UNIVERSITY OF TAMPERE
School of Management

The Geographic Determinants of Housing Supply Revisited: Evidence from Finland

Economics
Master's Thesis
January 2018
Supervisor: Jani-Petri Laamanen

Sami Jysmä

ABSTRACT

University of Tampere
School of Management
JYSMÄ, SAMI: The Geographic Determinants of Housing Supply Revisited: Evidence from Finland

Master's Thesis, 46 pages + 2 annexes (6 pages)
Economics
January 2017
Keywords: housing market, geography, housing supply elasticity

---

The purpose of this thesis is to 1) create a city-level dataset of geographical data similar to that provided by Saiz (2010) within a Finnish context, 2) reproduce the analyses regarding the effects of geography on local housing supply elasticity conducted in the Saiz (2010) study with the newly generated data, and 3) expand both the theoretical and empirical model applied in these analyses to incorporate more geographical heterogeneity.

The thesis studies the local housing supply elasticity in 42 Finnish cities within the time period of 2003 to 2015. Using a cross-sectional instrumental variable approach, with similar findings in a fixed effects panel regression, this thesis finds that geographical variables have significant effects on local housing supply elasticity. For example, a one standard deviation increase in the share of water within a 30 km radius from a city centre lowers the housing supply elasticity by 0.075–0.107 units for an average city, depending on the exact model specification. This can be contrasted with the estimated housing supply elasticity for such a city, which ranges from 0.467 to 0.728, depending on the specification.

TIIVISTELMÄ

Tampereen yliopisto
Johtamiskorkeakoulu
JYSMÄ, SAMI: The Geographic Determinants of Housing Supply Revisited: Evidence from Finland

Pro gradu -tutkielma, 46 sivua + 2 liitettä (6 sivua)
Taloustiede,
tammikuu 2017

Tämän pro gradu -tutkielman tarkoituksena on 1) tuottaa samankaltainen kaupunkitason maantieteellinen tietoaineisto Suomen kontekstissa kuin Saiz'n (2010) tutkimuksessa, 2) toistaa kyseisen tutkimuksen empiiriset analyysit maantieteellisten muuttujien vaikutuksesta paikalliseen asuntotarjonnan hintajoustoon ja 3) laajentaa tutkimuksessa hyödynnettävää teoreettista ja empiiristä mallia siten, että se ottaa huomioon suuremman määrän maantieteellistä heterogeenisuutta.

Tässä tutkielmassa tarkastellaan 42 suomalaisen kaupungin paikallista asuntotarjonnan hintajoustoa aikavälillä 2003 – 2015. Käyttäen instrumenttimuuttujamenetelmää poikkileikkausaineistolle ja kiinteiden vaikutusten regressiota paneeliaineistolle tutkielmassa havaitaan maantieteellisillä muuttujilla olevan merkittävää vaikutusta paikalliseen asuntotarjonnan joustoon. Esimerkiksi yhden keskihajonnan kokoisen lisäyksen veden osuudessa 30 km säteellä kaupungin keskustasta estimoidaan alentavan asuntotarjonnan joustoa 0,075 – 0,107 yksikköä riippuen käytetystä mallista. Tätä voidaan verrata estimoituun kaupunkien keskimääräiseen asuntotarjonnan joustoon, joka on 0,467 –0,728 riippuen mallista.

TABLE OF CONTENTS

1. INTRODUCTION

In recent years, especially after the late 00's housing market collapse in the US and the subsequent sub-prime mortgage crisis, there has been an increasing interest in studying the dynamics of housing markets. One sub-trend in this line of research has focused on the determinants of local housing supply elasticity. Many plausible underlying causes behind city-level variation in the elasticity have been proposed in the literature (see e.g. Green et al. 2005). These include the cost of capital, both the growth rate and level of a city's population, it's density, and the level of transportation costs within the city. Moreover, since at least the research by Rose (1989a, 1989b), the role played by the topography of a city within the local housing market has been an object of interest.

One innovative contribution to this field was made by Albert Saiz (2010) in his article The Geographic Determinants of Housing Supply published in The Quarterly Journal of Economics. In the article, Saiz introduces a newly constructed dataset of undevelopable land within a 50 km radius from US metropolitan central cities. This was made possible by increased availability of high-quality, satellite measured geographic information system (GIS) data and software. After introducing the new dataset, Saiz estimates the effects of these geographical constraints on inverse housing supply elasticities for the 95 US metropolitan statistical areas that had a population over 500,000 in 2000. The regression was estimated with 30 year changes in housing market variables in a 2SLS cross-sectional setting. From the estimation results, he then calculates long-run metropolitan-level housing supply elasticities. (Saiz, 2010)

Measurements and analysis similar to that of Saiz have not yet been implemented with Finnish data. Thus, the first and second objectives of this master's thesis are, respectively, to produce similar geographical measurements and replicate some of Saiz's research in a Finnish context. Due to data limitations, however, I conduct a similar 2SLS cross-section regression, but with 12 year relative changes in housing market variables and a set of 42 cities. Even with these differences, the replication results are fairly similar to those of Saiz (2010), thus providing additional evidence on the effects of geographical features on housing supply.

Furthermore, the model Saiz uses to estimate housing supply elasticities does not take into account heterogeneities such as differences in the sources of land unavailability or differences in the distribution of these sources near each city. Thus, the third objective of this thesis is to construct and estimate a more general model that allows these heterogeneities to have their own effects on local housing supply elasticities. The results for these estimations provide evidence that these heterogeneities do, in fact, matter: for example, steep slopes do not seem to have a significant effect on housing supply.

The thesis itself is organised as follows: first, I will review some previous empirical research that has been conducted on housing supply elasticities in general and on the effects of geography on housing supply elasticities in particular. Second, I present the theoretical framework that Saiz (2010) applies to study the relationships between geographic constraints and local housing supply. Moreover, I will introduce new developments to the framework that allow me to analyse the impacts of the prevalence and relative position of different types of geographic features.

After the framework has been established, I will describe the newly produced data set on geographic constraints for Finnish urban areas and its construction, as well as the data used in this thesis to measure housing prices and housing units. Then I will move on to introduce and discuss both the estimation method and the estimation results in more detail both compared to Saiz and in the context of the extended framework. For some of the extended framework estimates, I also provide marginal effects of the geographical variables on housing supply elasticities, as these are nonlinear combinations of both regression coefficients and variables.

Moreover, based on the regression estimates, I conduct further calculations in order to arrive at local housing supply elasticity estimates and measures of average effects by one standard deviation increases in the geographical variables. These calculations suggest that the effects of geographical variables on housing supply elasticities have also economic significance. For example, a one standard deviation increase in the share of water within a 30 km radius from a city centre lowers the housing supply elasticity by 0.075–0.107 units for an average city, depending on the exact model specification. This can be contrasted with the estimated housing supply elasticity for such a city, which ranges from 0.467 to 0.728, depending on the specification.

2. PREVIOUS RESEARCH

In recent years, there has been an increased interest in the empirical estimation of the elasticity of housing supply. This is most likely contributable both to the often-expressed view in the literature that housing supply remains an understudied issue (see e.g. Malpezzi & Maclennan 2001, Green et al 2005, Ball et al. 2010) and to the increase in the availability of housing market data (Gyourko 2009).

Within this literature, there are several aspects in which individual studies differ from one another. For example, early studies, such as Muth (1960) and Follain (1979), tended to apply a reduced form approach, where a system of equations is solved for an endogenous variable (e.g. equating supply with demand) and then estimated. More recent studies have mostly relied on a structural approach.

Moreover, different studies have applied different spatial scales: some have approached the issue from an international or national perspective, whereas others have estimated housing supply elasticities at the regional or local levels. As pointed out by Ball, Meen and Nygaard (2010), estimation results should be scale-dependent as for example planning constraints differ between regions which then are aggregated to the national level.

Furthermore, similar issues could plausibly arise from different temporal settings as well: the effect of at least some supply constraints can be expected to be time-dependent and therefore different time scales can be expected to lead to differing results. In this respect, Saiz (2010) differs significantly from other recent studies, as only a cross-sectional regression with the relative change of housing prices between 1970 and 2000 as the dependent variable is used, whereas most recent studies are conducted in a panel-data setting.

In addition to differences in estimation methods and the applied spatial and temporal scales, studies differ in how they examine the effects of possible sources of heterogeneity in housing supply elasticities. Some studies (e.g. Green, Malpezzi and Mayo 2005 and Oikarinen et al. 2015) apply a two-stage strategy. They first estimate regional or city level housing supply elasticities and then carry out a cross-sectional regression with the estimated elasticity as the dependent variable. In this setting, possible sources of variation in the housing supply elasticity are added as explanatory variables in the cross-sectional regression. Other studies implement a one-step approach, where the variables of interest are already incorporated to the empirical model used to estimate the elasticity itself (e.g. Saiz 2010, Meen and Nygaard 2011). This is achieved either by introducing the variables of interest to the model on their own or as interactions with either a housing price variable or a housing quantity variable, depending on the specification used.

Of the plausible variables affecting housing supply elasticity, the effect of geography, even though its importance is widely acknowledged (Meen and Nygaard 2011), has been studied in a limited number of papers. Earlier studies (e.g. Rose 1989a, 1989b; Malpezzi 1996 and Malpezzi et al. 1998) tended to focus on proxies for the proportion of a city's surroundings lost to geographic constraints and their effect on housing prices instead of supply elasticities. Starting with Saiz (2010) satellite generated data has become the norm with studies such as Meen and Nygaard (2011), Wang et al (2012), Oikarinen et al. (2015) and Dong (2016) applying similarly constructed variables. These studies use either the land area or the land share lost to geographical features to measure geographical constraints on housing markets. All of these newer studies find that geographical features have a significant impact on the elasticity of housing supply.

As this thesis aims to replicate and further develop the research conducted by Saiz (2010), it is important to note that the estimation setting applied by Saiz differs in several aspects from the majority of related research. First, Saiz (2010) estimates a cross-sectional regression, using 30 year relative increases in housing prices and housing units. Most recent research has, however, been conducted in a structural panel regression setting. Second, Saiz (2010) introduces interactions of the explanatory housing market variable with geographical and regulatory variables into the regression equation in order to decipher the effects of these two variables on housing supply elasticities. In addition to Saiz (2010), these features can be found together, to my knowledge, only in Meen and Nygaard (2011). Furthermore, Saiz (2010), instead of estimating the elasticity of housing supply directly, estimates the inverse of the elasticity by having relative price change as the endogenous variable and relative change in housing supply as the main explanatory variable. Thus, Saiz (2010) only attains estimates for the housing supply elasticities through indirect calculations. Taken together, these aspects make the estimation setting – again, to my knowledge – unique in the literature. Therefore, comparisons of the results from Saiz (2010) and related research should be made with caution.

3. FRAMEWORK

In order to study the linkages between geographic constraints and housing prices, Saiz (2010) develops a conceptual framework based on the Alonso-Muth-Mills model (Alonso 1964; Mills 1967; Muth 1969). The model describes a city's housing market, while taking into account the land availability of the city's surrounding area. Here, I will first introduce Saiz's model following his 2010 article, and then develop it further so that it will be possible to take into account different sources of land unavailability and variation in the distribution of these constraints within the city radius.

3.1. The Saiz model

Saiz (2010) assumes perfect competition both between urban developers and between housing consumers. That is to say that developers act as price takers and prices are determined by consumers competing for housing locations. Consumers are assumed to be homogeneous, and consumption is a sum of city specific amenities A_k and wages w_k with rent r and commuting costs td deducted. Commuting costs are a product of monetary costs t of commute per distance unit and commuting distance d. (Saiz 2010, 1262)

A no arbitrage assumption is made, and as all consumers are homogeneous, this means that each consumer within a city receives the same utility level. Hence, with a utility function that only depends on consumption, for city k we have $U(C_k) = U(A_k + w_k - \gamma r' - td) = \overline{U_k}$, where r' is the rent per housing land consumption and γ is the amount of housing land consumed, which is assumed to be constant, so that $r = \gamma r'$. As amenities and wages are assumed to be city specific constants, this implies that rents can be expressed as a linear function of commuting distance. Hence,

$$r(d) = r_0 - td, \tag{1}$$

where r_0 is the rent level in the city centre. (Saiz 2010, 1262)

Saiz models the city's surrounding area as a circle with a radius Φ_k. The share of land available for development within the city is denoted as a share Λ_k of the city's circle. As the developable sector of the city houses all the inhabitants, we get $\Lambda_k \pi {\Phi_k}^2 = \gamma H_k$, where $\pi {\Phi_k}^2$ is the area of the city circle and H_k the number of households. Thus, it can be shown that $\Phi_k = \sqrt{\gamma H_k / \pi \Lambda_k}$. (Saiz 2010, 1262)

3.1.1. Housing supply

As Saiz (2010) assumes perfect competition in the construction sector, housing is sold at cost. Thus, $P(d) = CC + LC(d)$, where housing price $P(d)$ at distance d from the city centroid is a sum of construction costs CC and the cost of land at that distance $LC(d)$. Assuming complete financial markets, housing prices are equal to the discounted value of rents. Thus, we get $r(d) = iCC + iLC(d)$, where i is the discount factor. Assuming that $LC(\Phi_k) = 0$ and applying equation 1, we get

$$r_0 = iCC + t\sqrt{\frac{\gamma H_k}{\pi \Lambda_k}}. \tag{2}$$

Using equation 2, it can be shown that the average rents can be expressed as the rents at a distance of two-thirds of the city's radius from the centroid to the perimeter, or that $\tilde{r_k} = r(\frac{2}{3}\Phi_k)$. Thus, Saiz derives the housing supply equation for average housing values in city k to be (full derivations can be found in Appendix II of Saiz (2010, 1287-1291))

$$\tilde{P}_k^S = CC + \frac{1}{3i}t\sqrt{\frac{\gamma H_k}{\pi \Lambda_k}}. \tag{3}$$

(Saiz 2010, 1262-1263)

Following this expression of housing supplyt, Saiz derives a proposition from his model that the inverse elasticity of housing supply, β_k^S, is decreasing in land availability, or more formally

$$\frac{\partial \beta_k^S}{\partial \Lambda_k} < 0, \tag{4}$$

where $\beta_k^S \equiv \partial ln\tilde{P}_k^S / \partial lnH_k$ (as the derivation of this result is not provided by Saiz, I give it in Derivation 1 of Appendix 1). (Saiz 2010)

3.1.2. The empirical model

Saiz (2010) defines the initial share of construction costs on housing prices as $\sigma_k \equiv CC/\tilde{P}_k$ and assumes that $\frac{\partial \tilde{P}_k}{\partial H_k} = \frac{\partial LC(H_k)}{\partial H_k}$, which is to say that the effect of changes in housing demand on housing prices comes through changes in land values (LC). Hence, the expression of average housing prices as the sum of construction costs and land values ($\tilde{P}_k = CC + LC(H_k)$) can be reformulated as

$$\frac{d\tilde{P}_k}{\tilde{P}_k} = \frac{d(CC + LC(H_k))}{\tilde{P}_k} = \frac{dCC}{\tilde{P}_k} + \frac{dLC(H_k)}{\tilde{P}_k} = \frac{CC}{\tilde{P}_k}\frac{dCC}{CC} + \left[\frac{dLC(H_k)}{dH_k}\frac{H_k}{\tilde{P}_k}\right]\frac{dH_k}{H_k},$$

which then becomes

$$dln\tilde{P}_k = \sigma_k dlnCC + \beta_k^S dlnH_k \tag{5}$$

(Saiz 2010, 1266).

Now, the relationship between the inverse housing supply elasticity and land availability shown in equation 4 can be linearly approximated as $\beta_k^S \approx \tilde{\beta}^S + (1 - \Lambda_k)\beta^{LAND}$, where $\tilde{\beta}^S$ is the intercept of the relation and $(1 - \Lambda_k)\beta^{LAND}$ gives the linearly approximated contribution of land availability to the inverse supply elasticity β_k^S (this is shown in Derivation 2 of Appendix 1 of this thesis). Applying this approximation to equation 5 and adding some control variables, we get the empirical form of equation 5 with unavailable land share $(1 - \Lambda_k)$

$$dln\tilde{P}_k^S = \sigma_k dlnCC + \tilde{\beta}^S dlnH_k + \beta^{LAND}(1 - \Lambda_k)dlnH_k + \sum_s R_{s,k} + \varepsilon_k, \tag{6}$$

where $\sum_s R_{s,k}$ are regional fixed effects dummies and ε_k is an error term. (Saiz 2010, 1266-1268)

3.2. Further development

Taking the Saiz (2010) model as a starting point, here I will take some steps towards allowing more within city heterogeneity in geographical constraints.

3.2.1. Allowing variation in developable land distribution

As can be seen from the Saiz model, it implicitly assumes that Λ_k is constant between the city centroid and the outer brim of the city's land area. Let $\Lambda(d)$ denote the share of developable land at distance d from the city centroid, thus the assumption can be expressed more formally as $\Lambda(d) = \Lambda_k,\ \forall d \in [0, \Phi_k]$. This, of course, may not be the case in reality. Moreover, it is plausible that geographic constraints closer to the city centre have larger effects on the housing market than similar constraints nearer the perimeter. Therefore, it could be of interest to derive an empirically estimable model that allows $\Lambda(d)$ to variate within the city radius. However, as topographical features tend to be both spatially autocorrelated and correlated with each other, this would risk adding multicollinearity to the model. Therefore, I will instead aim to incorporate a measure of average distance from the city centre to the land areas unavailable for development. This will allow me to take into account at least some of the variation in unavailable land distribution without causing multicollinearity.

Now, assume that $\Lambda(d)$ takes the form of a constant interval step function with n intervals within the radius Φ_k, so that each interval has the length $l = \frac{\Phi_k}{n}$ and within each interval j $\Lambda(d) = \Lambda_{j,k}$, a constant. Henceforth, I will refer to the circular areas defined by intervals j as zones. Now, average distance $\bar{d}_k$ to developable land can be expressed as the weighted average of the average zonal distances from the city centre. The weights are zonal shares of the total developable land. More formally for *n* zones: $\bar{d}_k = \sum_{j=1}^{n}(\Lambda_{j,k}A_j/\Lambda A)\, d_j$, where A_j and A are zonal and total surface areas respectively, Λ is the total share of developable land and d_j is the average distance from the city centre to zone j.

Relying otherwise on the standard Saiz model, it can be shown (see Derivation 3 in Appendix 1) that the average housing prices in city k will be (analogously to equation 3)

$$\tilde{P}_k^S = CC + \frac{1}{i}t\left(\sqrt{\frac{\gamma H_k}{\Lambda_k \pi}} - \bar{d}_k\right). \tag{7}$$

Equation 7 simplifies to its equivalent in the standard Saiz model, when $\Lambda(d) = \Lambda_k$, $\forall d \in [0, \Phi_k]$, or in other words, when $\Lambda(d)$ is uniformly distributed. When this is the case, $\bar{d}_k = \frac{2}{3}\sqrt{\frac{\gamma H_k}{\Lambda_k \pi}}$.

3.2.2. Decomposing the unavailable land measure

Furthermore, the measure for land unavailability that Saiz uses is a union of the area shares of three different sources of constraints: water bodies, areas with slopes over 15 %, and wetlands. This means that $(1 - \Lambda_k) = u_{wa} \cup u_{sl} \cup u_{we}$, where u_{wa}, u_{sl} and u_{we} are the unavailable land shares due to water, steep slopes and wetland respectively. In practice this simplifies to $(1 - \Lambda_k) = u_{wa} + u_{sl} + u_{we}$, as these different components do not overlap. While all these sources are likely to cause constraints on urban development, their effects might differ. Hence, instead of using the total area share within these constraints as my measure of land unavailability, I will treat them as heterogeneous sources of constraint. In order to achieve this, I will use the following aggregation model

$$(1 - \Lambda_k) = \alpha_{wa} u_k^{wa} + \alpha_{sl} u_k^{sl} + \alpha_{we} u_k^{we}, \tag{8}$$

where α_{wa}, α_{sl} and α_{we} are aggregation coefficients. Defining $\beta_{wa} = \alpha_{wa}\beta^{LAND}$, $\beta_{sl} = \alpha_{sl}\beta^{LAND}$ and $\beta_{we} = \alpha_{we}\beta^{LAND}$, I get

$$(1 - \Lambda_k) = \frac{\beta_{wa}}{\beta^{LAND}} u_k^{wa} + \frac{\beta_{sl}}{\beta^{LAND}} u_k^{sl} + \frac{\beta_{we}}{\beta^{LAND}} u_k^{we}, \tag{9}$$

which I will use later in the empirical model.

3.2.3. The empirical model

Intuitively $\frac{\partial \beta_k^S}{\partial \Lambda_k} < 0$ holds for the extended model as well (see Derivation 4 in Appendix 1). Thus, the first order approximation of equation 4 holds for this model as it is. Moreover, it can be shown (see Derivation 5 in Appendix 1) that a similar proposition holds as well: $\frac{\partial \beta_k^S}{\partial \overline{ud}} > 0$, where $\overline{ud}$ is the average distance to undevelopable land from the city centre. This, however, only holds when $(1 - \Lambda_k) > 0$. When $(1 - \Lambda_k) = 0$, $\frac{\partial \beta_k^S}{\partial \overline{ud}} = 0$ as well. Therefore, the effect of the average distance measure $\overline{ud}$ on the inverse supply elasticity is catalysed by the share of undevelopable land. This can be expressed through an interaction term. Thus, I can form a similar first order linear approximation for these relationships as the one used in the original Saiz model: $\beta_k^S = \tilde{\beta}^S + \beta^{LAND}(1-\Lambda_k) + \beta_{ud}\overline{ud}(1-\Lambda_k)$. Moreover, as I do not have access to city-level data on construction costs, I will leave them out. Hence the empirical log-linearized form of the model can be (analogously to equation 6) expressed as

$$dln\tilde{P}_k^S = \beta_0 + \tilde{\beta}^S dlnH_k + \beta^{LAND}(1-\Lambda_k)dlnH_k + \beta_{ud}\overline{ud}(1-\Lambda_k)dlnH_k + \sum_s R_{s,k} + \varepsilon_k, \quad (10)$$

where β_0 is the regional constant for region 0 (the constant in the regression results). Substituting the aggregation model of equation 9 to equation 10 I get

$$dln\tilde{P}_k^S = \beta_0 + \tilde{\beta}^S dlnH_k + \beta_{wa}u_k^{wa}dlnH_k + \beta_{sl}u_k^{sl}dlnH_k + \beta_{we}u_k^{we}dlnH_k + \beta_{ud}\overline{ud}(1-\Lambda_k)dlnH_k + \sum_s R_{s,k} + \varepsilon_k. \quad (11)$$

Moreover, note that the average distance to undevelopable land $\overline{ud}$ increases whenever the distance to a single component of unavailable land increases, so that: $\frac{\partial\overline{ud}}{\partial\overline{ud}_{wa}} > 0$, $\frac{\partial\overline{ud}}{\partial\overline{ud}_{sl}} > 0$ and $\frac{\partial\overline{ud}}{\partial\overline{ud}_{we}} > 0$, where the variables in the denominators are the average distances from the city centre to waterbodies, steep slopes and wetlands, respectively. Hence, $\overline{ud}$ can be linearly approximated as

$$\overline{ud} = \alpha_{wa}^{ud}\overline{ud}_{wa} + \alpha_{sl}^{ud}\overline{ud}_{sl} + \alpha_{we}^{ud}\overline{ud}_{we}. \quad (12)$$

Defining $\beta_{wa}^{ud} = \alpha_{wa}^{ud}\beta_{ud}$, $\beta_{sl}^{ud} = \alpha_{sl}^{ud}\beta_{ud}$, $\beta_{we}^{ud} = \alpha_{we}^{ud}\beta_{ud}$ and substituting these to equation 12 I get

$$\overline{ud} = \frac{\beta_{wa}^{ud}}{\beta_{ud}}\overline{ud}_{wa} + \frac{\beta_{wa}^{ud}}{\beta_{ud}}\overline{ud}_{sl} + \frac{\beta_{wa}^{ud}}{\beta_{ud}}\overline{ud}_{we} \quad (13)$$

Now, combining this with equation 11 gives me

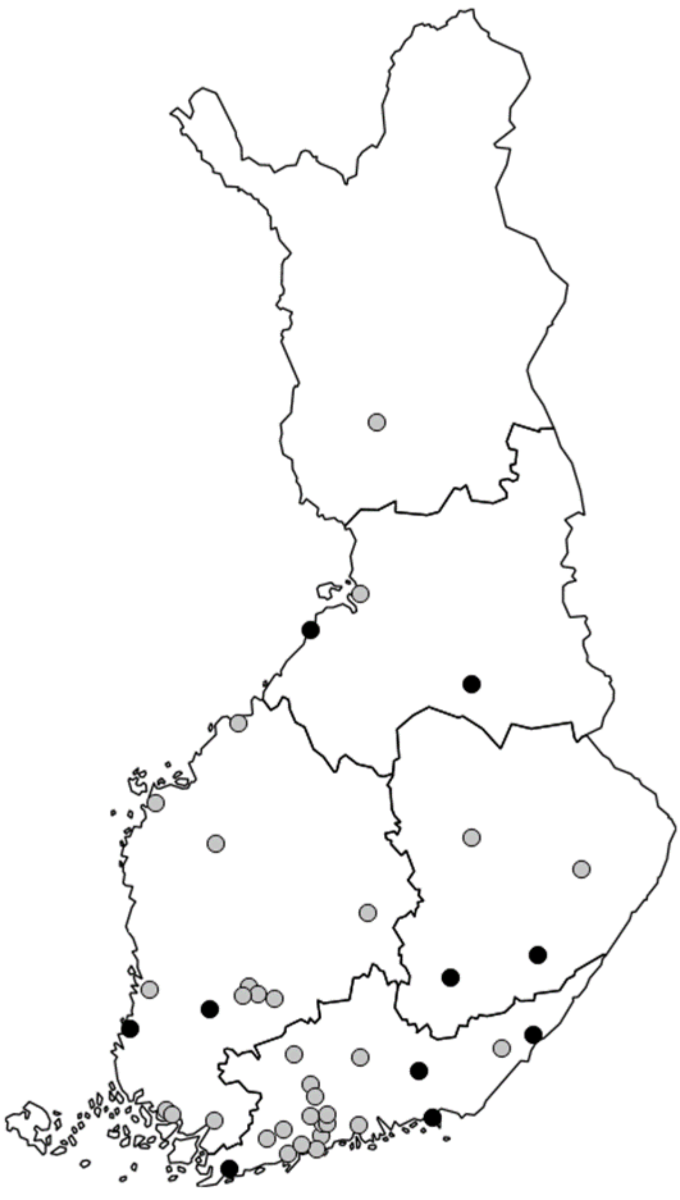
$$dln\tilde{P}_k^S = \beta_0 + \tilde{\beta}^S dlnH_k + \beta_{wa} u_k^{wa} dlnH_k + \beta_{sl} u_k^{sl} dlnH_k + \beta_{we} u_k^{we} dlnH_k + \beta_{wa}^{ud} \overline{ud}_{wa} u_k^{wa} dlnH_k + \beta_{wa}^{ud} \overline{ud}_{sl} u_k^{sl} dlnH_k + \beta_{wa}^{ud} \overline{ud}_{we} u_k^{we} dlnH_k + \sum_s R_{s,k} + \varepsilon_k, \qquad (14)$$

which is the final empirical equation I will apply in the empirical part of this paper.

4. DATA

In this section of the thesis, I will introduce and describe the data used in the empirical estimations in chapter 5. The data has been collected for those 42 Finnish cities that have had a population of over 25 000 according to the 2016 population figures. I use the 2017 administrative boundaries for all city level data. First, I will introduce and discuss the newly created geographical dataset on land unavailability for Finnish cities based on the work of Saiz (2010). After the new data set has been discussed, I will present and review the housing market data utilized in this thesis.

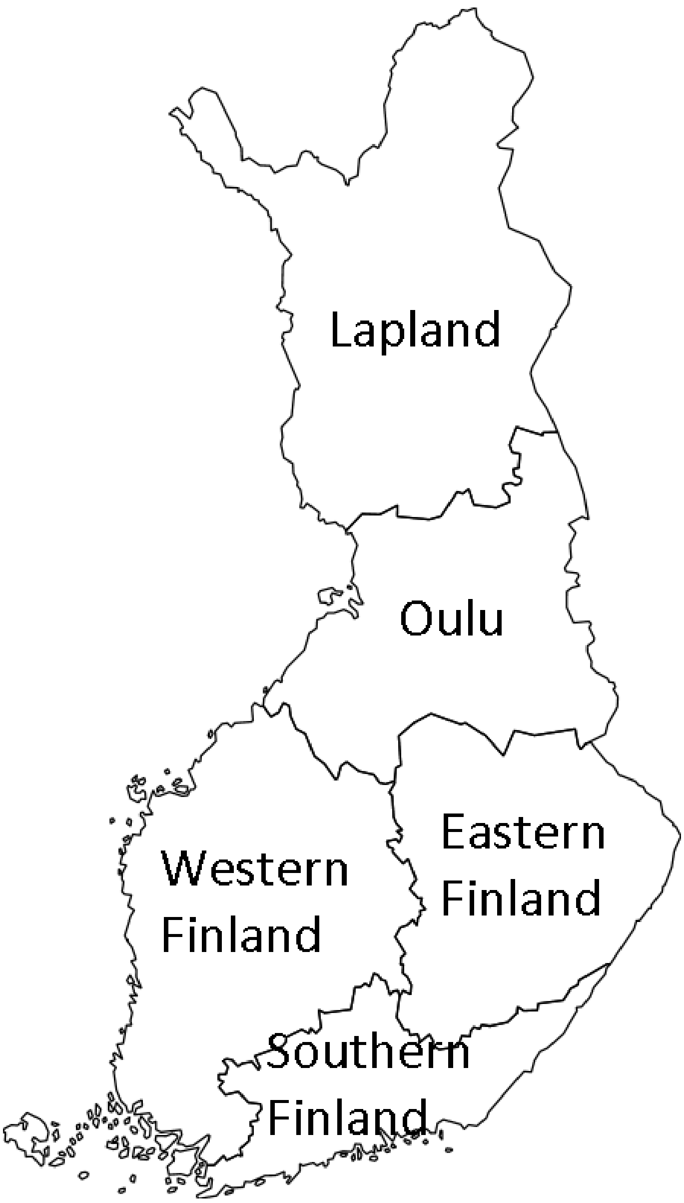
FIGURE 1. Positions of the 42 sample cities within Finland.



In figure 1 the locations of the 42 sample cities within Finnish national and provincial borders can be seen. Furthermore, those cities that have faced negative population change between 2003 and 2015 are marked with black (10 cities).

Like Saiz (2010), I use regional dummies ($\sum_s R_{s,k}$) in the regression equations to control for regional fixed effects. The regional division applied in this study is the provincial division that was in place in Finland between 1997 and 2009. However, the northern Finland provinces of Oulu and Lapland are merged into one region, as otherwise the province of Lapland would only comprise of the city of Rovaniemi. The locations of the provinces can be seen in Figure 2.

FOGURE 2. Provinces of Finland.



4.1. Sources of land unavailability

As in Saiz (2010), in order to measure land unavailability, data for three different sources of geographic constraints are utilized. These consist of satellite measured data for waterbodies and altitude. However, as no data for all wetlands in Finland seems to be readily available, I will instead use data for swamp areas. Even though swamps are very common in Finland and most likely constitute the vast majority of local wetlands, this might lead to different results compared to the application of a total wetland measure. This is because the between-cities distribution of swamps might differ from the equivalent distribution of other wetlands. For example, swamps might be more common in the north of Finland relative to other wetlands than is the case in the south of Finland.

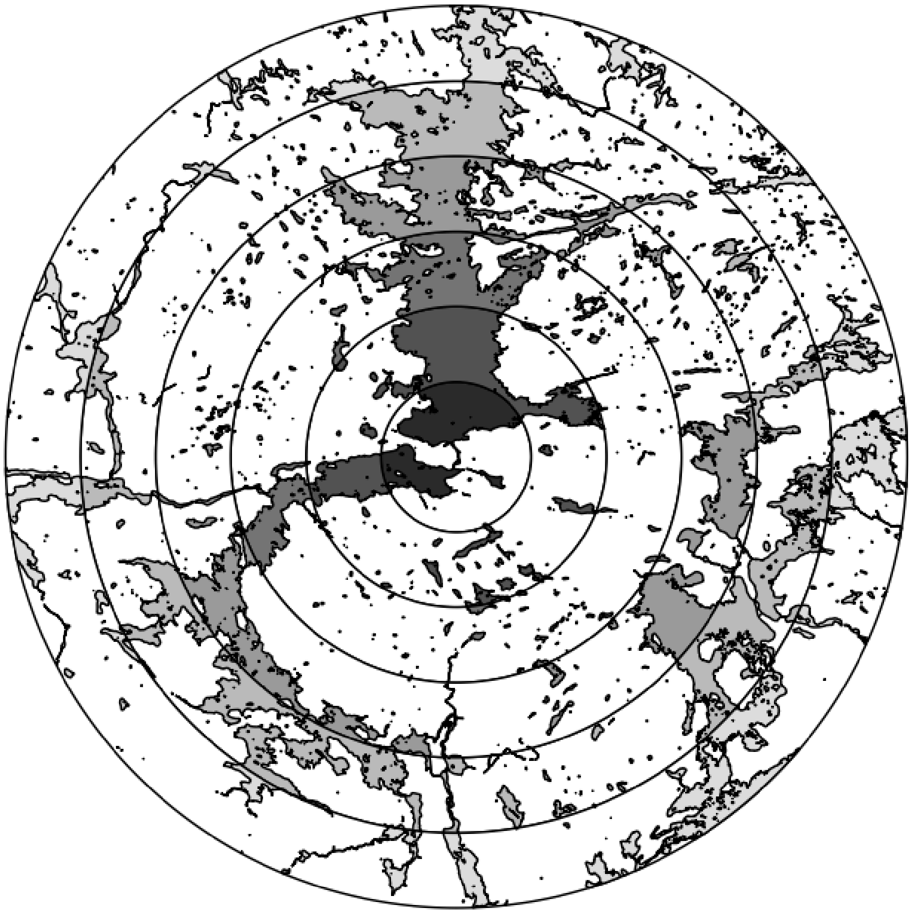
All data on these topographical features has been provided by the National Land Survey of Finland. This raw data has then been manipulated with geographic information system (GIS) software to obtain necessary measurements for the empirical analyses. For example, altitude data, which is provided in 10 m x 10 m cells, has been converted into data on slopes using the standard Horn's formula (Horn 1981 p. 20-21).

Saiz (2010) measures these geographic features within a 50 km radius from the centres of US metropolitan areas. In this thesis, however, I will utilize a 30 km radius from a city centre point provided by the National Land Survey of Finland. The shorter measurement distance is appropriate in the Finnish context, as metropolitan areas have significantly fewer inhabitants than in the US. Moreover, Finnish cities tend to be more densely populated than US cities with similar population size (Oikarinen et al. 2015, p. 25). This in turn means that, other things equal, US cities need larger land areas to house a certain number of inhabitants than Finnish cities. Taken together, these stylized facts imply that Finnish cities cover significantly smaller areas than their US counterparts.

In addition to total surface areas within a 30 km radius, for each source of geographic constraint, the surface areas within six circular zones or bands from the city centroids are measured. All the zones have a constant width of five kilometres. An illustrative example of this can be seen below in figure 3, where the waterbodies (grey areas) within a 30 km distance from

Tampere city centre have been divided into six zones. The zones in the figure are represented with zonal borders and different shades of grey for clarity.

FIGURE 3. Waterbodies and circular zones around Tampere city centre.



This zonal division allows me to apply the zonal version of the Saiz model, introduced in section 3.2, by making it possible to calculate an approximate average distance from the city centroid to the geographical constraints. In other words, this makes it possible to approximate the position of the mass of the measured topographical features in relation to the distance from the city centre.

Below in table 1, some descriptive statistics are provided for geographical variables. As can be seen, the water share within a 30-km radius is notably larger on average than either the share of steep slopes or swamps within the same area. This is to be expected, as most of the cities lie either close to the Baltic see or are located in the immediate vicinity of multiple lakes.

Moreover, Finland is a comparatively flat country: there are no major mountain ranges within its borders.

TABLE 1. Descriptive statistics for geographical variables

Variable	Mean (standard dev.)	Coefficient of variation
Water share, 30-km radius	0.218 (0.141)	0.647
Steep slope share, 30-km radius	0.083 (0.048)	0.578
Swamp share, 30-km radius	0.065 (0.057)	0.877
Total share, 30-km radius	0.366 (0.120)	0.327
Average distance to water areas	20.686 (1.567)	0.076
Average distance to steep slopes	19.618 (1.083)	0.055
Average distance to swamps	20.927 (1.330)	0.064
Water share 0-5km zone	0.166 (0.151)	0.910
Water share 5-10km zone	0.188 (0.150)	0.798
Water share 10-15km zone	0.189 (0.141)	0.746
Water share 15-20km zone	0.218 (0.149)	0.683
Water share 20-25km zone	0.223 (0.150)	0.673
Water share 25-30km zone	0.239 (0.156)	0.653

TABLE 1. (Continued) Descriptive statistics for geographical variables

Variable	Mean (standard dev.)	Coefficient of variation
Steep slope share 0-5km zone	0.094 (0.112)	1.191
Steep slope share 5-10km zone	0.088 (0.056)	0.636
Steep slope share 10-15km zone	0.090 (0.057)	0.633
Steep slope share 15-20km zone	0.084 (0.051)	0.607
Steep slope share 20-25km zone	0.083 (0.049)	0.590
Steep slope share 25-30km zone	0.079 (0.045)	0.570
Swamp share 0-10km zone	0.023 (0.028)	1.217
Swamp share 5-10km zone	0.049 (0.055)	1.122
Swamp share 10-15km zone	0.060 (0.060)	1.000
Swamp share 15-20km zone	0.064 (0.060)	0.938
Swamp share 20-25km zone	0.069 (0.059)	0.855
Swamp share 25-30km zone	0.070 (0.045)	0.643

As can be noted from the table, the coefficient of variation for the total unavailable land share is significantly smaller than that of water, steep slopes or swamps. Hence, it can be deduced that these three variables are most likely negatively correlated. This is confirmed by correlation coefficients, which are -0.186 for water bodies and swamp, -0.374 for water and steep slopes and -0.553 for swamps and steep slopes. This is as expected, due to the finite observation area: the sum of these three components must be equal or below 1. Moreover, there are geographical reasons behind these correlation coefficients: both swamps and water bodies are rare in areas with high elevation differences.

As measured by the coefficient of variation, the average distance measures do not variate nearly as much across cities as the surface area share measures. Therefore, even though the average distance measure for a uniformly distributed $\Lambda(d)$, 20 km, is quite close to the mean of average distance measures for both water bodies and swamps, they differ from it statistically significantly at the 5% level.

More interesting, however, is that the mean average distance to steep slopes is very low, only 19.618 km. In the low lying nature of most of Finnish landscape, at least part of this could be attributable to both long eskers and river valleys crossing through the observation circle. As eskers and river valleys are approximately constant in width, they would contribute around the same absolute amount of steep slope areas to each zone, thus causing the average distance measure to move towards the city centroid. However, as the data on steep slopes is generated by an algorithm from 10 m x 10 m altitude measurement cells, this might also indicate problems in the data generation process.

For the zonal statistics, it can be seen that the coefficient of variation is smaller the further away from the city centre the zone lies. This is natural, as the zonal – and thus, observation – areas increase with the distance to the city centre, so the average variation within a zone should decrease. This is analogous to an increase in sample size. However, interestingly, we see a sharper drop in the coefficient of variation between the 0-5 km slope zonal share and the 5-10 km slope zonal share than we do in water bodies or swamp areas. This might suggest that some city centres are located on relatively elevated ground whereas others are located on flat land. There might be historical reasons behind this, as cities were most likely founded on easily accessible locations; both rivers (especially in lowlands) and long eskers (in more rugged terrain) might have played roles as pathways in the past.

In addition to the descriptive statistics in table 1, below in table 2 I also provide the newly produced data-set for city-level measures of the geographical variables used in the empirical estimations in chapter 5. Here, I highlight some interesting details from the table. For example, it is noteworthy that Helsinki, which is widely considered to have a highly inelastic housing supply, has the second highest water and total unavailable land shares of the 42 cities. Only Raahe, a seaside town in the North of Finland has a larger share of water and unavailable land within the 30 km radius.

TABLE 2. Geographical variables at the city level

City	Water share, 30 km	Swamp share, 30 km	Slope share, 30 km	Total share, 30 km	Average distance, water (km)	Average distance, swamp (km)	Average distance, slope (km)	Average distance, water & swamp (km)	Average distance, total (km)
Espoo	0.301	0.023	0.121	0.445	21.689	18.217	19.092	21.445	20.805
Hämeenlinna	0.112	0.089	0.068	0.270	20.597	20.222	20.641	20.431	20.484
Helsinki	0.490	0.014	0.083	0.587	19.912	23.110	21.186	20.001	20.169
Hyvinkää	0.036	0.067	0.101	0.203	21.563	18.944	20.201	19.867	20.032
Imatra	0.195	0.044	0.058	0.296	19.525	21.900	18.906	19.963	19.757
Järvenpää	0.028	0.048	0.112	0.189	22.622	21.457	21.348	21.887	21.567
Joensuu	0.320	0.070	0.061	0.451	19.997	20.232	21.250	20.039	20.204
Jyväskylä	0.169	0.058	0.116	0.343	19.190	22.777	17.044	20.101	18.976
Kaarina	0.200	0.014	0.123	0.337	20.419	23.553	19.135	20.622	20.077
Kajaani	0.224	0.211	0.028	0.462	19.641	21.507	19.901	20.545	20.506
Kangasala	0.233	0.042	0.086	0.362	19.389	20.710	19.986	19.593	19.686
Kerava	0.106	0.035	0.105	0.246	25.202	22.256	19.699	24.463	22.430
Kirkkonummi	0.421	0.022	0.110	0.553	21.907	17.795	18.812	21.706	21.131
Kokkola	0.376	0.109	0.004	0.490	21.024	19.260	19.120	20.627	20.743
Kotka	0.519	0.039	0.025	0.583	19.721	21.210	21.184	19.824	19.883
Kouvola	0.083	0.071	0.080	0.233	20.567	21.605	21.216	21.045	21.073
Kuopio	0.280	0.045	0.083	0.408	17.957	22.663	19.430	18.603	18.771
Lahti	0.102	0.039	0.137	0.277	19.958	21.098	20.139	20.271	20.206
Lappeenranta	0.243	0.066	0.076	0.385	20.410	20.340	19.446	20.395	20.206
Lohja	0.145	0.034	0.195	0.374	10.151	21.508	19.983	20.410	20.187
Mikkeli	0.201	0.079	0.112	0.391	21.501	20.877	19.606	21.325	20.835
Nokia	0.173	0.040	0.092	0.305	19.106	21.241	19.759	19.503	19.580
Nurmijärvi	0.035	0.052	0.130	0.217	21.238	21.533	20.709	21.416	20.992
Oulu	0.276	0.151	0.003	0.431	19.807	22.013	18.187	20.587	20.569
Pori	0.249	0.073	0.016	0.338	23.091	21.880	20.029	22.817	22.684
Porvoo	0.254	0.027	0.104	0.384	22.699	19.186	18.898	22.365	21.429
Raahe	0.510	0.133	0.001	0.644	20.200	21.544	16.749	20.477	20.470
Raasepori	0.428	0.028	0.107	0.563	21.490	19.047	19.246	21.338	20.942
Rauma	0.406	0.036	0.013	0.455	20.930	21.345	16.982	20.964	20.854
Riihimäki	0.043	0.086	0.094	0.222	20.764	21.028	19.234	20.940	20.220

TABLE 2. (Continued) Geographical variables at the city level

City	Water share, 30 km	Swamp share, 30 km	Slope share, 30 km	Total share, 30 km	Average distance, water (km)	Average distance, swamp (km)	Average distance, slope (km)	Average distance, water & swamp (km)	Average distance, total (km)
Rovaniemi	0.063	0.312	0.028	0.403	18.368	20.208	19.839	19.900	19.896
Salo	0.050	0.044	0.133	0.227	20.447	20.371	19.186	20.411	19.702
Sastamala	0.068	0.075	0.057	0.200	19.316	21.027	20.064	20.209	20.167
Savonlinna	0.367	0.058	0.111	0.537	20.381	19.402	19.198	20.248	20.030
Seinäjoki	0.024	0.171	0.007	0.203	19.383	19.470	20.421	19.460	19.495
Tampere	0.216	0.039	0.094	0.349	19.234	20.546	20.776	19.435	19.796
Turku	0.210	0.015	0.113	0.338	21.312	22.900	19.290	21.417	20.707
Tuusula	0.089	0.039	0.115	0.243	25.181	22.820	20.511	24.471	22.597
Vaasa	0.359	0.048	0.005	0.412	21.187	21.452	18.851	21.218	21.190
Vantaa	0.262	0.018	0.105	0.385	23.001	20.025	19.663	22.807	21.950
Vihti	0.089	0.036	0.186	0.311	19.859	19.816	18.820	19.846	19.234
Ylöjärvi	0.195	0.042	0.100	0.336	18.881	20.835	20.298	19.224	19.541

Also of interest is that many of the smaller cities in the north of the Helsinki capital region, such as Järvenpää, Kerava, Nurmijärvi and Tuusula have total unavailable land shares well below the average of 0.366. What makes this observation interesting, is that a perceived trend towards urban sprawl was very much present in the public discourse in Finland during the sample period.

Average distance measures for water bodies and swamps seem to be quite evenly distributed around the benchmark uniform distribution's 20 km. Some figures do jump out however. For example, the average distance to water bodies from Kerava is over 25 km, which is due to the Baltic sea crossing the observation circle of the city just at its southern extremity in an otherwise relatively water bodiless area. On the other side of the distribution lies Kuopio, which city centre is almost completely surrounded by lake Kallavesi. Hence, the average distance to water bodies from the city centroid is less than 18 km.

The longest average distances to swamps are from the city centres of Kaarina and Helsinki, more than 23 km. Interestingly, these cities have the smallest total shares (0.014) of swamp within their observation circles. In contrast, the shortest distances to swamps are from Kirkkonummi and Espoo, both of which are located in the Helsinki capital region. As both of these cities have very low shares of swamps within their 30 km radiuses, this might be attributable to a national park located next to them. Of course, these extremes might be at least partially attributable to the low shares of swamps: small variations and measurement errors could lead to large swings in the average distance figures for these cities.

Average distance measures for steep slopes tend to be below the benchmark uniform distribution's 20 km. Especially noteworthy are the very low average distances of Raahe and Jyväskylä, only 16.982 km and 17.044 km respectively. These low figures might be at least partially explained by long, uniform geographical features: Raahe is located at a riverside in the low-lying region of Ostrobothnia and Jyväskylä is situated on an esker, next to the oblong lake Päjiänne.

4.2. Housing market data

For housing market variables I use data from Statistics Finland. Both yearly housing prices and housing supply data span from 2003 to 2015. Hence, the time span is significantly shorter than the 1970-2000 time period Saiz uses in his study. This might lead to highly differing results, especially as the cross-section estimation uses only a two period change in housing variables. Other things equal, this study should expect to find housing supply more inelastic than in Saiz (2010), as there should be more supply constraints at work in the shorter term.

The data for housing prices consists of average housing prices per square meter for privately financed apartments, terraced and semi-detached houses that are administrated through housing corporations and sold within each year. The price data from Statistics Finland comes in nominal form so I use the harmonised consumer price index (HCPI, 2000 = 100) to convert the data into real prices. Moreover, the quantities of yearly transactions are provided. I use this data as there is no readily available data for the average level of total housing prices. If a no arbitrage condition holds, the changes of the prices I have should reflect the changes of the more general housing price level. However, as a no arbitrage condition is a rather strong assumption, this adds uncertainty to the estimation results. For example, if demand for detached housing increases so that their prices increase more than those of other forms of housing, the price measure used will underestimate the price increases. If housing supply reacts to the increase in prices of detached houses, the estimates for supply elasticity will be artificially elastic. If the prices of detached houses increase less than those of other forms of housing, the estimates will be too inelastic. During the sample period, the price index for detached housing increased (20.1%) less than the index of housing prices used in this thesis (31.2%) in the Helsinki metropolitan area (Helsinki, Espoo and Vantaa) as a whole. In the rest of Finland combined, the opposite was true (19.4% vs 14.6%). Reassuringly however, for the whole of Finland relative increase in real prices of detached housing between 2003-2015 (19.5%) was fairly close to that of the price measure used (22.0%). Moreover, the possible bias from regional differences is mitigated by regional dummies in the regression analysis.

As a measurement of housing supply, I use the number of households within a city's administrative borders. As the main purpose of this thesis is to replicate Saiz's research and Saiz uses the number of households as well (Saiz 2010, p. 1269), this seems fitting. However, it is noteworthy that the estimation results remain very similar when the quantity of housing units is used instead.

Below in table 3 some descriptive statistics are provided for housing market variables. Notably, the coefficient of variation is almost identical for the number of households in 2003 and 2015, while at the same time there is significant variation in the growth rate of the quantity of households between cities. This is explained by the fact that within the 42-city sample of this study, while large central cities have experienced faster household growth than peripheral small ones, medium sized towns and cities close to larger cities have grown even faster. Here, we can see the urban sprawl described in the previous section in action. Due to this kind of heterogeneity in growth rates, the coefficient of variation for the number of households has remained at the same level. This has happened, as on the average, even though the mean number of households has risen significantly, some cities below the average in 2003 have moved closer to the average in 2015 on relative terms.

Descriptive statistics for average housing price measurements are also provided. However, as can be seen from the last row of table 3, the number of transactions these prices are based on varies greatly. Therefore, as the descriptive statistics for average housing prices are calculated using non-weighted data, these statistics should be considered somewhat unreliable. However, there seems to have been divergence in the housing prices, as the coefficient of variation of average real housing prices has increased from 2003 to 2015. This is also supported by the large variation in the relative changes in housing prices.

TABLE 3. Descriptive statistics for housing market variables

Variable	Mean (standard dev.)	Coefficient of variation
Number of households in 2003	37090.38 (46664.53)	1.258
Number of households in 2015	42313.83 (53211.41)	1.258
Relative change in households from 2003 to 2015	0.126 (0.068)	0.540
Average real housing price per sqm in 2003	1189.543 (307.270)	0.258
Average real housing price per sqm in 2015	1470.472 (436.792)	0.295
Relative change in real average housing price per sqm from 2003 to 2015	0.203 (0.091)	0.449
Average of quantities sold in 2003 and 2015	1273.381 (1907.963)	1.498

In addition to the descriptive statistics in table 3, the same housing market variables are provided in table 4 for all the 42 cities. Moreover, those cities that have experienced a decline in population during the sample period have been marked with an asterisk. As is to be expected, these cities have also faced negligible growth in the number of households. Interestingly, however, none of the cities have undergone actual declines in household figures. This might be at least partially attributable to a tendency in declining cities of the youth to move to growing urban areas, while their parents stay behind. Thus, the number of households would stay roughly the same, while population would decrease. This is a reassuring result for the application of the number of households as a proxy for housing supply: the number of housing units themselves are expected to decline only with a long lag, as real estate deteriorates slowly (Goodman 2005).

TABLE 4. Housing market variables at the city level

City	Number of households in 2003	Number of households in 2015	Relative change in households	Avg. real housing price in 2003	Avg. real housing price in 2015	Relative change in avg. housing price	Avg. of quantities sold in 2003 and 2015
Espoo	95561	117424	0.206	1963	2576	0.272	3731.5
Hämeenlinna	30353	34065	0.115	1187	1469	0.214	1068
Helsinki	286167	321381	0.116	2355	3150	0.291	11467
Hyvinkää	19746	22349	0.124	1242	1544	0.218	713.5
Imatra*	14702	14731	0.002	891	808	-0.097	306
Järvenpää	16234	19171	0.166	1500	1943	0.258	699.5
Joensuu	34135	38882	0.130	1128	1405	0.219	940.5
Jyväskylä	57885	68679	0.171	1250	1480	0.169	2007
Kaarina	11642	14152	0.195	1128	1496	0.282	377.5
Kajaani*	17730	18565	0.046	907	1034	0.131	478.5
Kangasala	11087	13239	0.177	1103	1453	0.276	317
Kerava	14106	16915	0.182	1464	1791	0.202	595
Kirkkonummi	12534	15854	0.235	1422	1824	0.250	372.5
Kokkola	18410	20959	0.130	913	1361	0.399	346.5
Kotka*	27155	28084	0.034	845	1032	0.200	831.5
Kouvola*	42693	43614	0.021	859	954	0.105	967.5
Kuopio	51735	58736	0.127	1217	1516	0.220	1653
Lahti	56067	61930	0.099	1045	1386	0.283	2022.5
Lappeenranta	34302	37152	0.080	1151	1426	0.214	782
Lohja	19456	21728	0.110	1220	1411	0.146	469.5
Mikkeli*	25769	27669	0.071	1073	1380	0.252	633.5
Nokia	12415	14957	0.186	1111	1355	0.198	465.5
Nurmijärvi	13617	16798	0.210	1459	1723	0.166	420
Oulu	75279	94226	0.224	1290	1363	0.055	2609.5
Pori	40839	43707	0.068	929	1124	0.191	1167
Porvoo	20124	22855	0.127	1373	1906	0.328	525.5
Raahe*	10984	11477	0.044	809	960	0.171	164
Raasepori*	12851	13641	0.060	1072	1161	0.079	262
Rauma*	19182	19802	0.032	923	1100	0.176	556
Riihimäki	12658	14405	0.129	1077	1270	0.165	490.5
Rovaniemi	26108	30292	0.149	947	1224	0.257	789.5
Salo	24458	26155	0.067	953	957	0.004	729
Sastamala*	11408	12008	0.051	858	1017	0.171	171.5
Savonlinna*	18358	18426	0.004	848	1067	0.231	438
Seinäjoki	23334	29375	0.230	1005	1442	0.361	760
Tampere	101632	119334	0.161	1416	1857	0.271	4189.5
Turku	91712	100153	0.088	1246	1573	0.233	3612.5
Tuusula	13337	15986	0.181	1500	1787	0.175	361.5
Vaasa	29665	33602	0.125	1136	1462	0.252	1021
Vantaa	81639	98885	0.192	1591	2029	0.243	3248
Vihti	10127	12391	0.202	1372	1480	0.076	389
Ylöjärvi	10600	13427	0.236	1185	1463	0.210	333

Notes: * city has experienced a decline in population between 2003 and 2015. Base year for real prices is 2000.

Other interesting observations from table 4 can be made. For example, both Oulu and Vihti have experienced above average (0.126) growth in the number of households (0.224 and 0.202

respectively), yet have housing price increases (0.055 and 0.076) well below the mean (0.203) of the 42 cities. Oulu is situated in a very flat landscape, but also has faced a severe economic downturn within the sample period. Vihti has a below average share of undevelopable land within its radius, but the real price increase can also be unreliable due to the relatively small number of transactions (389 on average) the figure is based upon.

Moreover, some cities, such as Helsinki, Turku and Lappeenranta have experienced both below average increases in the number of households and above average increases in real housing prices. Both Helsinki and Lappeenranta also have above average unavailable land measures. With the below average shares of unavailable land in cities like Vihti that have experienced above average household growth and below average price increases, this suggests qualitatively that the geographical variables might indeed have an impact on housing supply elasticities. This proposition will be examined more quantitatively in the next chapter.

5. RESULTS

In this chapter I will provide estimation results for Finnish city level housing supply equations. First I will introduce the methodology used. Second, I introduce results that follow the empirical estimations of Saiz (2010) as closely as possible. Then I will move on to estimate the expanded empirical versions of the Saiz model introduced in section 3.2.3. I will also estimate some of the specifications for only those cities that have experienced non-negative population changes in the sample period. Furthermore, I discuss some issues with the applied methodology and, as a robustness test, conduct a fixed city and year effects panel regression without instrumental variables for the period 2003-2015.

After this, I provide and discuss both descriptive statistics and city-level figures for the calculated housing supply elasticities in different model specifications. Moreover, as the supply elasticities are non-linear combinations of the regression coefficients and geographical variables in all but the panel configuration, I also provide estimates of marginal effects and the effects on housing supply elasticities by one standard deviation changes from the average variable levels in order to analyse the economic significance of the estimated results.

5.1. Methodology

The methodology used in the cross-section regression estimations is, as with Saiz (2010), two-stage least squares (2SLS) instrumental variable estimation. In the following estimations, the relative change of households ($dlnH_k$) between 2003 and 2015 is instrumented with three demand shifters: the share of pensioners of total population in 2003, distance from the city centre to Helsinki (the capital city) and distance from the city centre to Tampere (the central city of the second largest metropolitan area). The share of pensioners in 2003 is intended to act as an proxy for overall city vitality in that year as a large proportion of pensioners within a city indicates that the city is not attractive to younger cohorts. In addition to these variables, in later model specifications also interactions between the geographical and instrumental variables are employed as instruments. This is justifiable, as geographical variables are both interacted with the instrumented variable in the regression equation and can be seen as exogenous to the price increases.

The dependent variable in the estimations is the relative change in real average housing prices per square meter between 2003 and 2015. Thus, as pointed out in section 3.1.4 of this thesis, inverse housing supply elasticities can be straightforwardly derived from the regression coefficients and, therefore, also the housing supply elasticities themselves are calculatable.

Moreover, a weight adjustment to the residuals is made using the average of 2003 and 2015 quantities of the transactions to which the average housing price measurement is based on. This is due to the large between-city variation in the number of transactions, which means that the average housing values are more reliable for those cities which have had more transactions.

As the cities are mostly located in the south of Finland and some of their observation circles overlap, I also tested for spatial correlation in the regression residuals of different specifications using a distance based spatial weight matrix. In the specifications with no geographical variables, the residuals were spatially correlated at the 1 % level according to the Moran's I statistic. However, in those specifications that include geographical variables, the residuals are not statistically significantly spatially correlated. Thus, endogeneity issues due to spatial correlation seem to be insignificant in the specifications with geographical variables. Interestingly, this also indicates that an important part of spatial correlation in housing market dynamics in Finland might be attributable to autocorrelation in geographical features.

5.2. Replicating Saiz (2010)

Below in table 5 I present some estimation results from Saiz (2010, 1269) for different model configurations. These are found in columns 1, 3, 4, 5 and 7. I also provide estimation results for the Finnish data for configurations *as similar as possible* to those used by Saiz. These are presented in columns 2, 6 and 8.

However, as pointed out in chapter 4.2, the time interval in these estimations differs from that applied by Saiz: due to data limitations, this thesis has a 12 year interval, whereas Saiz utilizes a 30 year difference. As at least some supply constraints can be less critical in the longer term (for example, in the case of a long-lasting land shortage, swamps can be drained, slopes can be terraced and shallow water areas filled), the difference in the sample intervals should mean that, other things equal, the regression coefficients should be larger in this thesis than in Saiz

(2010). This is because the regression coefficients are negatively related to housing supply elasticities.

Moreover, in addition to the variables I have already described in the previous chapter, Saiz also utilizes city level variables for relative increases in construction costs ($dlnCC_k$), the share of construction costs to the price of housing in 1970 (σ_k) and, in some configurations, the logarithm of the Wharton Regulation Index (Gyourko et al. 2008) that aims to measure the strictness of local land use policies (lnWRI).

As Finland is a small and relatively homogenous country, where construction companies tend to operate nationwide, construction costs should be very similar across cities (Oikarinen et al. 2015). However, through the same argument, the share of construction costs to the price of housing in the first year of the study should variate across cities, as the housing price levels in cities differ significantly. Furthermore, Saiz uses the product of these two variables ($\sigma_k dlnCC$) as a city specific constant by deducting it from the relative increase of housing prices ($dln\tilde{P}_k^S$) and using this difference as the dependent variable in the empirical estimations (Saiz 2010, 1269). Thus, the dependent variable in this study suffers from a measurement error, making the coefficient estimates less precise.

The effect of the missing interacted ΔlnH X lnWRI variable on the results can be analysed as omitted variable bias. This is possible even though the variable itself is most likely endogenous as pointed out by Saiz (2010, 1272). As can be seen by comparing Saiz's estimation results in columns 3 and 4 of table 6, the bias seems to be positive for the regression coefficients of both ΔlnH and Unavailable land X ΔlnH, as the coefficients are larger in the configuration without lnWRI X ΔlnH (column 3). This should be taken into account when comparing results in column 6 to those in columns 4 and 5. In these estimations this would mean that the regression coefficients of this study would be smaller with the specification used by Saiz (2010) – at least if the missing variable would have a positive regression coefficient and be positively correlated with the existing housing market regressors as in the dataset used by Saiz (2010). In columns 7 and 8 the situation is somewhat more complicated, as Unavailable land X ΔlnH appears not only by itself, but in an interaction with the logarithmic population level as well.

TABLE 5. Estimation results compared to Saiz (2010).

	ΔlnP (supply), Saiz: 1970-2000, This study: 2003-2015							
	(1) Saiz	(2) This study	(3) Saiz	(4) Saiz	(5) Saiz	(6) This study	(7) Saiz	(8) This study
ΔlnH	0.650 (0.107)***	0.314 (0.209)	0.336 (0.116)***	0.060 (0.215)		-0.530 (0.261)**		-0.223 (0.364)
Unavailable land X ΔlnH			0.560 (0.118)***	0.511 (0.214)***	0.516 (0.116)***	2.445 (0.728)***	-5.329 (0.904)***	-1.748 (3.348)
ln(T_0 population) X unavailable land X ΔlnH							0.481 (0.117)***	0.263 (0.217)
lnWRI X ΔlnH				0.237 (0.130)*	0.268 (0.068)***		0.301 (0.066)***	
Region 1 (Saiz: Midwest, This study: Western FI)	-0.099 (0.054)*	-0.018 (0.027)	-0.041 (0.052)	-0.015 (0.055)	-0.009 (0.050)	0.015 (0.024)	0.002 (0.049)	0.011 (0.022)
Region 2 (Saiz: South, This study: Oulu & Lapland)	-0.236 (0.065)***	-0.156 (0.056)***	-0.170 (0.062)***	-0.129 (0.069)*	-0.116 (0.050)**	-0.163 (0.061)***	-0.115 (0.048)**	-0.148 (0.062)**
Region 3 (Saiz: West, This study: Eastern FI)	0.016 (0.076)	-0.013 (0.026)	0.057 (0.072)	0.059 (0.072)	0.069 (0.063)	-0.002 (0.020)	0.035 (0.046)	0.005 (0.019)
Constant	0.550 (0.055)***	0.207 (0.041)***	0.594 (0.052)***	0.528 (0.058)***	0.601 (0.046)***	0.177 (0.034)***	0.601 (0.045)***	0.191 (0.034)***
First stage F-statistic		139.85				59.08		80.66

Notes: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Saiz: N=382, this study: N=42.

Even with these differences in estimation specifications, however, the results are fairly similar. Importantly, the regression coefficient of Unavailable land X ΔlnH is significant at the 1 % level in all but the last specification. This suggests that the unavailability of land does indeed have an impact on housing supply elasticities in the Finnish case as well. As was expected, the coefficients themselves are larger than in Saiz (2010). Due to the above mentioned differences in estimation specifications and data, it is however difficult to determine if the effects of land unavailability are in fact larger in Finland than in the US.

The last model specification in table 5 (columns 7 and 8) suggests that, unlike in the US, the effects of the unavailability of land on housing supply elasticities might not be dependent on city size in Finland. This is in line with a similar result by Oikarinen, Peltola and Valtonen (2014). The result might be attributable to smaller absolute differences in city populations in Finland compared to those in the US.

5.3. Results for the extended model

On the next page in table 6 I provide estimation results for the extended model. For convenience, I reproduce two of the columns from table 5 as well: columns 2 and 6 appear in table 6 as columns 1 and 2 respectively.

In table 7 I provide estimation results for the same model specifications as those in table 6, but only for the subsample of the 32 cities for which population change has been non-negative. This is done because the effect of geography on housing markets is plausibly dependent on whether housing demand expands or diminishes: constraints might not be effective when population decreases. This is significant in the context of this thesis, as ten of the 42 cities in the sample have faced declines in population during the sample period, which I use as an – admittedly crude – proxy for declining housing demand. Unfortunately, corresponding estimates for these ten cities could not be estimated reliably due to problems with multicollinearity. Moreover, in table 8 I provide some marginal effect estimates at average variable values for convenience, as marginal effects on supply elasticities are nonlinear.

TABLE 6. Estimation results for the extended model, all cities

	ΔlnP (supply), 2003-2015				
	(1)	(2)	(3)	(4)	(5)
ΔlnH	0.314 (0.209)	-0.530 (0.261)**	-1.133 (0.680)*	-0.524 (0.290)**	-0.650 (0.570)
Unavailable land, total X ΔlnH		2.445 (0.728)***		4.100 (3.615)	
Unavailable land, water X ΔlnH			2.743 (0.879)***		3.749 (3.049)
Unavailable land, slope X ΔlnH			4.650 (3.800)		-28.981 (12.235)**
Unavailable land, swamp X ΔlnH			8.304 (3.669)**		51.007 (13.220)***
Unavailable land, total X AD, total X ΔlnH				-0.082 (0.195)	
Unavailable land, water X AD, water X ΔlnH					-0.074 (0.174)
Unavailable land, slope X AD, slope X ΔlnH					1.664 (0.712)**
Unavailable land, swamp X AD, swamp X ΔlnH					-2.278 (0.759)***
Region 1: Western FI	-0.018 (0.027)	0.015 (0.024)	0.005 (0.027)	0.012 (0.024)	0.026 (0.024)
Region 2: Oulu & Lapland	-0.156 (0.056)***	-0.163 (0.061)***	-0.276 (0.079)***	-0.164 (0.060)***	-0.116 (0.058)**
Region 3: Eastern FI	-0.013 (0.026)	-0.002 (0.020)	-0.017 (0.020)	-0.007 (0.025)	0.007 (0.027)
Constant	0.207 (0.041)***	0.177 (0.034)***	0.193 (0.033)***	0.179 (0.034)***	0.170 (0.030)***
First stage F-statistic	139.85	59.08	54.65	39.82	151.23

Notes: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. N=42. AD = average distance.

TABLE 7. Estimation results for the extended model, cities with non-negative population change

	ΔlnP (supply), 2003-2015				
	(1)	(2)	(3)	(4)	(5)
ΔlnH	0.022 (0.262)	-0.426 (0.285)	-0.796 (0.701)	-0.398 (0.330)	-0.259 (0.670)
Unavailable land, total X ΔlnH		1.478 (0.623)**		4.303 (3.639)	
Unavailable land, water X ΔlnH			2.168 (0.736)***		4.801 (3.089)
Unavailable land, slope X ΔlnH			0.978 (3.533)		-17.039 (10.751)
Unavailable land, swamp X ΔlnH			7.439 (3.845)*		54.638 (19.126)***
Unavailable land, total X AD, total X ΔlnH				-0.137 (0.194)	
Unavailable land, water X AD, water X ΔlnH					-0.161 (0.190)
Unavailable land, slope X AD, slope X ΔlnH					0.854 (0.667)
Unavailable land, swamp X AD, swamp X ΔlnH					-2.554 (1.151)**
Region 1: Western FI	-0.028 (0.024)	-0.006 (0.021)	-0.019 (0.024)	-0.011 (0.020)	-0.005 (0.021)
Region 2: Oulu & Lapland	-0.160 (0.056)***	-0.170 (0.061)***	-0.354 (0.055)***	-0.174 (0.060)***	-0.142 (0.124)
Region 3: Eastern FI	-0.041 (0.016)**	-0.036 (0.010)***	-0.058 (0.014)***	-0.045 (0.018)**	-0.039 (0.024)
Constant	0.258 (0.045)***	0.230 (0.036)***	0.237 (0.037)***	0.227 (0.036)***	0.227 (0.039)***
First stage F-statistic	118.90	88.30	172.82	304.24	452.00

Notes: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. N=32. AD = average distance.

As can be seen from columns 3 and 5 of tables 6 and 7, the effects of the area shares of geographical constraints within a 30-km radius seems to come mostly through water bodies and swamps: the coefficient for the area share of steep slopes is statistically insignificant at the 10% level. Moreover, the same applies to the coefficient of the steep slope interacted share and average distance variable. This might be attributable to the fact that there are no mountain ranges near any of the 42 cities. Moreover, most of Finland is relatively flat, so areas with steep slopes are not very common and might not pose very large constraints on urban development.

Interestingly, however, the area share of swamps has a larger coefficient than that of water bodies. This would mean that a percentage point change in the share of swamp area within city radius would have a larger impact on the elasticity of housing than a percentage point change in water area, which seems counterintuitive. However, a possible explanation behind this could be that the effects of water bodies and swamps are, in fact, nonlinear and have decreasing marginal effects on the elasticity of housing supply. Thus, as there is significantly less swamp area near the cities on average than there is water, swamplands could have a larger marginal impact on housing supply. The same might apply to the effects of the interacted unavailable land share and average distance variables: the more there is a certain source of land unavailability, the less its average distance matters.

Interestingly, in column 5 of both tables, the regression coefficients for variables containing steep slope measures are of unexpected sign: the coefficients for unavailable land share are negative and those of the variable with average distance measure are positive. However, in table 8, these are not statistically significant, which might be more reliable, as the cities with negative population change are left out.

Below in table 8, results for the marginal effects of geographical variables on housing supply elasticity are provided for the six specifications from tables 6 and 7. Model variants are named by their corresponding table and column within that table, so that 6.2 refers to the model specification in column 2 of table 6. As the effects on the elasticity are nonlinear, these marginal coefficients are calculated from the regression results in tables 6 and 7 at average variable values i.e. for a city that would have mean value of water, slope and swamp at mean average distance and would have experienced a mean relative increase in the number of households. For convenience, I also provide the significance levels of corresponding regression coefficients.

TABLE 8. Marginal effects in different specifications

	Model specification, table.column					
	6.2	6.3	6.5	7.2	7.3	7.5
Unavailable land, total	-0.864***			-0.679**		
Unavailable land, water		-0.934***	-1.315		-0.977***	-2.248
Unavailable land, slope		-1.583	10.165**		-0.441	7.979
Unavailable land, swamp		-2.826**	-17.891***		-3.354*	-25.585***
Unavailable land, water X AD, water			0.026			0.075
Unavailable land, slope X AD, slope			-0.584**			-0.400
Unavailable land, swamp X AD, swamp			0.799***			1.196***
Supply elasticity at average values	0.594	0.583	0.592	0.678	0.671	0.684

Notes: Significance levels of original regression coefficients provided for convenience. * significant at 10%; ** significant at 5%; *** significant at 1%. Column headers refer to model specifications, so that e.g. 6.2 refers to the specification in table 6 column 2.

5.4. Discussion and robustness

The validity of the instrumental variables approach hinges critically on the soundness of the assumptions that the applied instruments are both strong and do not correlate with the error term of the second regression equation i.e. they satisfy the exclusion restriction. The F-statistic for the instrument group in the first stage regressions are 139.851 for all cities and 118.902 for cities with non-decreasing populations. As these figures are well above the rule of thumb value of 10 proposed by Staiger and Stock (1997), the instruments may be considered strong. In table 9 I report only these results, as more elaborate model specifications only add interactions with $dlnH_k$ and geographical variables in the second stage regression, and interactions with instruments and geographical variables in the first stage. For other specifications, the first stage results are reported in tables 14 and 15 in Appendix 2. However, even in the most elaborate model specification, the smallest first stage F-value for an instrument group is 97.231. The 2SLS first stage results both for all cities and for the subset of cities which have experienced non-negative population change are reported in table 9 for the simple model $dln\tilde{P}_k^S = \beta_0 + \beta_k^S dlnH_k + \sum_s R_{s,k} + \varepsilon_k$.

TABLE 9. First stage results for 2SLS

	ΔlnH (relative change in the number of households)	
	All cities (N=42)	Cities with > 0 population change (N=32)
Region 1: Western FI	-0.033** (0.014)	-0.031** (0.013)
Region 2: Oulu & Lapland	-0.046 (0.032)	-0.013 (0.031)
Region 3: Eastern FI	-0.010 (0.019)	-0.001 (0.020)
Distance to Helsinki (in 1000 km)	0.360*** (0.072)	0.329*** (0.059)
Distance to Tampere (in 1000 km)	-0.368*** (0.010)	-0.377*** (0.062)
Share of pensioners in 2003	-0.016*** (0,001)	-0.014*** (0,001)
Constant	0.464*** (0.022)	0.444*** (0.021)
Adjusted R^2	0.865	0.888
First stage F-statistic	139.851	118.902
Notes: Significance levels: * significant at 10%; ** significant at 5%; *** significant at 1%.		

The exclusion restriction is satisfied if the instrumental variables are not correlated with the error term of the second stage regression i.e. they are correlated with the dependent variable only through their correlation with the instrumented variable. It is plausible that the instruments used both in this thesis and by Saiz (2010) might be endogenous: the instruments are used as demand shifters to instrument for the relative change in the number of households, yet if households with different preferences concerning housing services react to these demand shifters differently, the instruments may be correlated with relative price changes. For example, if younger people prefer costlier housing (such as smaller apartments, as the prices are per square meter) than older ones, and are more prone to move close to the capital city so that the share of younger cohorts there increases, the demand for costlier housing will go up, plausibly effecting the average price level through more than just the relative increase in the number of households as the supply of smaller apartments increases. As the distance to the capital city instrument does not take these heterogeneities in preferences into account, it may suffer from endogeneity. The same applies to all other instruments used both in this study (distance to the main city of the second largest metropolitan area, the share of pensioners in 2003) and in Saiz (2010) (average hours of sun in January, the industrial composition of the cities in 1973, and change in the number of immigrants between 1970 and 2000 divided by total population in 1970).

As a robustness check, I also conducted fixed effects panel regressions with year effects and city-clustered standard errors for a panel of the same 42 cities that ranges from 2003 to 2015 as means to check the robustness of the cross-section regression results. In order to estimate long term effects, logarithmic levels of the number of households and prices are used instead of differences. Importantly, the results for the most extensive model specification stay fairly similar. These estimation results are provided in table 10 on the next page. It is noteworthy that only the most extensive model specification in column 5 has any statistically significant coefficients (apart from many city and year dummies, not provided in table 10). However, these coefficients resemble those in column 5 in tables 6 and 7, even when one takes into account that the housing market variables are in level form and thus marginal effects are different. This provides additional evidence for the existence and magnitude of geographical effects on housing supply., especially in the case of the most elaborate model.

TABLE 10. Estimation results for the extended model, panel regression with city fixed effects and year effects.

	lnP (supply)				
	(1)	(2)	(3)	(4)	(5)
lnH	0.355 (0.409)	-0.169 (0.521)	0.505 (0.961)	-0.176 (0.530)	0.784 (0.395)**
Unavailable land, total X lnH		1.467 (1.247)		0.093 (4.065)	
Unavailable land, water X lnH			0.899 (1.293)		6.147 (3.336)*
Unavailable land, slope X lnH			-0.453 (5.225)		-17.342 (13.756)
Unavailable land, swamp X lnH			-3.468 (4.195)		90.619 (13.670)***
Unavailable land, total X AD, total X lnH				0.067 (0.189)	
Unavailable land, water X AD, water X lnH					-0.238 (0.160)
Unavailable land, slope X AD, slope X lnH					0.747 (0.727)
Unavailable land, swamp X AD, swamp X lnH					-4.562 (0.674)***
Constant	3.378 (4.534)	2.389 (5.005)	1.130 (0.017)	2.486 (5.029)	-0.161 (2.156)

Notes: City-clustered standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. N=42.

5.5. On elasticities and economic significance

I now move on from the regression estimates of inverse supply elasticities to provide and discuss estimates of the housing supply elasticities themselves. First, I provide some descriptive statistic on housing supply elasticities from different model specifications, after which I present city-level estimations for the same specifications.

In addition, in order to evaluate the economic significance of the estimation results, I calculate estimates for how one standard deviation increases in the geographical variables would affect the housing supply elasticities at average variable values.

Below, in table 11 I provide some descriptive statistics for estimated housing supply elasticities derived from different model specifications. The average elasticity estimates range from 0.467

to 0.728, which is substantially below figures reported by Saiz (2010): 1.75 population weighted average and 2.5 unweighted. However, these estimates are close to those reported by Oikarinen et al. (2015), who arrived at an unweighted average of 0.544 with Finnish data on 15 cities.

Note that the specifications 6.2, 6.3 and 6.5 are analogous to those in 7.2, 7.3 and 7.5, but so that the latter ones are estimated using only the 32 cities which have had non-negative population changes within the sample time-interval. Hence, one can observe, as expected, that the mean of housing supply elasticities is higher in those specifications which include only growing cities.

Moreover, it can be noted that specifications 7.2, 7.3 and 7.5 have smaller coefficients of variation than their counterparts in 6.2, 6.3 and 6.5. This is as expected, as the cities with negative population change have also had relatively inelastic housing supply. Thus, removing them from the sample has lowered the average distance from the distribution mean.

TABLE 11. Descriptive statistics for elasticities from different specifications

Specification of elasticity Table.Column	Mean (standard dev.)	Coefficient of variation
6.2	0.645 (0.394)	0.611
6.3	0.467 (0.675)	1.445
6.5	0.619 (0.449)	0.725
7.2	0.728 (0.381)	0.523
7.3	0.723 (0.395)	0.546
7.5	0.649 (0.283)	0.436

Below in table 12, city-level housing supply elasticity estimates are provided for six different model specifications. In most cases, the cities with negative population change (marked with an asterisk) have below average housing supply elasticities in the first four model specifications. Furthermore some of these cities have very low levels of elasticity, for example, the maximum estimated elasticities for Imatra and Savonlinna are 0.012 and 0.021 respectively. Nevertheless, two cities with negative population change, Kajaani and Raahe, experience large variation in their housing supply elasticity estimates: for both cities, the estimates range from below -1.9 to above 0.7 for Kajaani and Raahe. However, as the coefficients are for the pooled regression (that is, the regression includes both the cities with negative and non-negative population growth), the regression coefficients and thus the calculated housing elasticities can be considered less suitable than those in the regressions without the cities with diminishing population.

Moreover, it can be noted that many smaller cities near larger ones, such as Järvenpää, Kerava and Kirkkonummi near Helsinki, Nokia and Ylöjärvi near Tampere, and Kaarina near Turku, have housing supply elasticities that are more elastic than is the average. Also, they have more elastic supply than their corresponding central cities. This situation might be partially behind the (again, perceived) trend within the studied period towards urban sprawl: it is conceivable that the central cities with their inelastic housing supply would have "pushed" some home buyers to move to places with more elastic housing supply. After all, these places could absorb larger increases in housing demand without as large price increases, thus being able to accommodate large portions of the demand that has been directed towards a certain region. More quantitatively, this can be seen in that relative housing supply growth during the study period and the estimated housing supply elasticities are positively correlated: the lowest correlation coefficient is for the 6.5 specification, for which it is 0.576.

TABLE 12. Estimated elasticities for different model specifications

	Model specification, table.column					
City	6.2	6.3	6.5	7.2	7.3	7.5
Espoo	0.705	0.724	0.727	0.742	0.772	0.726
Hämeenlinna	0.601	0.524	0.532	0.509	0.448	0.424
Helsinki	0.411	0.421	0.412	0.413	0.401	0.395
Hyvinkää	0.716	0.647	0.559	0.577	0.559	0.456
Imatra*	0.011	0.010	0.012			
Järvenpää	1.003	0.971	0.788	0.809	0.861	0.734
Joensuu	0.522	0.510	0.461	0.578	0.540	0.475
Jyväskylä	0.699	0.663	1.302	0.719	0.720	1.050

TABLE 12. (Continued) Estimated elasticities for different model specifications

	Model specification, table.column					
City	6.2	6.3	6.5	7.2	7.3	7.5
Kaarina	0.783	0.895	0.895	0.820	1.023	0.957
Kajaani*	1.105	-2.250	0.634			
Kangasala	0.696	0.727	0.655	0.729	0.747	0.647
Kerava	0.956	0.999	1.118	0.831	0.906	0.962
Kirkkonummi	0.635	0.651	0.688	0.730	0.717	0.725
Kokkola	0.466	0.425	0.404	0.494	0.397	0.358
Kotka*	0.162	0.155	0.171			
Kouvola*	0.120	0.110	0.123			
Kuopio	0.542	0.562	0.597	0.586	0.612	0.601
Lahti	0.519	0.489	0.494	0.435	0.449	0.432
Lappeenranta	0.380	0.350	0.386	0.331	0.306	0.309
Lohja	0.503	0.454	0.473	0.453	0.474	0.480
Mikkeli*	0.346	0.326	0.339			
Nokia	0.803	0.861	0.788	0.815	0.881	0.763
Nurmijärvi	1.186	1.086	0.953	1.010	1.082	0.962
Oulu	1.708	1.902	2.810	2.093	2.439	1.604
Pori	0.320	0.318	0.345	0.296	0.285	0.300
Porvoo	0.555	0.561	0.602	0.513	0.522	0.519
Raahe*	0.734	-1.931	0.504			
Raasepori*	0.262	0.250	0.273			
Rauma*	0.151	0.152	0.154			
Riihimäki	0.723	0.615	0.764	0.595	0.538	0.570
Rovaniemi	1.822	0.832	0.654	1.745	1.109	0.501
Salo	0.346	0.340	0.336	0.308	0.331	0.308
Sastamala*	0.270	0.263	0.262			
Savonlinna*	0.021	0.021	0.021			
Seinäjoki	1.250	0.799	0.697	1.181	0.673	0.517
Tampere	0.658	0.688	0.567	0.673	0.708	0.581
Turku	0.404	0.427	0.422	0.382	0.424	0.407
Tuusula	0.961	0.969	0.989	0.832	0.905	0.950
Vaasa	0.496	0.537	0.549	0.505	0.478	0.475
Vantaa	0.749	0.812	0.792	0.745	0.802	0.777
Vihti	0.902	0.810	0.958	0.852	0.976	0.953
Ylöjärvi	0.906	0.955	0.793	0.982	1.046	0.836

Notes: * city has experienced a decline in population between 2003 and 2015. Column headers refer to model specifications, so that e.g. 7.2 refers to the specification in table 7 column 2.

In order to gain insight into the economic significance of the estimation results, I have calculated changes in supply elasticities when geographic variables increase by one standard deviation from their average values. These figures can be found for six different model specifications in table 13 below. Moreover, where applicable, I provide the statistical significance levels from the regression estimates. Note, however, that for specifications 6.5 and 7.5 this is not sensible due to the interacted nature of the variables in these model versions.

As is to be expected from the regression results, unavailable area shares of water and swamp have large negative effects on housing supply elasticity in specifications 7.3 and 8.3. Although the results stay fairly strong for water bodies in specifications 7.5 and 8.5, these effects are somewhat mitigated by the interacted term with average distance. This does not seem to be the case with swamps, however, where the effect is negligible in 8.5 and reversed in 7.5. Moreover, the effect of a one standard deviation increase in average distance to swamplands seems to have a major positive impact on housing supply elasticity. Again, these discrepancies might point to nonlinearities in the underlying processes.

TABLE 13. Effects of one standard deviation increase in geographic variables at average variable values.

	Model specification, table.column					
	7.2	7.3	7.5	8.2	8.3	8.5
Unavailable land, total	-0.088***			-0.062**		
Unavailable land, water		-0.107***	-0.092		-0.101***	-0.075
Unavailable land, slope		-0.067	-0.056		-0.070	0.005
Unavailable land, swamp		-0.127**	-0.060		-0.151**	-0.031
Average distance, water			0.009			0.026
Average distance, slope			-0.048			-0.032
Average distance, swamp			0.079			0.126
Supply elasticity at average values	0.594	0.583	0.592	0.678	0.671	0.684

Notes: Significance levels of original regression coefficients provided for columns other than 6.5 and 7.5. * significant at 10%; ** significant at 5%; *** significant at 1%. Column headers refer to model specifications, so that e.g. 7.2 refers to the specification in table 7 column 2.

6. CONCLUSION

This thesis began by introducing three objectives for itself: generating data similar to that of Saiz (2010), replicating his analysis as closely as possible in the Finnish context, and further developing the applied framework and analysis to include more heterogeneity in the geographic variables.

Using satellite generated data, I created geographic variables on waterbodies, swamps and steep slopes for 42 Finnish cities, including both area shares within a 30km radius from the city centre and average distance measures to those areas.

With this newly generated data set, I replicated the analysis of Saiz (2010). The replication results (in table 6) are fairly similar to those in the aforementioned study, even though the number of cities, their size and the studied time interval differ significantly between the studies. Hence, this might provide evidence that geographic features have an effect on city-level housing supply elasticity. As the endogeneity of the applied instrumental variables is plausible, I also conducted a fixed effects panel regression with year effects as a robustness check. These results provided evidence in favour of the most extensive model.

Moreover, I further developed the analytical framework to include variation both in the sources of geographical constraints and in the position of these features relative to city centres. Empirically estimating models with these new additions provided more information on the sources of the effects on housing supply elasticity: At least in the Finnish context, it seems that steep slopes do not pose strict constraints on housing development, whereas waterbodies and swamps have more substantial effects both through their prevalence and relative position. Moreover, the findings indicate that these effects are not only statistically significant, but have a substantial impact on local housing supply elasticities in Finland.

REFERENCES

Alonso, W. *Location and Land Use* (Cambridge, MA: Harvard University Press, 1964).

Ball, Michael, Geoffrey Meen and Christian Nygaard, "Housing Supply Price Elasticities Revisited: Evidence from International, National, Local and Company Data," *Journal of Housing Economics*, 19 (2010), 255-268.

Dong, Yilin. "A Note on Geographical Constraints and Housing Markets in China," *Journal of Housing Economics*, 33 (2016), 15-21.

Follain, James, "The Price Elasticity of Long-run supply of new housing construction," *Land Economics*, 55 (1979), 190-191.

Goodman, Allen, "Central Cities and Housing Supply: Growth and Decline in US Cities," *Journal of Housing Economics*, 14 (2005), 315-335.

Green, Richard, Stephen Malpezzi and Stephen Mayo, "Metropolitan-Specific Estimates of the Price Elasticity of Supply of Housing, and Their Sources," *American Economic Review*, 95 (2005), 334-339.

Gyourko, Joseph, Albert Saiz and Anita A. Summers, "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 45 (2008), 693-729.

Gyourko, Joseph, "Housing Supply," *Annual Review of Economics*, 1 (2009), 295-318.

Horn, Berthold, "Hill Shading and the Reflectance Map," *Proceedings of the IEEE*, 69 (1981), 14-47.

Malpezzi, Stephen, "Housing Prices, Externalities, and Regulation in U.S. Metropolitan Areas," *Journal of Housing Research*, 7 (1996), 209-242.

Malpezzi, Stephen, Gregory H. Chun and Richard K. Green, "New Place-to-place Housing Price Indexes for U.S. Metropolitan Areas, and Their Determinantas," *Real Estate Economics*, 26 (1998), 235-274.

Malpezzi, Stephen and Duncan Maclennan, "The Long-Run Price Elasticity of Supply of New Residential Construction in the United States and the United Kingdom," *Journal of Housing Economics*, 10 (2001), 278-306.

Meen, Geoffrey and Christian Nygaard, "Local Housing Supply and the Impact of History and Geography," *Urban Studies*, 48 (2011), 3107-3124.

Mills, Edwin, "An Aggregate Model of Resource Allocation in a Metropolitan Area," *American Economic Review*, 57 (1967), 197-210.

Muth, Richard, *Cities and Housing* (Chicago: University of Chicago Press, 1969).

Muth, Richard, "The Demand for Non-Farm Housing", in *The Demand for Durable Goods*, A. Harberger, ed. (Chicago: University of Chicago Press, 1960).

Oikarinen, Elias, Risto Peltola and Eero Valtonen, "Regional Variation in the Elasticity of Supply of Housing, and Its Determinants: The Case of a Small Sparsely Populated Country," *Regional Science and Urban Economics*, 50 (2015), 18-30.

Rose, Louis, "Urban Land Supply: Natural and Contrived Restrictions," *Journal of Urban Economics*, 25 (1989), 325-345.

Rose, Louis, "Topographical Constraints and Urban Land Supply Indexes," *Journal of Urban Economics*, 26 (1989), 335-347.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, 125 (2010), 1253-1296.

Staiger, Douglas, and James H. Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65 (1997), 557-586.

Wang, Songtao, Su Han Chan and Bohua Xu, ”The Estimation and Determinants of the Price Elasticity of Housing Supply: Evidence from China,” *Journal of Real Estate Research*, 34 (2012), 311-344.

APPENDIX 1 – MATHEMATICAL DERIVATIONS

Derivation 1

As $\beta_k^S \equiv \frac{dln\tilde{P}_k^S}{dlnH_k} = \frac{\partial \tilde{P}_k^S}{\partial H_k}\frac{H_k}{\tilde{P}_k^S}$ and from equation 3 we have that $\tilde{P}_k^S = CC + \frac{1}{3i}t\sqrt{\frac{\gamma H_k}{\pi \Lambda_k}}$, we can express the inverse supply elasticity as $\beta_k^S = \frac{\frac{1}{6i}t\sqrt{\frac{\gamma H_k}{\pi}}}{CC\Lambda_k^{\frac{1}{2}} + \frac{1}{3i}t\sqrt{\frac{\gamma H_k}{\pi}}}$. Partially differentiating this with respect to Λ_k gives us $\frac{\partial \beta_k^S}{\partial \Lambda_k} = \frac{-\frac{1}{12i}t\sqrt{\frac{\gamma H_k}{\pi}}CC\Lambda_k^{\frac{1}{2}}}{\left(CC\Lambda_k^{\frac{1}{2}} + \frac{1}{3i}t\sqrt{\frac{\gamma H_k}{\pi}}\right)^2} < 0$.

Derivation 2

The first order Taylor approximation of β_k^S near the point $\Lambda_k = a$ can be expressed as $\beta_k^S \approx \beta_k^S(a) + \beta_k^{S'}(a) \cdot (\Lambda_k - a) = \beta_k^S(a) + \beta_k^{S'}(a) \cdot (\Lambda_k - a + 1 - 1) = \beta_k^S(a) + (1-a)\beta_k^{S'}(a) + (\Lambda_k - 1)\beta_k^{S'}(a)$. Defining $\tilde{\beta}^S \equiv \beta_k^S(a) + (1-a)\beta_k^{S'}(a)$ and $\beta^{LAND} \equiv -\beta_k^{S'}(a)$ we have$\beta_k^S \approx \tilde{\beta}^S + (1-\Lambda_k)\beta^{LAND}$, where $\tilde{\beta}^S$ is the intercept of the relation and $(1-\Lambda_k)\beta^{LAND}$ gives the approximate contribution of land availability to the inverse supply elasticity β_k^S.

Derivation 3

As Saiz notes (2010, p. 1287), a share of $2\pi\Lambda(d)d/\,\gamma H_k$ households live at distance d from the city centre. As $\Lambda(d)$ variates with distance, this implies that the average rent level within the city is $\tilde{r}_k = \left(\frac{1}{\gamma H_k}\right)\int_0^{\Phi_k} 2\pi\Lambda(x)x \cdot r(x) \cdot dx$. Substituting equation 1, we have

$$\tilde{r}_k = \left(\frac{1}{\gamma H_k}\right)\int_0^{\Phi_k} (2\pi\Lambda(x)x) \cdot (r_0 - tx) \cdot dx\,. \qquad (A1)$$

As $\Lambda(x)$ is a constant step function with n intervals, step points at $x_j, j = 0, 1, \dots, n$ and a value of $\Lambda_{j,k}$ at interval j, equation A1 can be expressed as

$$\tilde{r}_k = \left(\frac{\pi}{\gamma H_k}\right) \sum_j \Lambda_{j,k} \int_{x_{j-1}}^{x_j} (2xr_0 - 2x^3 t) \cdot dx\,. \qquad (A2)$$

Integrating and simplifying, equation A2 becomes

$$\tilde{r}_k = \left(\frac{\pi}{\gamma H_k}\right) \sum_j \Lambda_{j,k} \left[\left(x_j{}^2 - x_{j-1}{}^2\right) r_0 - \frac{2}{3} t \left(x_j{}^3 - x_{j-1}{}^3\right) \right]. \qquad (A3)$$

Noting that $\left(x_j - x_{j-1}\right) = l$ which is the interval length, $x_j = jl$ and that $l = \frac{\Phi_k}{n} = \sqrt{\frac{\gamma H_k}{\Lambda_k \pi n^2}}$, we get

$$\tilde{r}_k = \frac{1}{\Lambda_k n^2} \left(\sum_j (2j-1) \Lambda_{j,k}\, r_0 - t \frac{2}{3} l \sum_j (3j^2 - 3j + 1)\, \Lambda_{j,k} \right). \qquad (A4)$$

Now, as $\Lambda(x)$ is a step function, the total share of developable land $\Lambda_k = \frac{developable\ land}{total\ land} = \frac{\int_0^{\Phi_k} 2\pi x \Lambda(x) dx}{\pi {\Phi_k}^2} = \frac{\sum_j \Lambda_{j,k}(x_j^2 - x_{j-1}^2)}{{\Phi_k}^2} = \frac{l^2 \sum_j \Lambda_{j,k}(2j-1)}{{\Phi_k}^2} = \frac{\sum_j \Lambda_{j,k}(2j-1)}{n^2}$. Substituting this result into equation A4 and simplifying, we have

$$\tilde{r}_k = \left(r_0 - t \frac{2}{3} l \frac{\sum_j (3j^2 - 3j + 1) \Lambda_{j,k}}{\sum_j (2j-1)\, \Lambda_{j,k}}\right). \qquad (A5)$$

Now, the average distance to developable land $\bar{d}_k = \sum_{j=1}^{n} (\Lambda_{j,k} A_j / \Lambda_k A)\, d_j$, where the average distance to zone j is $d_j = \frac{\int_{x_{j-1}}^{x_j} x \cdot 2\pi x \cdot dx}{\int_{x_{j-1}}^{x_j} 2\pi x \cdot dx}$, the total area of zone j is $A_j = \int_{x_{j-1}}^{x_j} 2\pi x \cdot dx$ and the total area of developable land is $\Lambda_k A = \sum_j \Lambda_{j,k} \int_{x_{j-1}}^{x_j} 2\pi x \cdot dx$. Taken together, these mean that $\bar{d}_k$ can be expressed as $\bar{d}_k = \frac{\sum_j (\frac{1}{3} x_j^3 - \frac{1}{3} x_{j-1}^3)}{\sum_j (\frac{1}{2} x_j^2 - \frac{1}{2} x_{j-1}^2)}$. Noting that $\left(x_j - x_{j-1}\right) = l$, this becomes $\bar{d}_k = \frac{2}{3} l \frac{\sum_j (3j^2 - 3j + 1) \Lambda_{j,k}}{\sum_j (2j-1) \Lambda_{j,k}}$. Substituting this into equation A5, we get

$$\tilde{r}_k = r_0 - t\bar{d}_k. \qquad (A6)$$

Notably, this corresponds to the rent level at average distance $\bar{d}$ from the city centre. As with Saiz (2010, p. 1263) $r_0 = iCC + t\sqrt{\frac{\gamma H_k}{\Lambda_k \pi}}$ and $P(d) = \frac{r(d)}{i}$. Combining these with equation A6, we get the final housing supply equation as

$$\tilde{P}_k^S = CC + \frac{1}{i} t\left(\sqrt{\frac{\gamma H_k}{\Lambda_k \pi}} - \bar{d}_k\right).$$

Derivation 4

The inverse of housing supply elasticity is $\beta_k^S \equiv \frac{dln\tilde{P}_k^S}{dlnH_k} = \frac{\partial \tilde{P}_k^S}{\partial H_k}\frac{H_k}{\tilde{P}_k^S}$. Defining $z_k \equiv \frac{\bar{d}_k}{\Phi_k}$ and noting that this has a value between 0 and 1, equation 7 becomes $\tilde{P}_k^S = CC + \frac{1-z_k}{i} t\sqrt{\frac{\gamma H_k}{\Lambda_k \pi}}$. Thus, the inverse elasticity of supply can be expressed as $\beta_k^S = \frac{\frac{1-z_k}{2i} t\sqrt{\frac{\gamma H_k}{\pi}}}{CC\Lambda_k^{\frac{1}{2}} + \frac{1-z_k}{i} t\sqrt{\frac{\gamma H_k}{\pi}}}$. Therefore, $\frac{\partial \beta_k^S}{\partial \Lambda_k} = \frac{-\frac{1-z_k}{4i} t\sqrt{\frac{\gamma H_k}{\pi}} CC\Lambda_k^{\frac{1}{2}}}{\left(CC\Lambda_k^{\frac{1}{2}} + \frac{1-z_k}{i} t\sqrt{\frac{\gamma H_k}{\pi}}\right)^2} < 0.$

Derivation 5

First, note that $\frac{\partial \beta_k^S}{\partial \overline{ud}_k} = \frac{\partial \beta_k^S}{\partial z_k} \cdot \frac{\partial z_k}{\partial \bar{d}_k} \cdot \frac{\partial \bar{d}_k}{\partial \overline{ud}_k}$. Now, it is straight forward to derive that $\frac{\partial \beta_k^S}{\partial z_k} = -\frac{t}{2i}\sqrt{\frac{\gamma H_k}{\Lambda_k \pi}} \frac{CC}{\tilde{P}_k^{S^2}} < 0$ and $\frac{\partial z_k}{\partial \bar{d}_k} = \frac{1}{\Phi_k} > 0$, so it is sufficient to show $\frac{\partial \bar{d}_k}{\partial \overline{ud}_k}$ to be negative for $\frac{\partial \beta_k^S}{\partial \overline{ud}_k}$ to be positive. Note that the average distance to undevelopable land is $\overline{ud}_k = \sum_j \frac{(1-\Lambda_{j,k})A_j}{(1-\Lambda_k)A}\ d_j$. Hence, the average distance to developable land is $\bar{d}_k = \sum_j (\Lambda_{j,k} A_j / \Lambda_k A)\ d_j = \frac{1}{\Lambda_k A}\sum_j (\Lambda_{j,k} A_j d_j - A_j d_j + A_j d_j) = \frac{1}{\Lambda_k A}(\sum_j A_j d_j - \sum_j (1 - \Lambda_{j,k}) A_j d_j) = \sum_j \frac{A_j d_j}{\Lambda_k A} - \sum_j \frac{(1-\Lambda_{j,k}) A_j d_j}{\Lambda_k A} = \sum_j \frac{A_j d_j}{\Lambda_k A} - \frac{(1-\Lambda_k)}{\Lambda_k}\sum_j \frac{(1-\Lambda_{j,k}) A_j d_j}{(1-\Lambda_k)A} = \sum_j \frac{A_j d_j}{\Lambda_k A} - \frac{(1-\Lambda_k)}{\Lambda_k}\overline{ud}_k$.Thus, $\frac{\partial \bar{d}_k}{\partial \overline{ud}_k} = -\frac{(1-\Lambda_k)}{\Lambda_k} < 0$, when $(1-\Lambda_k) > 0$, and becomes 0 when $(1-\Lambda_k) = 0$.

APPENDIX 2 – FIRST STAGE RESULTS

TABLE 14. First stage results for 2SLS-regressions with all cities.

	ΔlnH, 2003-2015, model specification: table.column					
	6.2	6.5	6.9	7.3	7.4	7.5
A = Distance to Helsinki (in 1000 km)	0.360***	0.829***	0.956***	0.832***	0.770***	0.715*
	(0.065)	(0.113)	(0.110)	(0.125)	(0.258)	(0.376)
B = Distance to Tampere (in 1000 km)	-0.368***	-0.771***	-1.094***	-0.713**	-0.339	-1.663**
	(0.067)	(0.274)	(0.299)	(0.305)	(0.396)	(0.733)
C = Share of pensioners in 2003	-0.016***	-0.014***	-0.012***	-0.013***	-0.017***	-0.013**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.005))
A X totalshare		-1.211***	-3.601***	-0.886		
		(0.257)	(0.869)	(1.929)		
B X totalshare		1.176	1.706	-2.707		
		(0.725)	(1.589)	(2.860)		
C X totalshare		-0.007	0.014	0.025		
		(0.006)	(0.010)	(0.023)		
A X totalshare X lnpop03			0.170***			
			(0.058)			
B X totalshare X lnpop03			0.042			
			(0.096)			
C X totalshare X lnpop03			-0.002			
			(0.001)***			
A X totalshare X adtotal				-0.018		
				(0.095)		
B X totalshare X adtotal				0.189		
				(0.134)		
C X totalshare X adtotal				-0.002		
				(0.001)		
A X watershare					-1.197***	-0.284
					(0.384)	(2.662)
B X watershare					0.861	13.701**
					(0.796)	(5.022)
C X watershare					-0.004	-0.131***
					(0.006)	(0.039)
A X slopeshare					-0.974	-22.051***
					(1.291)	(7.441)
B X slopeshare					-2.320	0.584
					(2.025)	(21.175)
C X slopeshare					0.020	0.295
					(0.012)	(0.173)
A X swampshare					-1.414	35.171**
					(1.915)	(12.664)
B X swampshare					0.515	-109.175**
					(2.114)	(38.127)
C X swampshare					0.007	0.511**
					(0.014)	(0.201)
A X watershare X adwater						-0,035
						(0.140)
B X watershare X adwater						-0.522**
						(0.226)
C X watershare X adwater						0.006***
						(0.002)
A X slopeshare X adslope						1.161**
						(0.442)
B X slopeshare X adslope						-0.004
						(1.135)

TABLE 14. (Continued) First stage results for 2SLS-regressions with all cities.

	ΔlnH, 2003-2015, model specification: table.column					
	6.2	6.5	6.9	7.3	7.4	7.5
C X slopeshare X adslope						-0.015 (0.009)
A X swampshare X adswamp						-1.809** (0.667)
B X swampshare X adswamp						5.691** (1.973)
C X swampshare X adswamp						-0,026** (0.011)
F-statistic	139.85	59.08	80.66	54.65	39.82	151.23

Notes: Totalshare is the total share of undevelopable land within a city circle. Watershare, swampshare and slopeshare are the equivalent shares of water, swamp and steep slopes respectively. Adtotal is the average distance to the total undevelopable land area within a city circle. Adwater, adswamp and adslope are the equivalent distances of water, swamp and steep slopes respectively. * significant at 10%; ** significant at 5%; *** significant at 1%. N=42.

TABLE 15. First stage results for 2SLS-regressions, cities with non-negative population change.

	ΔlnH, 2003-2015, model specification: table.column				
	8.1	8.2	8.3	8.4	8.5
A = distance to Helsinki (in 1000 km)	0.329*** (0.067)	0.744*** (0.110)	0.779*** (0.110)	0.406* (0.217)	-0.588 (1.008)
B = distance to Tampere (in 1000 km)	-0.377*** (0.068)	-1.082*** (0.256)	-0.960*** (0.304)	-0.684 (0.457)	-4.227** (1.444)
C = share of pensioners in 2003	-0.014*** (0.001)	-0.010*** (0.002)	-0.010*** (0.003)	-0.012** (0.005)	0.011 (0.011)
A X total30		-1.126*** (0.247)	-0.530 (1.935)		
B X total30		2.103*** (0.693)	-2.758 (2.554)		
C X total30		-0.016*** (0.005)	0.016 (0.020)		
A X total30 X adtotal			-0.035 (0.096)		
B X total30 X adtotal			0.228* (0.124)		
C X total30 X adtotal			-0.001* (0.001)		
A X water30				-0.564 (0.351)	15.758** (6.640)
B X water30				1.431 (1.011)	26.540 (15.113)
C X water30				-0.011 (0.009)	-0.256 (0.139)
A X slope30				-0.009 (1.141)	-66.569* (32.914)
B X slope30				1.108 (2.686)	-60.692** (17.891)
C X slope30				-0.007 (0.019)	0.843** (0.243)
A X swamp30				1.092 (1.792)	6.601 (22.272)
B X swamp30				-1.288 (1.710)	-221.022* (103.319)

TABLE 15. (Continued) First stage results for 2SLS-regressions, cities with non-negative population change.

	ΔlnH, 2003-2015, model specification: table.column				
	8.1	8.2	8.3	8.4	8.5
C X swamp30				0.003	1.268*
				(0.022)	(0.606)
A X water30 X adwater					-0.681**
					(0.278)
B X water30 X adwater					-0.934
					(0.611)
C X water30 X adwater					0.010
					(0.006)
A X slope30 X adslope					3.653*
					(1.893)
B X slope30 X adslope					3.898**
					(1.104)
C X slope30 X adslope					-0.049**
					(0.014)
A X swamp30 X adswamp					0.241
					(1.427)
B X swamp30 X adswamp					11.259*
					(5.233))
C X swamp30 X adswamp					-0.068*
					(0.033)
F-statistic	118.90	88.30	172.82	304.24	452.00

Notes: Totalshare is the total share of undevelopable land within a city circle. Watershare, swampshare and slopeshare are the equivalent shares of water, swamp and steep slopes respectively. Adtotal is the average distance to the total undevelopable land area within a city circle. Adwater, adswamp and adslope are the equivalent distances of water, swamp and steep slopes respectively. * significant at 10%; ** significant at 5%; *** significant at 1%. N=32.