0.423***	0.031	-0.408	-0.742
Std. Error	0.02	0.043	0.56	1.539
No. observations	19527	19527	19082	18997
Kleibergen-Paap F-stat			13.1	3.09
OTHER				
Coefficient	0.338***	-0.005	0.195	0.171
Std. Error	0.027	0.061	0.333	0.366
No. observations	17051	17051	16640	16607
Kleibergen-Paap F-stat			42.9	30
Distance band	20 kms	20 kms	20 kms	20 kms
Year fixed effects	YES	YES	YES	YES
Plant fixed effects		YES	YES	YES
IV			YES	YES
Scheme trends				YES

Sources: BSD and ARD-ONS via SDS. Standard errors clustered at firm level. * p<0.1, ** p<0.05, *** p<0.01

Table B.10: Firm labour productivity results

	(1)	(2)	(3)	(4)
ALL SECTORS				
Coefficient	0.242***	0.030*	0.136	0.093
Std. Error	0.006	0.015	0.104	0.125
No. observations	106501	106501	105533	105203
Kleibergen-Paap F-stat			352	270
MANUFACTURING				
Coefficient	0.083***	0.046**	0.145	0.151
Std. Error	0.008	0.023	0.154	0.159
No. observations	33624	33624	33460	33391
Kleibergen-Paap F-stat			240	267
CONSTRUCTION				
Coefficient	0.132***	0.015	0.412	0.435
Std. Error	0.018	0.037	0.349	0.608
No. observations	5365	5365	5316	5285
Kleibergen-Paap F-stat			30.7	28.1
CONSUMER SERVICES				
Coefficient	0.253***	0.047	0.216	0.19
Std. Error	0.011	0.035	0.197	0.245
No. observations	27598	27598	27327	27221
Kleibergen-Paap F-stat			93.1	80.2
PRODUCER SERVICES				
Coefficient	0.378***	0.03	-0.402	-0.904
Std. Error	0.015	0.027	0.398	1.403
No. observations	19871	19871	19585	19497
Kleibergen-Paap F-stat			14.2	1.97
OTHER				
Coefficient	0.277***	0.001	0.118	-0.155
Std. Error	0.013	0.04	0.26	0.274
No. observations	20043	20043	19845	19809
Kleibergen-Paap F-stat			59.2	37
Distance band	20 kms	20 kms	20 kms	20 kms
Year fixed effects	YES	YES	YES	YES
Plant fixed effects		YES	YES	YES
IV			YES	YES
Scheme trends				YES

Sources: BSD and ARD-ONS via SDS. Standard errors clustered at firm level. * p<0.1, ** p<0.05, *** p<0.01

Table B.11: Firm wage bill per worker results

	(1)	(2)	(3)	(4)
ALL SECTORS				
Coefficient	0.093***	-0.002	0.092	0.013
Std. Error	0.004	0.015	0.115	0.141
No. observations	60543	60543	53323	53166
Kleibergen-Paap F-stat			157	123
MANUFACTURING				
Coefficient	0.033***	0.021	-0.059	-0.136
Std. Error	0.006	0.022	0.146	0.151
No. observations	21081	21081	19318	19286
Kleibergen-Paap F-stat			151	200
CONSTRUCTION				
Coefficient	0.064***	0.004	0.273	0.337
Std. Error	0.013	0.058	0.222	0.387
No. observations	3373	3373	2938	2916
Kleibergen-Paap F-stat			24	13.2
CONSUMER SERVICES				
Coefficient	0.079***	0.003	0	0.016
Std. Error	0.007	0.024	0.168	0.221
No. observations	15561	15561	13165	13126
Kleibergen-Paap F-stat			113	75
PRODUCER SERVICES				
Coefficient	0.183***	-0.015	-0.031	1.757
Std. Error	0.012	0.029	0.334	10.488
No. observations	11418	11418	9861	9813
Kleibergen-Paap F-stat			1.54	0.0444
OTHER				
Coefficient	0.122***	-0.013	0.518	0.452
Std. Error	0.012	0.032	0.449	0.588
No. observations	9110	9110	8041	8025
Kleibergen-Paap F-stat			34.5	17
Distance band	20 kms	20 kms	20 kms	20 kms
Year fixed effects	YES	YES	YES	YES
Plant fixed effects		YES	YES	YES
IV			YES	YES
Scheme trends				YES

Sources: BSD and ARD-ONS via SDS. Standard errors clustered at firm level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table B.12: Firm gross output results

B.3 Labour market outcomes

YEAR	MANUFACT.	CONSTRUCT.	CONSUMER SERVICES	PRODUCER SERVICES	OTHER	ALL SECTORS
2002	7,638	1,615	7,331	7,380	7,249	31,213
2003	7,871	1,747	8,579	8,582	8,027	34,806
2004	7,336	1,706	9,375	8,371	8,147	34,935
2005	7,445	1,692	10,300	8,911	8,764	37,112
2006	6,765	1,921	10,249	9,267	9,087	37,289
2007	4,806	1,800	8,671	9,392	7,122	31,791
2008	4,773	1,850	8,429	9,244	7,094	31,390
Total	46,634	12,331	62,934	61,147	55,490	238,536

Source: ASHE–ONS. Observations within 30 kms of home and work wards.

Table B.13: Number of observations by year and industrial sector

OBSERVATIONS	**Individual home-work ward id**	**Individuals allowed to move wards**
Total	238,536	238,536
Number of individuals	114,370	79,286
Average number of T	2.09	3.01

Source: ASHE–ONS. Observations within 30 kms of home and work wards.

Table B.14: Number of individuals for each panel

VARIABLE		MEAN	STANDARD DEVIATION	
			Individual home-work ward id	**Individuals allowed to move wards**
Basic hourly earnings	*Overall*	10.75	6.96	6.96
	Between		6.8	6.64
	Within		1.59	2.11
Basic weekly pay	*Overall*	372.34	263.11	263.11
	Between		259.58	255.5
	Within		48.3	69
Basic weekly hours	*Overall*	33.78	9.33	9.33
	Between		9.81	9.86
	Within		2.58	3.65
Hourly earnings	*Overall*	11.21	7.3	7.3
	Between		7.18	6.89
	Within		1.87	2.54
Gross weekly pay	*Overall*	402.95	277.5	277.5
	Between		272.52	266.67
	Within		62.08	84.54
Total weekly hours	*Overall*	35.29	10.37	10.37
	Between		10.64	10.65
	Within		3.21	4.27
Log of basic hourly earnings	*Overall*	2.23	0.51	0.51
	Between		0.51	0.5
	Within		0.11	0.15
Log of basic weekly pay	*Overall*	5.69	0.73	0.73
	Between		0.76	0.76
	Within		0.14	0.22
Log of basic weekly hours	*Overall*	3.46	0.41	0.41
	Between		0.45	0.46
	Within		0.12	0.17
Log of hourly earnings	*Overall*	2.27	0.51	0.51
	Between		0.51	0.5
	Within		0.12	0.16
Log of gross weekly pay	*Overall*	5.77	0.72	0.72
	Between		0.75	0.75
	Within		0.15	0.22
Log of total weekly hours	*Overall*	3.5	0.42	0.42
	Between		0.45	0.45
	Within		0.12	0.17
Log of accessibility (timevar employment) from work ward	*Overall*	15.59	0.71	0.71
	Between		0.72	0.71
	Within		0.03	0.13
Log of accessibility (timevar employment) from home ward	*Overall*	15.5	0.64	0.64
	Between		0.64	0.65
	Within		0.03	0.1
Log of accessibility (2001 employment) from work ward	*Overall*	15.52	0.7	0.7
	Between		0.72	0.71
	Within		0.01	0.12
Log of accessibility (2001 employment) from home ward	*Overall*	15.43	0.63	0.63
	Between		0.64	0.64
	Within		0.01	0.09
Travel time *Between* work and home wards	*Overall*	0.26	0.24	0.24
	Between		0.25	0.23
	Within		0.00	0.09
Log of travel time *Between* work and home wards	*Overall*	-1.95	1.33	1.33
	Between		1.34	1.28
	Within		0.01	0.51

Source: ASHE–ONS, CASWEB and authors' own calculations. Observations within 30 kms of home and work wards.

Table B.15: Number of individuals for each panel

VARIABLE		MEAN	STANDARD DEVIATION	
			Individual home-work id	**Individuals allowed to move wards**
Distance to closest improvements from work ward	*Overall*	16.45	7.45	7.45
	Between		7.48	7.34
	Within		0.00	1.87
Distance to closest improvements from home ward	*Overall*	15.97	7.44	7.44
	Between		7.46	7.42
	Within		0.00	1.33
Proportion of females	*Overall*	0.45	0.5	0.5
	Between		0.5	0.5
	Within		0.00	0
Age	*Overall*	39.19	11.97	11.97
	Between		12.08	12.45
	Within		1.17	1.53
Work ward mean population age in 2001	*Overall*	38.23	3.33	3.33
	Between		3.34	3.19
	Within		0.00	1.17
Work ward proportion of high degrees in 2001	*Overall*	0.22	0.13	0.13
	Between		0.13	0.13
	Within		0.00	0.03
Work ward unemployment rate in 2001	*Overall*	0.07	0.04	0.04
	Between		0.04	0.04
	Within		0.00	0.01
Work ward proportion of living in social housing in 2001	*Overall*	0.24	0.15	0.15
	Between		0.15	0.15
	Within		0.00	0.05
Work ward density (addresses over area) in 2001	*Overall*	1654.06	1649.58	1649.58
	Between		1727.82	1655.39
	Within		0.00	514.62
Work ward proportion of commuters by motor in 2001	*Overall*	0.66	0.18	0.18
	Between		0.19	0.18
	Within		0.00	0.04
Work ward average distance traveled *Between* work and home 2001	*Overall*	11.64	4.02	4.02
	Between		4.03	3.92
	Within		0.00	1.11
Home ward mean population age in 2001	*Overall*	38.52	3.18	3.18
	Between		3.23	3.14
	Within		0.00	0.91
Home ward proportion of high degrees in 2001	*Overall*	0.19	0.1	0.1
	Between		0.11	0.1
	Within		0.00	0.02
Home ward unemployment rate in 2001	*Overall*	0.06	0.03	0.03
	Between		0.03	0.03
	Within		0.00	0.01
Home ward proportion of living in social housing in 2001	*Overall*	0.21	0.14	0.14
	Between		0.15	0.14
	Within		0.00	0.04
Home ward density (addresses over area) in 2001	*Overall*	1497.02	1458.33	1458.33
	Between		1559.06	1531.83
	Within		0.00	367
Home ward proportion of commuters by motor in 2001	*Overall*	0.73	0.13	0.13
	Between		0.14	0.13
	Within		0.00	0.03
Home ward average distance traveled *Between* work and home 2001	*Overall*	12.53	3.75	3.75
	Between		3.74	3.7
	Within		0.00	0.83

Source: ASHE–ONS, CASWEB and authors' own calculations. Observations within 30 kms of home and work wards.

Table B.16: Number of individuals for each panel

PANEL A: **Log of basic weekly pay**					
	(1)	(2)	(3)	(4)	(5)
Log of accessibility	0.320**	0.502***	0.232	0.033	0.326
from work ward	[0.141]	[0.189]	[0.281]	[0.261]	[0.202]
Log of accessibility	-0.071	0.119	0.003	-0.201	-0.032
from home ward	[0.108]	[0.174]	[0.186]	[0.221]	[0.214]
Log of travel time	0.056*	0.097**	0.01	-0.035	0.173**
between work and home	[0.032]	[0.049]	[0.073]	[0.082]	[0.075]
Observations	38,421	8,834	44,614	42,498	42,309
Kleibergen-Paap F-stat	879	881	784	120	1,079
Work-ward no of clusters	1,851	1,245	2,539	2,085	2,754
Home-ward no of clusters	2,983	1,694	3,316	3,172	3,322
PANEL B: **Log of total weekly hours worked**					
	(1)	(2)	(3)	(4)	(5)
Log of accessibility	0.049	0.461	0.422	0.115	0.782
from work ward	[0.127]	[0.390]	[0.379]	[0.266]	[0.516]
Log of accessibility	0.195**	0.106	0.131	-0.2	-1.315
from home ward	[0.096]	[0.166]	[0.206]	[0.179]	[1.037]
Log of travel time	-0.043	0.101	0.145	-0.065	0.088
between work and home	[0.039]	[0.114]	[0.251]	[0.067]	[0.089]
Observations	38,452	8,802	44,587	42,997	42,122
Kleibergen-Paap F-stat	837	863	792	120	1,065
Work-ward no of clusters	1,851	1,238	2,535	2,082	2,749
Home-ward no of clusters	2,990	1,694	3,316	3,183	3,326
Group	MANUF	CONST	CONS S	PROD S	OTHER

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ASHE–ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table B.17: Robustness: by industrial sector – 2002-2008

PANEL A: Log of basic weekly pay				
	(1)	(2)	(3)	(4)
Log of accessibility	-0.133	0.474***	0.325*	0.274
from work ward	[0.406]	[0.138]	[0.184]	[0.211]
Log of accessibility	0.17	-0.086	0.037	-0.143
from home ward	[0.303]	[0.139]	[0.130]	[0.140]
Log of travel time	0.176**	-0.006	0.026	0.117*
between work and home	[0.075]	[0.069]	[0.040]	[0.067]
Observations	20,702	75,671	38,322	39,681
Kleibergen-Paap F-stat	877	925	598	601
Work-ward no of clusters	2,140	3,033	2,545	2,363
Home-ward no of clusters	2,604	3,588	3,196	3,173
PANEL B: Log of total weekly hours worked				
	(1)	(2)	(3)	(4)
Log of accessibility	0.322	0.323***	0.378	0.584
from work ward	[0.394]	[0.113]	[0.286]	[0.424]
Log of accessibility	-0.286	-0.012	0.049	-0.707
from home ward	[0.337]	[0.150]	[0.145]	[0.706]
Log of travel time	-0.029	-0.082*	0.358	0.049
between work and home	[0.063]	[0.047]	[0.267]	[0.064]
Observations	20,439	75,458	38,299	40,471
Kleibergen-Paap F-stat	874	921	604	569
Work-ward no of clusters	2,131	3,029	2,543	2,368
Home-ward no of clusters	2,590	3,586	3,198	3,187
Group	Skill1	Skill2	Skill3	Skill4

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ASHE–ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table B.18: Robustness: by skill level (4 categories) – 2002-2008

PANEL A: **Log of basic weekly pay**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of accessibility	0.354***	0.322***	0.322**	0.329**	0.468***	0.182	0.219	0.427***
from work ward	[0.123]	[0.120]	[0.154]	[0.166]	[0.124]	[0.209]	[0.177]	[0.101]
Log of accessibility	0.035	0.075	-0.084	-0.059	-0.114	0.131	0.045	-0.086
from home ward	[0.114]	[0.125]	[0.130]	[0.116]	[0.111]	[0.118]	[0.116]	[0.108]
Log of travel time	0.062	0.033	0.038	0.059	0.070**	0.051	0.042	0.054
between work and home	[0.049]	[0.042]	[0.038]	[0.044]	[0.033]	[0.035]	[0.050]	[0.044]
Observations	27,759	24,485	152,346	155,620	50,087	41,477	130,018	138,628
Kleibergen-Paap F-stat	1,461	686	518	458	1,977	1,186	520	329
Work-ward no of clusters	1,517	860	3,040	2,691	2,790	1,505	3,073	2,046
Home-ward no of clusters	969	1,557	2,855	3,481	1,624	3,215	2,200	3,635
PANEL B: **Log of total weekly hours worked**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of accessibility	0.007	0.052	0.502**	0.515***	0.483	0.169	0.253**	0.510**
from work ward	[0.150]	[0.173]	[0.205]	[0.200]	[0.312]	[0.154]	[0.129]	[0.217]
Log of accessibility	0.122	0.18	-0.377	-0.317	-0.373	0.103	-0.049	-0.329
from home ward	[0.137]	[0.196]	[0.323]	[0.286]	[0.437]	[0.160]	[0.098]	[0.302]
Log of travel time	0.023	-0.081	0.041	0.055	0.097	-0.018	-0.032	0.058
between work and home	[0.057]	[0.060]	[0.084]	[0.084]	[0.095]	[0.036]	[0.043]	[0.099]
Observations	27,722	24,372	152,670	156,020	50,315	41,427	130,077	138,965
Kleibergen-Paap F-stat	1,456	685	511	456	1,974	1,191	511	328
Work-ward no of clusters	1,518	858	3,039	2,690	2,786	1,503	3,068	2,045
Home-ward no of clusters	968	1,553	2,858	3,489	1,624	3,217	2,202	3,645
Home median	L access	All	H access	All	L density	All	H density	All
Work median	All	L access	All	H access	All	L density	All	H density

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ASHE–ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table B.19: Robustness: by initial level of accessibility and density (medians) – 2002-2008

PANEL A: **Log of basic weekly pay**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of accessibility	0.282**	0.360**	0.353**	0.313**	0.304*	0.352**	0.290*	0.272*
from work ward	[0.116]	[0.150]	[0.173]	[0.158]	[0.168]	[0.150]	[0.148]	[0.159]
Log of accessibility	-0.113	0.082	0.06	-0.063	0.237*	-0.096	-0.094	0.06
from home ward	[0.087]	[0.149]	[0.137]	[0.108]	[0.134]	[0.138]	[0.116]	[0.113]
Log of travel time	0.105***	0.038	0.003	0.07	0.088	0.143**	0.037	0.011
between work and home	[0.031]	[0.031]	[0.054]	[0.058]	[0.059]	[0.064]	[0.036]	[0.026]
Observations	75,556	57,996	104,549	122,109	78,798	111,394	101,307	68,711
Kleibergen-Paap F-stat	1,285	2,102	515	451	474	190	989	1,464
Work-ward no of clusters	2,995	1,794	2,842	1,757	2,815	1,852	2,752	1,699
Home-ward no of clusters	1,933	3,329	1,891	3,494	1,949	3,696	1,875	2,803
PANEL B: **Log of total weekly hours worked**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of accessibility	0.246*	0.252	0.497	0.442**	0.260*	0.583*	0.372*	0.168
from work ward	[0.145]	[0.207]	[0.305]	[0.204]	[0.152]	[0.301]	[0.201]	[0.128]
Log of accessibility	-0.105	-0.174	-0.353	-0.236	0.288**	-0.375	-0.316	0.027
from home ward	[0.107]	[0.196]	[0.488]	[0.310]	[0.144]	[0.421]	[0.281]	[0.109]
Log of travel time	0.055	0.091	-0.004	-0.022	-0.015	0.191	0.023	-0.045
between work and home	[0.084]	[0.104]	[0.065]	[0.043]	[0.046]	[0.159]	[0.078]	[0.035]
Observations	75,929	57,975	104,463	122,417	79,363	111,838	101,029	68,554
Kleibergen-Paap F-stat	1,288	2,402	509	448	478	190	993	1,488
Work-ward no of clusters	2,996	1,791	2,837	1,757	2,812	1,850	2,751	1,698
Home-ward no of clusters	1,934	3,333	1,892	3,508	1,951	3,699	1,875	2,809
Home median	L unrate	All	H unrate	All	L motorc	All	H motorc	All
Work median	All	L unrate	All	H unrate	All	L motorc	All	H motorc

2-way clustering (work-ward and home-ward). All specifications include sector-year dummies, μ_{ihw} fixed-effects, instrument and all the controls and trends. Source: ASHE–ONS, DfT, CASWEB and authors own calculations. * p<0.1, ** p<0.05, *** p<0.01

Table B.20: Robustness: by initial level of unemployment rate and commuters by motor (medians) – 2002-2008

	Depvar: **Log of travel time to workplace**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log of accessibility (1997 employment)	0.0585	0.0528	0.0115	-0.0163	-0.0042	-0.0125	0.1219	0.0845	
	0.0066***	0.0060***	0.0046*	0.0039***	0.0039	0.0040**	0.1223	0.1261	
Log of accessibility (annual employment)									-0.0569
									0.041
Student		-0.2406	-0.0566	-0.0599	-0.0643	-0.0641	-0.066	-0.0657	-0.0658
		0.0165***	0.0168***	0.0161***	0.0161***	0.0161***	0.0161***	0.0161***	0.0161***
Skill-other		0.0389	-0.0087	-0.0015	-0.0021	-0.0016	-0.0016	-0.0013	-0.0013
		0.0096***	0.0089	0.0086	0.0085	0.0085	0.0086	0.0086	0.0086
Skill-O-level		0.0895	-0.0075	0.0055	0.0041	0.0049	0.0036	0.0041	0.0041
		0.0099***	0.0092	0.0089	0.0089	0.0089	0.0089	0.0089	0.0089
Skill-teach & A-level		0.1891	0.0267	0.0363	0.0334	0.0352	0.0327	0.0334	0.0334
		0.0106***	0.0101**	0.0097***	0.0096***	0.0096***	0.0097***	0.0097***	0.0097***
Skill-high		0.3443	0.1042	0.0996	0.0952	0.0967	0.0943	0.0948	0.0948
		0.0108***	0.0107***	0.0103***	0.0103***	0.0102***	0.0103***	0.0103***	0.0103***
House- rented from a gov scheme		0.0538	0.1447	0.1259	0.1198	0.1196	0.1256	0.1254	0.1254
		0.0099***	0.0091***	0.0087***	0.0088***	0.0088***	0.0089***	0.0089***	0.0089***
House- privately rented		0.0531	0.0777	0.0555	0.0527	0.0521	0.0551	0.0549	0.0549
		0.0105***	0.0099***	0.0090***	0.0090***	0.0089***	0.0092***	0.0092***	0.0092***
Civil status-married		-0.0589	-0.065	-0.0469	-0.0474	-0.0463	-0.0452	-0.0453	-0.0453
		0.0084***	0.0075***	0.0071***	0.0071***	0.0071***	0.0072***	0.0072***	0.0072***
Civil status-separated		-0.0509	-0.0588	-0.0418	-0.0414	-0.0403	-0.0384	-0.039	-0.039
		0.0162**	0.0143***	0.0140**	0.0140**	0.0139**	0.0141**	0.0141**	0.0140**
Civil status-divorced		-0.0412	-0.055	-0.0359	-0.0349	-0.0337	-0.0343	-0.0344	-0.0344
		0.0114***	0.0100***	0.0098***	0.0098***	0.0098***	0.0098***	0.0098***	0.0098***
Civil status-widowed		-0.1187	-0.0999	-0.0793	-0.0778	-0.0789	-0.0757	-0.0755	-0.0754
		0.0273***	0.0248***	0.0247**	0.0247**	0.0246**	0.0250**	0.0250**	0.0250**
Family status-1 adult,1+ children		-0.0303	0.0054	-0.0261	-0.0263	-0.0242	-0.0366	-0.0327	-0.0329
		0.0592	0.0524	0.0499	0.0495	0.0491	0.0501	0.05	0.05
Family status-2 adults,no children		0.0144	0.0035	0.0124	0.012	0.0135	0.0139	0.0141	0.014
		0.0091	0.0081	0.0078	0.0078	0.0078	0.0079	0.0078	0.0078
Family status-2 adults,1+ children		0.0432	0.0514	0.0263	0.026	0.0286	0.0218	0.0251	0.0249
		0.0577	0.0512	0.0484	0.0481	0.0476	0.0486	0.0485	0.0485
Family status-3 adults,no children		-0.024	-0.0022	0.009	0.0085	0.0083	0.0104	0.0107	0.0106
		0.0108*	0.0098	0.0095	0.0094	0.0094	0.0096	0.0095	0.0095
Family status-3 adults,1+ children		0.0024	0.0295	0.0087	0.0094	0.0119	0.0035	0.0075	0.0074
		0.0586	0.0519	0.0493	0.049	0.0485	0.0495	0.0493	0.0493
No kids- 1 child in the house		-0.0388	-0.0333	0.0025	0.0021	-0.0012	0.0058	0.0025	0.0026
		0.0579	0.0512	0.0485	0.0482	0.0477	0.0487	0.0486	0.0486
No kids- 2 children		-0.0876	-0.062	-0.0207	-0.0209	-0.0236	-0.0176	-0.0211	-0.021
		0.058	0.0512	0.0485	0.0482	0.0477	0.0487	0.0485	0.0485
No kids- 3 or more		-0.1235	-0.0806	-0.0444	-0.0446	-0.0472	-0.0372	-0.0415	-0.0413
		0.0586*	0.0518	0.0492	0.0489	0.0484	0.0494	0.0492	0.0493
Occ-Professional			0.0171	0.0175	0.0177	0.0179	0.0169	0.0164	0.0164
			0.0091	0.0091	0.0091	0.0091*	0.0091	0.0091	0.0091
Occ-Associate Professional & Technical			0.0021	0.0004	0.0014	0.0015	0.0018	0.0018	0.0017
			0.0087	0.0086	0.0086	0.0085	0.0085	0.0085	0.0085
Occ-Administrative & Secretarial			-0.0167	-0.0191	-0.0174	-0.0168	-0.0149	-0.0149	-0.015
			0.0088	0.0086*	0.0086*	0.0085*	0.0085	0.0085	0.0085
Occ-Skilled Trades			-0.0575	-0.059	-0.0571	-0.0557	-0.0553	-0.0549	-0.055
			0.0100***	0.0097***	0.0097***	0.0097***	0.0096***	0.0096***	0.0096***
Occ-Personal Service			-0.0888	-0.088	-0.0861	-0.0852	-0.0843	-0.0842	-0.0842
			0.0112***	0.0109***	0.0109***	0.0109***	0.0109***	0.0109***	0.0109***
Occ-Sales & Customer Service			-0.0226	-0.0266	-0.0255	-0.0245	-0.0253	-0.025	-0.025
			0.0115	0.0113*	0.0113*	0.0113*	0.0113*	0.0113*	0.0113*
Occ-Process, Plant & Machine Operatives			-0.1695	-0.1689	-0.1665	-0.1636	-0.1663	-0.1661	-0.1662
			0.0107***	0.0104***	0.0104***	0.0104***	0.0104***	0.0105***	0.0105***
Occ-Elementary			-0.1313	-0.1278	-0.1255	-0.1227	-0.1238	-0.1231	-0.1232
			0.0104***	0.0102***	0.0102***	0.0102***	0.0102***	0.0102***	0.0102***
Live & work in same Local Authority			-0.6773	-0.7225	-0.7211	-0.723	-0.747	-0.7474	-0.7473
			0.0092***	0.0083***	0.0082***	0.0083***	0.0087***	0.0087***	0.0087***
Full time job dummy			-0.0778	-0.0696	-0.0698	-0.0699	-0.070	-0.070	-0.070
			0.0067***	0.0066***	0.0066***	0.0066***	0.0066***	0.0066***	0.0066***
Second job dummy			-0.0545	-0.0493	-0.05	-0.0487	-0.0487	-0.0486	-0.0486
			0.0116***	0.0113***	0.0112***	0.0113***	0.0113***	0.0113***	0.0113***
Public sector worker			0.0312	0.0229	0.0241	0.0243	0.02	0.0197	0.0197
			0.0085***	0.0083**	0.0083**	0.0083**	0.0083*	0.0083*	0.0083*
Distance to closest transport improvement					0.0015	0.0016			
					0.0008	0.0008*			
Mean age of the population					-0.0034	-0.0031			
					0.0011**	0.0011**			
% of households living in social housing					0.0238	-0.0777			
					0.0485	0.0505			
% of WAP with qualification level 4 o 5					0.1271	0.1584			
					0.0570*	0.0592**			
Average distance traveled to workplace					0.01160	0.00750			
					0.0016***	0.0016***			
% of households with cars or vans					-0.27740	-0.41340			
					0.0695***	0.0740***			
% of employees commuting by motor					0.37850	0.57500			
					0.0614***	0.0662***			
Area					0.0008	0.0014			
					0.0003**	0.0002***			
Observations	110,148	110,148	110,148	110,148	110,148	110,148	110,148	110,148	110,148
Adjusted R^2	0.0272	0.0573	0.2698	0.2955	0.2973	0.2993	0.2716	0.2718	0.2718
Personal characteristics		YES	YES	YES	YES	YES	YES	YES	YES
Job characteristics			YES	YES	YES	YES	YES	YES	YES
District dummies				YES	YES	YES			
Ward attributes/trends					YES	YES		YES	YES
Scheme dummies/trends						YES		YES	YES
Ward fixed-effects							YES	YES	YES

Clustered (ward) s.e. All specifications include year, monthly dummies & pseudo-fe (age, gender, white). * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Source: LFS-ONS, DfT, CASWEB and authors own calculations.

Table B.21: LFS results: Log of travel time to workplace

Depvar: **Employment status (employed/out of work)**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of accessibility (1997 employment)	-0.0118	-0.0008	0.0026	0.0007	0.0001	0.0232	0.0191	
	0.0018***	0.0011	0.0012*	0.0012	0.0013	0.0342	0.037	
Log of accessibility (annual employment)								0.0059
								0.0118
Student		-0.3516	-0.3506	-0.3488	-0.3484	-0.3468	-0.3468	-0.3468
		0.0044***	0.0042***	0.0039***	0.0039***	0.0038***	0.0038***	0.0038***
Skill-other		0.1623	0.1593	0.1582	0.1581	0.1558	0.1558	0.1558
		0.0028***	0.0027***	0.0027***	0.0027***	0.0026***	0.0026***	0.0026***
Skill-O-level		0.1942	0.1892	0.1873	0.1871	0.1843	0.1842	0.1842
		0.0029***	0.0028***	0.0027***	0.0027***	0.0027***	0.0027***	0.0027***
Skill-teach & A-level		0.2141	0.2109	0.2098	0.2098	0.2081	0.208	0.208
		0.0034***	0.0033***	0.0032***	0.0032***	0.0032***	0.0032***	0.0032***
Skill-high		0.2444	0.2415	0.2416	0.2415	0.2414	0.2415	0.2415
		0.0034***	0.0033***	0.0032***	0.0032***	0.0032***	0.0032***	0.0032***
House-rented from a gov scheme		-0.2521	-0.2468	-0.2444	-0.2445	-0.2458	-0.2457	-0.2457
		0.0031***	0.0030***	0.0030***	0.0030***	0.0030***	0.0030***	0.0030***
House-privately rented		-0.1095	-0.1066	-0.0993	-0.0996	-0.0985	-0.0985	-0.0985
		0.0035***	0.0035***	0.0033***	0.0033***	0.0031***	0.0031***	0.0031***
Civil status-married		0.0062	0.0038	0.0029	0.0029	0.0037	0.0038	0.0038
		0.0022**	0.0022	0.0023	0.0022	0.0023	0.0023	0.0023
Civil status-separated		0.0136	0.0128	0.0124	0.0121	0.0118	0.0118	0.0118
		0.0044**	0.0044**	0.0044**	0.0044**	0.0044**	0.0044**	0.0044**
Civil status-divorced		0.024	0.0212	0.0207	0.0205	0.0206	0.0206	0.0206
		0.0032***	0.0032***	0.0032***	0.0032***	0.0032***	0.0032***	0.0032***
Civil status-widowed		-0.047	-0.0481	-0.0491	-0.0492	-0.0472	-0.0473	-0.0473
		0.0074***	0.0074***	0.0074***	0.0074***	0.0074***	0.0074***	0.0074***
Family status-1 adult,1+ children		-0.0566	-0.0551	-0.0504	-0.0506	-0.0454	-0.0445	-0.0445
		0.0132***	0.0132***	0.0132***	0.0132***	0.0135***	0.0134***	0.0134***
Family status-2 adults,no children		0.0674	0.0658	0.0644	0.0644	0.0644	0.0643	0.0643
		0.0025***	0.0025***	0.0025***	0.0025***	0.0025***	0.0025***	0.0025***
Family status-2 adults,1+ children		0.0505	0.051	0.0554	0.0551	0.0591	0.0597	0.0598
		0.0127***	0.0127***	0.0127***	0.0127***	0.0130***	0.0129***	0.0129***
Family status-3 adults,no children		0.0603	0.0593	0.0574	0.0577	0.0604	0.0603	0.0603
		0.0032***	0.0032***	0.0031***	0.0031***	0.0031***	0.0031***	0.0031***
Family status-3 adults,1+ children		0.0863	0.0871	0.091	0.0909	0.0957	0.0965	0.0965
		0.0130***	0.0130***	0.0130***	0.0130***	0.0132***	0.0132***	0.0132***
No kids-1 child in the house		-0.0326	-0.0349	-0.0416	-0.0412	-0.044	-0.0448	-0.0448
		0.0127*	0.0127**	0.0127**	0.0127**	0.0130***	0.0130***	0.0130***
No kids-2 children		-0.0594	-0.0628	-0.07	-0.0698	-0.072	-0.0728	-0.0728
		0.0128***	0.0128***	0.0128***	0.0128***	0.0130***	0.0130***	0.0130***
No kids-3 or more		-0.1395	-0.1417	-0.1476	-0.1475	-0.1475	-0.1484	-0.1484
		0.0130***	0.0130***	0.0130***	0.0130***	0.0132***	0.0132***	0.0132***
Distance to closest transport improvement				-0.0002	-0.0002			
				0.0002	0.0002			
Mean age of the population				0.0011	0.0014			
				0.0004**	0.0004***			
% of households living in social housing				0.1303	0.1147			
				0.0159***	0.0165***			
% of WAP with qualification level 4 o 5				-0.0432	-0.0364			
				0.0168*	0.0176*			
Average distance traveled to workplace				-0.0039	-0.0044			
				0.0005***	0.0005***			
% of households with cars or vans				0.3109	0.2778			
				0.0228***	0.0239***			
% of employees commuting by motor				-0.0413	-0.0122			
				0.0189*	0.0211			
Area				-0.0001	-0.0001			
				0.0001*	0.0001			
Observations	1,182,811	1,182,811	1,182,811	1,182,811	1,182,811	1,182,811	1,182,811	1,182,811
Adjusted R^2	0.087	0.234	0.237	0.238	0.239	0.205	0.206	0.206
Personal characteristics		YES	YES	YES	YES	YES	YES	YES
District dummies			YES	YES	YES			
Ward attributes/trends				YES	YES		YES	YES
Scheme dummies/trends					YES		YES	YES
Ward fixed-effects						YES	YES	YES

Clustered (ward) s.e. All specifications include year, monthly dummies & pseudo-fe (age, gender, white). * $p<0.05$, ** $p<0.01$, *** $p<0.00$. Source: LFS-ONS, DfT, CASWEB and authors own calculations.

Table B.22: LFS results: Employment status (employed/out of work)

Depvar: **Employment status (employed/unemployed)**								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log of accessibility (1997 employment)	-0.0046	-0.0002	0.0006	0.0005	0.0005	-0.0178	-0.003	
	0.0007***	0.0005	0.0005	0.0006	0.0006	0.0163	0.0181	
Log of accessibility (annual employment)								-0.01
								0.0067
Student		-0.0384	-0.0391	-0.0395	-0.0394	-0.0406	-0.0406	-0.0406
		0.0032***	0.0031***	0.0032***	0.0032***	0.0032***	0.0032***	0.0032***
Skill-other		0.0377	0.0366	0.0361	0.0361	0.036	0.036	0.036
		0.0018***	0.0018***	0.0018***	0.0018***	0.0018***	0.0018***	0.0018***
Skill-O-level		0.0446	0.043	0.0419	0.0419	0.0416	0.0416	0.0416
		0.0018***	0.0018***	0.0018***	0.0018***	0.0018***	0.0018***	0.0018***
Skill-teach & A-level		0.0593	0.0585	0.0571	0.057	0.057	0.057	0.0569
		0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***
Skill-high		0.0587	0.0572	0.0553	0.0554	0.0556	0.0557	0.0557
		0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***	0.0019***
House-rented from a gov scheme		-0.1203	-0.1191	-0.1174	-0.1174	-0.1185	-0.1184	-0.1184
		0.0022***	0.0022***	0.0022***	0.0022***	0.0023***	0.0023***	0.0023***
House-privately rented		-0.0326	-0.0332	-0.0318	-0.0319	-0.0328	-0.0328	-0.0328
		0.0017***	0.0018***	0.0018***	0.0018***	0.0018***	0.0018***	0.0018***
Civil status-married		0.0301	0.0292	0.0288	0.0288	0.0284	0.0284	0.0284
		0.0013***	0.0013***	0.0013***	0.0013***	0.0013***	0.0013***	0.0013***
Civil status-separated		0.0094	0.0093	0.0092	0.0092	0.0093	0.0092	0.0092
		0.0029**	0.0029**	0.0029**	0.0029**	0.0029**	0.0029**	0.0029**
Civil status-divorced		0.0151	0.0141	0.0141	0.014	0.0137	0.0137	0.0137
		0.0019***	0.0018***	0.0018***	0.0018***	0.0019***	0.0019***	0.0019***
Civil status-widowed		0.0184	0.018	0.0177	0.0177	0.0186	0.0186	0.0186
		0.0041***	0.0041***	0.0041***	0.0041***	0.0041***	0.0041***	0.0041***
Family status-1 adult,1+ children		0.0009	0.0015	0.003	0.0031	0.0065	0.0075	0.0075
		0.0063	0.0063	0.0064	0.0064	0.0065	0.0065	0.0065
Family status-2 adults,no children		0.0183	0.0177	0.0172	0.0172	0.0176	0.0175	0.0175
		0.0014***	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***	0.0014***
Family status-2 adults,1+ children		0.0312	0.0313	0.0325	0.0325	0.0357	0.0366	0.0365
		0.0056***	0.0057***	0.0057***	0.0057***	0.0058***	0.0058***	0.0058***
Family status-3 adults,no children		0.0163	0.016	0.0152	0.0153	0.0159	0.0158	0.0158
		0.0017***	0.0017***	0.0017***	0.0017***	0.0017***	0.0017***	0.0017***
Family status-3 adults,1+ children		0.0325	0.0327	0.0338	0.0338	0.0374	0.0383	0.0383
		0.0058***	0.0058***	0.0058***	0.0058***	0.0059***	0.0060***	0.0060***
No kids-1 child in the house		-0.024	-0.0244	-0.0264	-0.0264	-0.0292	-0.0301	-0.0301
		0.0056***	0.0056***	0.0057***	0.0057***	0.0058***	0.0058***	0.0058***
No kids-2 children		-0.0297	-0.0306	-0.0329	-0.0329	-0.0356	-0.0365	-0.0364
		0.0056***	0.0057***	0.0057***	0.0057***	0.0058***	0.0059***	0.0058***
No kids-3 or more		-0.0457	-0.0463	-0.0482	-0.0482	-0.0506	-0.0515	-0.0515
		0.0059***	0.0059***	0.0059***	0.0059***	0.0061***	0.0061***	0.0061***
Distance to closest transport improvement				0.000	0.000			
				0.0001	0.0001			
Mean age of the population				0.0004	0.0005			
				0.0002*	0.0002**			
% of households living in social housing				0.04	0.0353			
				0.0077***	0.0081***			
% of WAP with qualification level 4 o 5				-0.0009	-0.0013			
				0.0082	0.0089			
Average distance traveled to workplace				-0.001	-0.0012			
				0.0002***	0.0002***			
% of households with cars or vans				0.1167	0.1104			
				0.0111***	0.0120***			
% of employees commuting by motor				-0.0385	-0.0328			
				0.0087***	0.0101**			
Area				0.0000	0.0000			
				0.0000	0.0000			
Observations	920,644	920,644	920,644	920,644	920,644	920,644	920,644	920,644
Adjusted R^2	0.036	0.085	0.087	0.088	0.088	0.070	0.070	0.070
Personal characteristics		YES	YES	YES	YES	YES	YES	YES
District dummies			YES	YES	YES			
Ward attributes/trends				YES	YES		YES	YES
Scheme dummies/trends					YES		YES	YES
Ward fixed-effects						YES	YES	YES

Clustered (ward) s.e. All specifications include year, monthly dummies & pseudo-fe (age, gender, white). * p<0.05, ** p<0.01, *** p<0.001. Source: LFS-ONS, DfT, CASWEB and authors own calculations.

Table B.23: LFS results: Employment status (employed/unemployed)

List of Tables

List of Figures

Bibliography

Alonso-Borrego, C., Garoupa, N., & Vázquez, P. (2011). *Does immigration cause crime? Evidence from Spain*. Working Paper 11-08, Universidad Carlos III Economics Series.

Amior, M. (2011). *Why is Earnings Inequality Larger in Expensive Cities?* Working paper, University College London.

Amiti, M. & Cameron, L. (2007). Economic geography and wages. *The Review of Economics and Statistics*, 89(1), 15–29.

Amuedo-Dorantes, C. & de la Rica, S. (2008a). *Complements or Substitutes? Immigrant and Native Task Specialization in Spain*. CReAM Discussion Paper Series 0816, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Amuedo-Dorantes, C. & de la Rica, S. (2008b). *Does Immigration Raise Natives Income? National and Regional Evidence from Spain*. Working Paper 2008-17, FEDEA.

Andersson, F., Burgess, S., & Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics*, 61(1), 112–128.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, First edition.

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of Monetary Economics*, 23(2), 177–200.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal*, 7(4), 465–506.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2010). ivreg2: Stata module for extended instrumental variables/2sls,gmm and ac/hac, liml and k-class regression.

Baum-Snow, N. (2007a). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2), 775–805.

Baum-Snow, N. (2007b). Suburbanization and transportation in the monocentric model. *Journal of Urban Economics*, 62(3), 405–423.

Baum-Snow, N. (2010). Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960-2000. *American Economic Review*, 100(2), 378–82.

Behrens, K., Lamorgese, A., Ottaviano, G. I. P., & Tabuchi, T. (2004). *Testing the Home Market Effect in a Multi-Country World: The Theory*. CEPR Discussion Papers 4468, Centre for Economic and Policy Research.

Bentolila, S., Dolado, J. J., & Jimeno, J. F. (2008). *Does immigration affect the Phillips curve? Some evidence for Spain*. Banco de España Working Paper 0814, Banco de España.

Bertoli, S. & Fernández-Huertas, J. (2011). *Multilateral Resistance to Migration*. IZA Discussion Papers 5958, Institute for the Study of Labor (IZA).

Bertoli, S., Fernández-Huertas, J., & Ortega, F. (2011). Immigration policies and the ecuadorian exodus. *World Bank Economic Review*, 25(1), 57–76.

Boarnet, M. G. (1998). Spillovers and the locational effects of public infrastructure. *Journal of Regional Science*, 38(3), 381–400.

Boix, R. & Veneri, P. (2009). *Metropolitan Areas in Spain and Italy*. IERMB Working Paper, Institut d'Estudis Regionals i Metropolitans de Barcelona.

Borjas, G. J. (1995). The economic benefits from immigration. *Journal of Economic Perspectives*, 9(2), 3–22.

Borjas, G. J. (2002). Homeownership in the immigrant population. *Journal of Urban Economics*, 52(3), 448–476.

Borjas, G. J. (2003). The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. *The Quarterly Journal of Economics*, 118(4), 1335–1374.

Borjas, G. J. (2006). Native internal migration and the labor market impact of immigration. *Journal of Human Resources*, 41(2).

Borjas, G. J. (2009). *The Analytics of the Wage Effect of Immigration*. NBER Working Paper 14796, National Bureau of Economic Research, Inc.

Brakman, S., Garretsen, H., & van Marrewijk, C. (2009). *The New Introduction to Geographical Economics*. Cambridge Books. Cambridge University Press.

Buonanno, P., Bianchi, M., & Pinotti, P. (2011). Do immigrants cause crime? *Journal of European Economic Association*, Forthcoming.

Burchfield, M., Overman, H. G., Puga, D., & Turner, M. A. (2006). Causes of sprawl: A portrait from space. *The Quarterly Journal of Economics*, 121(2), 587–633.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1), 22–64.

Card, D. (2007). *How Immigration Affects U.S. Cities*. CReAM Discussion Paper Series 0711, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Card, D. & DiNardo, J. (2000). Do immigrant inflows lead to native outflows? *American Economic Review*, 90(2), 360–367.

Carrasco, R., Jimeno, J. F., & Ortega, A. C. (2008). *The impact of immigration on the wage structure: Spain 1995-2002*. Economics Working Paper we080603, Universidad Carlos III, Departamento de Econom/'ia.

Chandra, A. & Thompson, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, 30(4), 457–490.

Ciccone, A. & Hall, R. E. (1996). Productivity and the density of economic activity. *American Economic Review*, 86(1), 54–70.

Combes, P.-P. (2011). The empirics of economic geography: How to draw policy implications? *Review of World Economics*, 147(3), 567–592.

Combes, P.-P., Duranton, G., & Gobillon, L. (2008a). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2), 723–742.

Combes, P.-P., Duranton, G., & Gobillon, L. (2011). The identification of agglomeration economies. *Journal of Economic Geography*, 11(2), 253–266.

Combes, P.-P., Duranton, G., Gobillon, L., & Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration Economics*, NBER Chapters (pp. 15–65). National Bureau of Economic Research, Inc.

Combes, P.-P. & Lafourcade, M. (2001). *Transport Cost Decline and Regional Inequalities: Evidence from France*. CEPR Discussion Papers 2894, Centre for Economic and Policy Research.

Combes, P.-P., Mayer, T., & Thisse, J.-F. (2008b). *Economic geography: The integration of regions and nations*. Princeton University Press.

Cortés, P. (2008). The effect of low-skilled immigration on U.S. prices: evidence from CPI data. *Journal of Political Economy*, 116(3), 381–422.

Coughlin, C. C. & Segev, E. (2000). Location determinants of new foreign-owned manufacturing plants. *Journal of Regional Science*, 40(2), 323–351.

Criscuolo, C., Haskel, J., & Martin, R. (2003). Building the evidence base for productivity policy using business data linking. *Economic Trends*, 600, 39–51.

Cuñat, V. & Garicano, L. (2010). *Did Good Cajas Extend Bad Loans? Governance, Human Capital and Loan Portfolio*. Mimeo, London School of Economics.

Cushing, B. & Poot, J. (2003). Crossing boundaries and borders: regional science advances in migration modelling. *Economics of Governance*, 83(1), 317–338.

Dearden, L., Fitzsimons, E., & Wyness, G. (2011). *The Impact of Tuition Fees and Support on University*. IFS Working Papers W11/17, Institute for Fiscal Studies.

Decressin, J. & Fatas, A. (1995). Regional labor market dynamics in Europe. *European Economic Review*, 39(9), 1627–1655.

Department of Transport (1998a). *A new Deal for Transport: Better for Everyone*. White paper - Department of Transport.

Department of Transport (1998b). *A New Deal for Trunk Roads in England: Understanding the New Approach*. Report - Department of Transport.

Department of Transport (2005). *Transport, Wider Economic Benefits, and Impacts on GDP*. Discussion paper.

Détang-Dessendre, C. & Gaigné, C. (2009). Unemployment duration, city size, and the tightness of the labor market. *Regional Science and Urban Economics*, 39(3), 266–276.

Di Addario, S. (2011). Job search in thick markets. *Journal of Urban Economics*, 69(3), 303–318.

Díaz-Serrano, L. & Raya, J. (2011). *Is there Discriminatory Mortgage Pricing against Immigrants in the Spanish Lending Market?* IZA Discussion Paper 5578, IZA Institute for the Study of Labor.

Donaldson, D. (2010). Railroads of the raj: Estimating the impact of transportation infrastructure. *NBER Working Papers*, 16487.

Duranton, G. (2008). Spatial economics. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Palgram Macmillan, 2^{nd} edition.

Duranton, G., Morrow, P., & Turner, M. A. (2011). Roads and trade: Evidence from the US. Mimeo, University of Toronto.

Duranton, G. & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In J. V. Henderson & J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, volume 4 of *Handbook of Regional and Urban Economics* chapter 48, (pp. 2063–2117). Elsevier.

Duranton, G. & Rodríguez-Pose, A. (2005). When economists and geographers collide, or the tale of the lions and the butterflies. *Environment and Planning A*, 37(10), 1695–1705.

Duranton, G. & Storper, M. (2006). Agglomeration and growth: a dialogue between economists and geographers. *Journal of Economic Geography*, 6(1), 1–7.

Duranton, G. & Turner, M. A. (2011a). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, 101(6), 2616–52.

Duranton, G. & Turner, M. A. (2011b). Urban growth and transportation. Mimeo, University of Toronto.

Dustmann, C., Frattini, T., & Preston, I. (2008a). *The Effect of Immigration along the Distribution of Wages*. CReAM Discussion Paper Series 0803, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Dustmann, C., Glitz, A., & Frattini, T. (2008b). The labour market impact of immigration. *Oxford Review of Economic Policy*, 24(3), 478–495.

El-Geneidy, A. M. & Levinson, D. M. (2006). *Access to destinations: Development of accessibility measures*. Report, Research Services Section. Mn/DOT 2006-16, University of Minnesota. Dept. of Civil Engineering and Minnesota. Dept. of Transportation.

Faber, B. (2009). *Market Integration and the Periphery: The Unintended Consequences of New Highways in a Developing Country*. Unpublished, London School of Economics.

Fernald, J. G. (1999). Roads to prosperity? Assessing the link between public capital and productivity. *American Economic Review*, 89(3), 619–638.

Fernández-Huertas, J., Ferrer, A., & Saiz, A. (2009). *Localización de los Inmigrantes y Preferencias Residenciales de la Poblacin Autóctona: ¿Nuevos guetos?* Working paper, Institute for Economic Analysis (CSIC), Universitat Autónoma de Barcelona, Spain.

Fingleton, B. (2006). The new economic geography versus urban economics: an evaluation using local wage rates in great britain. *Oxford Economic Papers*, 58(3), 501–530.

Frattini, T. (2008). *Immigration and Prices in the UK*. Job market paper, Univesity College of London.

Frisch, R. & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica: Journal of the Econometric Society*, (pp. 387–401).

Fujita, M., Krugman, P., & Venables, A. J. (2001). *The Spatial Economy: Cities, Regions, and International Trade*, volume 1 of *MIT Press Books*. The MIT Press.

Gandal, N., Hanson, G. H., & Slaughter, M. J. (2004). Technology, trade, and adjustment to immigration in Israel. *European Economic Review*, 48(2), 403–428.

García-Milá, T., McGuire, T. J., & Porter, R. H. (1996). The effect of public capital in state-level production functions reconsidered. *The Review of Economics and Statistics*, 78(1), 177–80.

García-Montalvo, J. (2010). *Land Regulations and Housing Prices: an Investigation for the Spanish Case*. Working paper, Fundación Banco Santander.

Gauthier-Loiselle, M. & Hunt, J. (2009). *How Much Does Immigration Boost Innovation?* CEPR Discussion Papers 7116, Centre for Economic and Policy Research.

Gautier, P. A. & Zenou, Y. (2010). Car ownership and the labor market of ethnic minorities. *Journal of Urban Economics*, 67(3), 392–403.

Gibbons, S., Lyytikäinen, T., Overman, H., & Sanchis-Guarner, R. (2011). *Productivity and employment impacts of transport improvements*. Working paper, London School of Economics.

Gibbons, S. & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1), 148–169.

Gibbons, S. & Machin, S. (2006). *Transport and Labour Market Linkages: empirical evidence, implications for policy and scope for further research*. Background paper for the Eddington report to the Department of Transport.

Gibbons, S. & Overman, H. G. (2009). *Productivity in Transport Evaluation Studies*. Report for the Department of Transport.

Glaeser, E. L. (2008). *Cities, Agglomeration, and Spatial Equilibrium*. Oxford University Press.

Glaeser, E. L. & Kohlhase, J. (2003). Cities, regions and the decline of transport costs. *Papers in Regional Science*, 83(1), 197–228.

Gobillon, L., Magnac, T., & Selod, H. (2011). The effect of location on finding a job in the paris region. *Journal of Applied Econometrics*, 26(7), 1079–1112.

Gobillon, L., Selod, H., & Zenou, Y. (2007). The mechanisms of spatial mismatch. *Urban Studies*, 44(12), 2401–2427.

Goerlich-Gisbert, F. J. & Cantarino-Martí, I. (2010). *Un ndice de rugosidad del terreno a escala municipal a partir de modelos de elevacin digital de acceso pblico*. Working Papers 20105, Fundacion BBVA / BBVA Foundation.

Gonzalez, L. & Ortega, F. (2009). *Immigration and Housing Booms: evidence from Spain*. IZA Discussion Papers 4333, Institute for the Study of Labor (IZA).

Gonzalez, L. & Ortega, F. (2010). How do very open economies absorb large immigration flows? Recent evidence from Spanish regions. *Labour Economics*, (pp. 149–187).

Graham, D. J. (2005). *Wider Economic Benefits of Transport Improvements: link between Agglomeration and Productivity*. Stage 1 report to the Department of Transport.

Graham, D. J. (2006). *Wider Economic Benefits of Transport Improvements: link between Agglomeration and Productivity*. Stage 2 report to the Department of Transport.

Graham, D. J. (2007). *Agglomeration Economies and Transport Investment*. OECD/ITF Joint Transport Research Centre Discussion Papers 2007/11, OECD Publishing.

Graham, D. J., Gibbons, S., & Martin, R. (2009). *Transport investments and the distance decay of agglomeration benefits*. Draft – Department of Transport.

Graham, D. J. & Melo, P. C. (2009). *Agglomeration Economies and Labour Productivity: Evidence from Longitudinal Worker Data for GBs Travel-to-Work Areas*. SERC Discussion Papers 0031, Spatial Economics Research Centre, LSE.

Gramlich, E. M. (1994). Infrastructure investment: A review essay. *Journal of Economic Literature*, 32(3), 1176–96.

Greulich, E., Quigley, J. M., & Raphael, S. (2005). The anatomy of rent burdens: Immigration, growth and rental housing. (30864).

Gutiérrez-i-Puigarnau, E. & van Ommeren, J. N. (2010). Labour supply and commuting. *Journal of Urban Economics*, 68(1), 82–89.

H. Hanson, G. (2005). Market potential, increasing returns and geographic concentration. *Journal of International Economics*, 67(1), 1–24.

Hanson, G. H. (2008). *The Economic Consequences of the International Migration of Labor*. NBER Working Paper 14490, National Bureau of Economic Research, Inc.

Harris, C. D. (1954). The, market as a factor in the localization of industry in the United States. *Annals of the association of American geographers*, 44(4), 315–348.

Hatton, T. J. & Tani, M. (2005). Immigration and inter-regional mobility in the UK, 1982-2000. *Economic Journal*, 115(507), F342–F358.

Head, K. & Mayer, T. (2006). Regional wage and employment responses to market potential in the eu. *Regional Science and Urban Economics*, 36(5), 573–594.

Henderson, J. V. (2003). Marshall's scale economies. *Journal of Urban Economics*, 53(1), 1–28.

Hering, L. & Poncet, S. (2010). Market access and individual wages: Evidence from China. *The Review of Economics and Statistics*, 92(1), 145–159.

Highways Agency (2009). *The Highways Agency: Framework Document*. Highways Agency - Department of Transport.

Hilber, C. A. L. & Vermeulen, W. (2009). Supply constraints and house price dynamics: Panel data evidence from england.

Holl, A. (2004a). Manufacturing location and impacts of road transport infrastructure: Empirical evidence from Spain. *Regional Science and Urban Economics*, 24(3), 341–363.

Holl, A. (2004b). Start-ups and relocations: Manufacturing plant location in portugal*. *Papers in regional Science*, 83(4), 649–668.

Holl, A. (2004c). Transport infrastructure, agglomeration economies, and firm birth: Empirical evidence from portugal. *Journal of Regional Science*, 44(4), 693–712.

Holl, A. (2006). A review of the firm-level role of transport infrastructure with implications for transport project evaluation. *Journal of Planning Literature*, 21(1), 3.

Holl, A. (2011). Market potential and firm-level productivity in Spain. *Journal of Economic Geography*, Forthcoming.

Holtz-Eakin, D. (1994). Public-sector capital and the productivity puzzle. *The Review of Economics and Statistics*, 76(1), 12–21.

Hymel, K. (2009). Does traffic congestion reduce employment growth? *Journal of Urban Economics*, 65(2), 127–135.

Ihlanfeldt, K. (2002). Spatial mismatch in the labor market and racial differences in neighborhood crime. *Economics Letters*, 76(1), 73–76.

Ihlanfeldt, K. R. & Sjoquist, D. L. (1990). Job accessibility and racial differences in youth employment rates. *American Economic Review*, 80(1), 267–76.

Inhanfeldt, K. R. (2006). A primer on spatial mismatch within urban labor markets. In *A companion to urban economics*, Blackwell companions to contemporary economics (pp. 404 – 417). Wiley-Blackwell.

Jiwattanakulpaisarn, P., Noland, R. B., Graham, D. J., & Polak, J. W. (2009). Highway infrastructure and state-level employment: A causal spatial analysis. *Papers in Regional Science*, 88(1), 133–159.

Jofré-Monseny, J., Marín-López, R., & Viladecans-Marsal, E. (2011). The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms. *Journal of Urban Economics*, 70(2-3), 61–74.

Kangasniemi, M., Mas, M., Robinson, C., & Serrano, L. (2009). *The Economic Impact of Migration: productivity Analysis for Spain and the United Kingdom*. MPRA Paper 17212, University Library of Munich, Germany.

Kleibergen, F. & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1), 97–126.

Krugman, P. & Venables, A. J. (1995). Globalization and the inequality of nations. *The Quarterly Journal of Economics*, (pp. 857–880).

Krugman, P. R. (1991a). *Geography and trade*. the MIT Press.

Krugman, P. R. (1991b). Increasing returns and economic geography. *Journal of Political Economy*, 99(3), 483–99.

Lach, S. (2007). Immigration and prices. *Journal of Political Economy*, 115(4), 548–587.

Lahr, M. L., Duran, R., & Varughese, A. (2005). *Estimating the Impact of Highways on Average Travel Velocities and Market Size*. EconWPA Urban/Regional 04003009, Center for Urban Policy Research.

Levinsohn, J. & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2), 317–341.

Li, Z. & Li, H. (2010). Road investment and inventory reduction: evidence from a large developing country. In *International Conference on Infrastructure Economics and Development organized by the World Bank, Agence Française de Développement, the Public Private Infrastructure Advisory Facility and the Institut D'économie Industrielle* (pp. 14–15).

Longhi, S., Nijkamp, P., & Poot, J. (2009). *Regional Economic Impacts of Immigration: a Review*. Tinbergen Institute Discussion Papers 09-047/3, Tinbergen Institute.

Manning, A. (2003). The real thin theory: monopsony in modern labour markets. *Labour Economics*, 10(2), 105–131.

Manning, A. (2006). A generalised model of monopsony. *Economic Journal*, 116(508), 84–100.

Marshall, A. (1890). *Principles of Economics*. London: Macmillan.

Martin, P., Mayer, T., & Mayneris, F. (2011). Spatial concentration and plant-level productivity in france. *Journal of Urban Economics*, 69(2), 182–195.

Martin, R. (1999). The new 'geographical turn' in economics: Some critical reflections. *Cambridge Journal of Economics*, 23(1), 65–91.

Martori, J., Hoberg, K., & Surinach, J. (2006). Población inmigrante y espacio urbano: indicadores de segregación y pautas de localización. *EURE-Revista Latinoamericana de Estudios Urbanos Regionales*, 32(97), 49–62.

Matas, A., Raymond, J.-L., & Roig, J.-L. (2010). Job accessibility and female employment probability: The cases of barcelona and madrid. *Urban Studies*, 47(4), 769–787.

Mazzolari, F. & Neumark, D. (2011). Immigration and product diversity. *Journal of Population Economics*, In press.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6), 1695–1725.

Melo, P. C., Graham, D. J., & Noland, R. B. (2010). Impact of transport infrastructure on firm formation: Evidence from the portuguese municipalities. *Transportation Research Record*, 2163, 133–143.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4), 683–701.

Mion, G. & Naticchioni, P. (2005). *Urbanization Externalities, Market Potential and Spatial Sorting of Skills and Firms*. CEPR Discussion Papers 5172, Centre for Economic and Policy Research.

Mion, G. & Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics*, 42(1), 28–55.

Mocetti, S. & Porello, C. (2010). How does immigration affect native internal mobility? new evidence from Italy. *Regional Science and Urban Economics*, 40(6), 427–439.

Moreno, R. & López-Bazo, E. (2007). Returns to local and transport infrastructure under regional spillovers. *International Regional Science Review*, 30(1), 47.

Mulalic, I., van Ommeren, J. N., & Pilegaard, N. (2010). Wages and commuting: Quasi-natural experiments' evidence from firms that relocate. *Tinbergen Institute Discussion Papers*, (10-093/3).

Nickell, S., Jones, P., & Quintini, G. (2002). A picture of job insecurity facing british men. *Economic Journal*, 112(476), 1–27.

Olley, G. S. & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6), 1263–97.

Ortega, F. & Peri, G. (2009). *The Causes and Effects of International Migrations: evidence from OECD Countries 1980-2005*. NBER Working Paper 14833, National Bureau of Economic Research, Inc.

Ortega, J. & Verdugo, G. (2011). *Immigration and the Occupational Choice of Natives: A Factor Proportions Approach*. IZA Discussion Papers 5451, Institute for the Study of Labor (IZA).

Ottaviano, G. I. (2008). *Infrastructure and economic geography: An overview of Theory and Evidence*. EIB Papers 6/2008, European Investment Bank - Economic and Financial Studies.

Ottaviano, G. I. & Peri, G. (2006). The economic value of cultural diversity: evidence from U.S. cities. *Journal of Economic Geography*, 6(1), 9–44.

Ottaviano, G. I. & Peri, G. (2008a). *Immigration and National Wages: clarifying the Theory and the Empirics*. NBER Working Paper 14188, National Bureau of Economic Research, Inc.

Ottaviano, G. I. & Peri, G. (2008b). *Task Specialization, Immigration, and Wages*. Discussion paper series CDP No 02/08, CREAM.

Ottaviano, G. I. & Peri, G. (2011). The effects of immigration on U.S. wages and rents: a general equilibrium approach. *Journal of the European Economic Association*, Forthcoming.

Overman, H. G. (2004). Can we learn anything from economic geography proper? *Journal of Economic Geography*, 4(5), 501–516.

Overman, H. G. & Puga, D. (2010). Labor pooling as a source of agglomeration: An empirical investigation. In *Agglomeration Economics*, NBER Chapters (pp. 133–150). National Bureau of Economic Research, Inc.

Ozgen, C., Nijkamp, P., & Poot, J. (2009). *The Effect of Migration on Income Convergence: meta-Analytic Evidence*. Tinbergen Institute Discussion Papers 09-022/3, Tinbergen Institute.

Patacchini, E. & Zenou, Y. (2005). Spatial mismatch, transport mode and search decisions in england. *Journal of Urban Economics*, 58(1), 62–90.

Pekkala-Kerr, S. & Kerr, W. R. (2011). Economic impacts of immigration: A survey. *Finnish Economic Papers*, 24(1), 1–32.

Peri, G. (2009). *The Effect of Immigration on Productivity: Evidence from U.S.* NBER Working Paper 15507, National Bureau of Economic Research, Inc.

Peri, G. & Requena, F. (2010). The trade creation effect of immigrants: Testing the theory on the remarkable case of Spain. *Canadian Journal of Economics*, 43, 1433–1459.

Peri, G. & Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, 1(3), 135–69.

Peri, G. & Sparber, C. (2011). Assessing inherent model bias: An application to native displacement in response to immigration. *Journal of Urban Economics*, 69, 82–91.

Puga, D. (2002). European regional policies in light of recent location theories. *Journal of Economic Geography*, 2(4), 373–406.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50(1), 203–219.

Quigley, J. M. (2008). Urban economics. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Palgram Macmillan, 2^{nd} edition.

Raphael, S. (1998). The spatial mismatch hypothesis and black youth joblessness: Evidence from the san francisco bay area. *Journal of Urban Economics*, 43(1), 79–111.

Redding, S. J. (2010). The empirics of new economic geography. *Journal of Regional Science*, 50(1), 297–311.

Redding, S. J. & Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, 62(1), 53–82.

Requena, F., Serrano, G., & Montaner, J. M. (2009). *The effects of immigration on the productive structure of Spanish regions*. Working Papers - Serie EC, Instituto Valenciano de Investigaciones Econmicas, S.A. (Ivie).

Rodríguez, C. & Bustillo, R. (2008). Modelling foreign real estate investment: The Spanish case. *The Journal of Real Estate Finance and Economics*, (pp. 17–27).

Rogers, C. L. (1997). Job search and unemployment duration: Implications for the spatial mismatch hypothesis. *Journal of Urban Economics*, 42(1), 109–132.

Rosenthal, S. S. & Strange, W. C. (2001). The determinants of agglomeration. *Journal of Urban Economics*, 50(2), 191–229.

Rosenthal, S. S. & Strange, W. C. (2008). Agglomeration and hours worked. *The Review of Economics and Statistics*, 90(1), 105–118.

Royuela, V. & Vargas, M. (2007). *Defining housing market areas using commuting and migration algorithms. Catalonia (Spain) as an applied case study*. IREA Working Paper 200707, University of Barcelona, Research Institute of Applied Economics.

Saiz, A. (2003). Room in the kitchen for the melting pot: immigration and rental prices. *The Review of Economics and Statistics*, 85(3), 502–521.

Saiz, A. (2007). Immigration and housing rents in American cities. *Journal of Urban Economics*, 61(2), 345–371.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics*, 125(3), 1253–1296.

Sosvilla-Rivero, S. (2008). *Immigration and Housing Prices in Spain*. Working Paper 2008-40, FEDEA.

Stillman, S. & Maré, D. C. (2008). *Housing Markets and Migration: Evidence from New Zealand*. Working Paper 08-06, Motu Economic and Public Policy Research.

Stock, J. H. & Yogo, M. (2002). *Testing for Weak Instruments in Linear IV Regression*. NBER Technical Working Papers 0284, National Bureau of Economic Research, Inc.

Strange, W. C. (2008). Urban agglomeration. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Palgram Macmillan, 2^{nd} edition.

Straub, S. (2011). Infrastructure and development: A critical appraisal of the macro-level literature. *The Journal of Development Studies*, 47(5), 683–708.

Talavull de la Paz, P. (2003). Determinants of housing prices in spanish cities. *Journal of Property Investment and Finance*, 21(2), 109–135.

Turner, M. A. (2009). The effects of transportation infrastructure on cities: a Review of the evidence. Lectures for EAERE Summer School, Venice 2009.

Uriel-Jiménez, E., Albert-Pérez, C., Benages-Candau, E., & Cucarella-Tormo, V. (2009). *El Stock de Capital en Viviendas en España y Su Distribución Territorial (1990-2007)*. Fundación BBVA, First edition.

van Ommeren, J. N. & Gutiérrez-i-Puigarnau, E. (2011). Are workers with a long commute less productive? An empirical analysis of absenteeism. *Regional Science and Urban Economics*, 41(1), 1–8.

Venables, A. J. (2007). Evaluating urban transport improvements: Cost-benefit analysis in the presence of agglomeration and income taxation. *Journal of Transport Economics and Policy*, 41(2), 173–188.

Venables, A. J. (2008). New economic geography. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Palgram Macmillan, 2^{nd} edition.

Venables, A. J. & Gasiorek, M. (1999). *The Welfare Implications of Transport Improvements in the Presence of Market Failures*. Report to The Standing Advisory Committee on Trunk Road Assessment (SACTRA).

Vickerman, R. (2002). *Transport and Economic Development*. Transport and economic development round table 119, European Conference of Ministers of Transport, OECD.

Vickerman, R. (2007). *Recent Evolution of Research into the Wider Economic Benefits of Transport Infrastructure Investments*. OECD/ITF Joint Transport Research Centre Discussion Papers 2007/9, OECD Publishing.

Vickerman, R., Spiekermann, K., & Wegener, M. (1999). Accessibility and economic development in europe. *Regional Studies*, 33(1), 1–15.

Warunsiri, S. & McNown, R. (2010). The returns to education in thailand: A pseudo-panel approach. *World Development*, 38(11), 1616–1625.

Zachariadis, M. (2010). *Immigration and International Prices*. Working Papers in Economics 3-2010, University of Cyprus Department of Economics.

Zachariadis, M. (2011). Immigration and international prices: An initial assessment. *Economics Letters*, 110(2), 83–85.

Zenou, Y. (2009). *Urban Labor Economics*. Cambridge University Press.