UNIVERSITY OF GOTHENBURG
School of Business, Economics and Law
Graduate School

MASTER'S THESIS

Determinants of House Prices in Sweden

Authors: **Radan Papousek, Martin Sodling**
Supervisor: **Aico van Vuuren**
Academic year: **2017/2018**

Bibliographic note

Papoušek Radan, Södling Martin. *Determinants of House Prices in Sweden.* Gothenburg 2018. 38 p. Master's thesis (MSc.) University of Gothenburg, School of Business, Economics and Law. Supervisor: Aico van Vuuren

Character count: 57 252

Abstract

The thesis analyses the main determinants of the Swedish house prices. We use panel data of 290 Swedish municipalities across 2003 - 2016 to estimate Spatial Durbin Model, which allows us to capture spatial dependencies in the data, thus obtaining unbiased estimates of the main drivers. Further, we run the cross-sectional regressions for every year to discover the dynamics in the determinants. This proves to be especially valuable having the period of the financial crisis of 2008 in our data set. We obtain a comprehensive picture of the spatial dependencies and spillovers from one municipality to the others. Proper analysis of the dynamics in the housing sector is vital for assessing policy implications made by important players such as the central banks. Since the price development has not been even among the individual municipalities, analysis on the regional level may give us better insight into the sector than country-level studies. We estimate the total effects as well as the direct and indirect effects of the spatially autocorrelated variables. We find that the main determinants for the Swedish house prices are construction costs and real income. We find that developers are not fully able to transfer the cost onto the final buyers. Real income is a stable driver of the house prices, dropping in importance only in 2009; we attribute it to the higher uncertainty in the markets, thus people withholding their house purchases. Comparing to other countries, availability of credit for households plays more significant role; possibly due to the fact that Swedish households rely more on credit and have more often floating-rate mortgages. Housing supply and unemployment have effect on prices only when there is a same shift also in the neighbouring municipalities. There is no visible difference for the spillover patterns of different municipalities. The spatial model proves to be a better suit for the estimation, since it outperforms the non-spatial model in out-of-sample forecast.

Keywords:	Spatial Durbin Model, spatial autocorrelation, Sweden, house prices, spillover, municipalities
Authors' e-mails:	papousek.radan@gmail.com martin.soedling@gmail.com
Supervisor's e-mail:	aico.van.vuuren@economics.gu.se

Acronyms

AC	Autocorrelation
BIS	Bank for International Settlements
DGP	Data generating process
FE	Fixed effects model
GMM	Generalised methods of moments
HAC	Heteroskedasticity & autocorrelation consistent
IPS	Im, Pesaran and Shin unit root test
LLC	Levin-Lin-Chu unit root test
OLS	Ordinary least squares
PCADF	Pesaran's cross-sectionally augmented Dickey–Fuller test
RE	Random effects model
RMSE	Root mean square error
SAR	Spatial autocorrelation model
SCB	Statistics Sweden
SEM	Spatial error model
VAR	Vector autoregression
VECM	Vector error correction
VIF	Variance inflation factors

Contents

Chapter 1

Introduction

There has been a vast amount of research done regarding the housing markets. House prices are an important aspect of the macroeconomic environment and the housing market is strongly interlinked with the bank sector and credit activity, thus there is a direct connection with the real economy and it is watched closely by policy makers. Proper analysis of the dynamics in the housing sector is vital for assessing policy implications made by important players such as the central banks. As seen in the past, high volatility in house prices, bubbles and bursts have serious impact on domestic economies, possibly even affecting the regional or global markets. Concerns are now arising with the low-interest-rate environment. Nowadays, central banks have been reluctant to increase the interest rates being aware of the situation in the housing market e.g., Sweden's Riksbank has had a negative repo rate since 2015 (Riksbank 2018). Households are sensitive to interest rates movements, since the mortgage payments are the main burden for the majority of them (SCB 2018). In economies such as the Swedish one, where fix rate mortgages are in minority and households debt is on record levels (BIS 2018, Finansinspektionen 2017), any change in monetary policy can have an unforeseen impact on the whole economy. Sweden's Riksbank has issued many statements warning of the increased significance of the household indebtedness for the financial and economic stability. We have seen a steep rise in the house prices raising questions about its sustainability. As suggested by Julian Livingston-Booth (2018*b*), imbalances have been emerging in the Swedish house market with residential property prices shooting up by more than 50 % compared to its pre-crisis peak.

Nonetheless, the price development has not been even among the individual municipalities. For instance, rents in Stockholm grew more than twice as rapidly as in Gothenburg or Malmö and are now more than double in absolute terms (Julian Livingston-Booth 2018*b*). We now see some stabilization as prices starts to decline. However, we can already see some implications of this correction e.g., downgrading of some equities (Julian Livingston-Booth 2018*a*) and projected GDP growth downgraded (Las 2018).

The majority of research has been conducted either on country level or on individual-property level. As the price development differs significantly between municipalities, neither of this accounts for the dynamics on a municipality level. Our aim is to improve the analysis by using up-to-date techniques and investigating the dependencies on more granular data; we take into account also the spatial dimension in modelling house prices. LeSage & Pace (2009) make the case that if one does not take spatial dependencies into account, where there are spatial dependencies, you risk getting biased and inefficient estimates[1].

We contribute to the existing literature by capturing spatial dependency and quantifying it. Such an approach gives additional insight into the real estate market and provides policy makers with more in-depth knowledge about the dynamics within Sweden. We estimate the *total effects* as well as the *direct* and *indirect effects* of the spatially autocorrelated variables. The indirect effect captures the impact of a change in a variable to, and from surrounding municipalities. There has been research suggesting that housing prices exhibit spatial clustering, hence applying a spatial model and estimating total effects may prove to be a more accurate depiction of the house prices dynamics and of the important drivers. We also provide the development of the variables of interest over time, giving additional perspective of the dynamics of the house prices.

The rest of the thesis is organized as follows. In chapter 2, we briefly summarize existing research in the field. Chapter 3 describes the data and employed methodology. In chapter 4, we discuss further our data, construction of a weight matrix, and final model specification that we choose for estimation. We report and discuss the results in chapter 5, and finally chapter 6 concludes the thesis.

[1] Shown in the Appendix A

Chapter 2

Literature Review

In this section we give a brief overview of spatial econometric modelling, then present research related to ours where the focus is on the variables we aim to estimate and the model we use.

Common techniques employed in the field is to utilize VAR models with different specifications. Such models are however not able to catch spatial dependencies. This is understandable for country-level analyses, where it may be safe to assume no significant spatial dependencies among countries; e.g., Tsatsaronis & Zhu (2004) document significant differences in house price development in a sample of developed countries. The authors suggest that house markets are "intrinsically local in character" (Tsatsaronis & Zhu 2004, p.67) e.g., due to factors such as income, demographic dynamics, availability and cost of land, VAT levels, inheritance taxes, financing availability, etc. This also leads (Tsatsaronis & Zhu 2004, p.68) to a conclusion that country-level analysis may not be optimal due to the local character of house markets. Moreover, many factors e.g., demographics, unemployment, or income may be strongly interlinked among the individual regions within a country. Thus when analysing the house market in one country, ignoring the spatial dependencies on regional level may lead to biased and inconsistent estimates if not accounted for (Anselin 1988, Dubin 1998), thus we choose to employ spatial econometric techniques in our analysis.

There has been growing number of papers employing spatial models e.g., Kim et al. (1998), Liao & Wang (2012), Wilhelmsson (2008). Nonetheless, the models are mainly run on cross-sectional data due to computational difficulties arising from panel data specifications and lack of data. Spatio-temporal models are gaining importance just recently thanks to the IT development. For instance, Holly et al. (2010) investigate changes in real house prices in the USA on panel data; Zheng & Hui (2016) examine liquidity impacts on house market performance in Hong Kong using a panel. We further expand the research utilizing the panel data on 290 Swedish municipalities across 2003-2016. We choose to conduct our analysis on panel data rather than cross-sectional one, since it allows us to control for unobservable heterogeneity - limiting omitted variable bias (Hsiao 2005). Also, in general panel data contains more degrees of freedom and there is higher variability which leads to higher efficiency and better inference of the parameters (Elhorst 2014, Hsiao 2005). Moreover, it can cope better with non-stationary data (Hsiao 2005, p.7). Our methodology is mainly based on Anselin et al. (2008), LeSage & Pace (2009), Zheng & Hui (2016). Nonetheless due to a lack of research - either ignoring spatial dependencies or using sub-optimal[1] model specification such as SAR - further comparison with results published in related literature has limited value.

2.1 Spatial Durbin Model and Total Effects

Elhorst (2014), Lesage (2008) contain fairly detailed descriptions of spatial models for both cross section and for panel data sets, and instructions on how to implement them. Figure 2.1 summarizes the spatial models. At the top, there is the General Nesting Spatial Model that allows for spatial interaction in all variables: the dependent, the error term and the independent variables. This model is deemed no good, since the parameters of this model are not identified (Elhorst 2014). Moving downwards, the amount of variables that are allowed spatial interaction is reduced until we reach the one at the bottom which is an ordinary OLS. Most researchers use either the Spatial Lag Model SAR, or the Spatial Error Model SEM (Elhorst 2014). If an OLS or an SEM estimation is employed when there is a data generating process (DGP) that is best modelled with the SAR specification, then this will produce a biased estimate. While using the Spatial Durbin Model (SDM) specification will in fact produce unbiased estimates for all DGP's that are represented by the SEM, SAR, SAC and of course SDM (LeSage & Pace 2009). Elhorst (2014) also suggest using the SDM specification instead of the SAR, as he finds SAR to be too restrictive.

When estimating a spatial model, the focus is on the direct and the indirect effects (LeSage & Pace 2009). As an example, say you estimate a SDM model and have an insignificant and negative estimate of the coefficient for WX, but a positive and significant indirect effect, where W is the weight matrix and X is a matrix of independent variables. This is not so strange if you interpret the indirect effect as the partial derivative of X. We use the example produced in Elhorst (2014), where we have an SDM model as in equation (2.1), where R represents intercept and error terms.

$$\mathbf{Y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta}) + \mathbf{R} \tag{2.1}$$

[1] Due to being to restrictive.

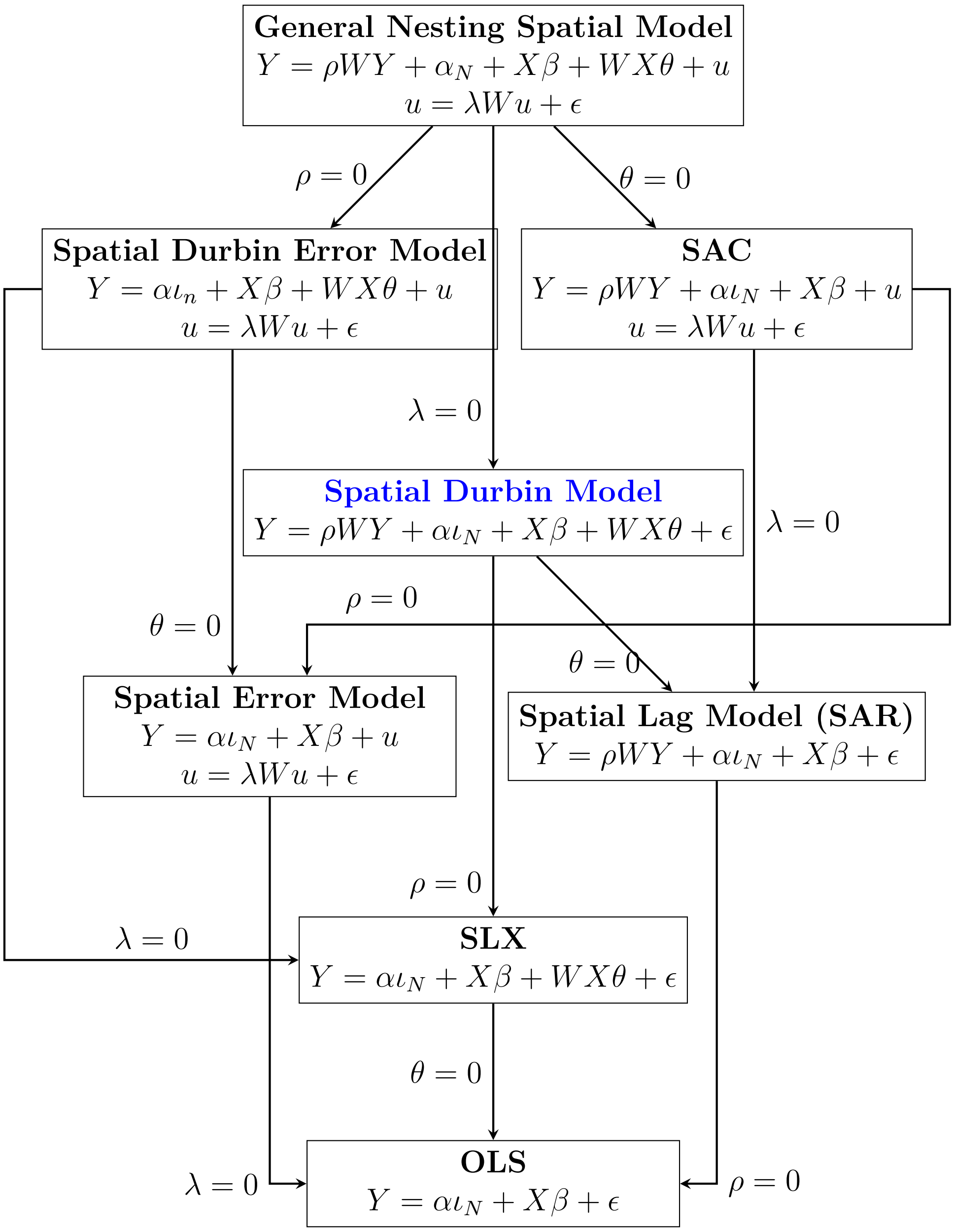


Figure 2.1: The relationships between different spatial dependence models for cross-section data. (Elhorst 2014, p.13)

If we take the partial derivative of $E[\mathbf{Y}]$ with respect to the kth independent variable, for entity 1 up to N, we get the direct effect and the indirect effect. Where the direct effect (the diagonal elements) is due to a change of certain explanatory variable for a certain entity, which can contain feedback effects. This can be shown in equation 2.2, when $|\rho| < 1$ and W is row-normalised (rows sum to one). q represents the order of the weight matrix. When $q = 0$, the weight matrix becomes an identity matrix: $W^0 = I$; when $q = 1$, W contains zeros on the diagonal (assuming that the initial specification does not allow for an entity to be a neighbour of oneself). For higher powers, $q > 1$, the diagonal is non-zero, allowing for feedback effects.

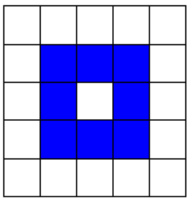
$$(\mathbf{I} - \rho\mathbf{W})^{-1} = I + \rho W + \rho^2 W^2 + \rho^3 W^3 ... \tag{2.2}$$

The indirect effects (off-diagonal elements) is the change this causes in the other units of the same explanatory variable, i.e. you can still have an insignificant θ and indirect effects. If, however, both $\rho = 0$ and $\theta = 0$ then there are no indirect effects or feedback effects. This can be seen in equation (2.3) and (2.4). The analysis of the effects usually concerns the averages of the direct, indirect and total effects (total being the sum of the direct and indirect). Where the average total effects can be interpreted as the average effect from an entity to other entities, but also from other entities to one entity. These two are equivalent in a numerical sense[2], but as mentioned, are interpreted differently (LeSage & Pace 2009).

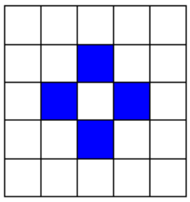
$$\begin{bmatrix} \frac{\partial E(\mathbf{Y})}{\partial x_{1k}} & \cdot & \frac{\partial E(\mathbf{Y})}{\partial x_{Nk}} \end{bmatrix} = \begin{bmatrix} \frac{\partial E(y_1)}{\partial x_{1k}} & \cdot & \frac{\partial E(y_1)}{\partial x_{Nk}} \\ \cdot & \cdot & \cdot \\ \frac{\partial E(y_N)}{\partial x_{1k}} & \cdot & \frac{\partial E(y_N)}{\partial x_{Nk}} \end{bmatrix} \tag{2.3}$$

$$= (\mathbf{I} - \rho\mathbf{W})^{-1} \begin{bmatrix} \beta_k & w_{12}\theta_k & \cdot & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdot & w_{2N}\theta_k \\ \cdot & \cdot & \cdot & \cdot \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdot & \beta_k \end{bmatrix} \tag{2.4}$$

The weight matrix specification plays a crucial role in the estimation algorithm. One way is to design it as either a Rook or a Queen contiguity matrix.



(a) Queen



(b) Rook

Figure 2.2: Contiguity matrices

There are distance based matrices, where you take the inverse of the distance, giving higher values to closer neighbours, or set some threshold: e.g., every one within a certain radius is a neighbour (making sure the radius is within a minimum to make sure there are no observations with zero links). Making use of k-nearest

[2] A brief discussion about the equivalence is provided in Appendix A.

neighbour is also possible. Where you give k-amount of links to each observation in a binary way: 1 if they are one of the k neighbours, zero otherwise. In most applications row standardization is employed. By row standardizing, for any given municipality, i: $\sum_{j=1}^{n} w_{ij} = 1$, i.e. every element is divided by its row sum: $w_{ij} = \sum_{j=1}^{n} w_{ij}$ (LeSage & Pace 2009, Michael D. Ward 2008, Roger S. Bivand 2008). This yields an asymmetric matrix (which has computational consequences) and it also yields an average estimate of all of the links; all the links have the same value. There are different orders of weight matrices. Order one weight matrix has zero and non-zero values on the off-diagonal and zeros on the diagonal. Having zeros on the diagonal means that you are not a neighbour to yourself. There are however specifications when you have higher order weight matrices, cases where you also try to estimate the feedback effects (LeSage & Pace 2009). Basing the weight matrices on other things than geographical distance is also possible, such as socio-economical distance. This can however lead to endogeneity issues, since "it is very likely that these elements are correlated with the final outcome" (Qu & Lee 2015, p.209).

There is no given best version of a weight matrix, it has to be chosen based on previous studies, what the mapping of the data looks like and economic logic. Concerning previous studies, made in bigger cities, Kim et al. (1998) have opted for the distance decay matrix, where the inverse of the euclidean distance is being used and then standardized Kim et al. (1998), Liao & Wang (2012), Zheng & Hui (2016).

2.2 Related Research

Since there has not been done much research on house market employing spatial models due to lack of data, computational difficulties, and given that it is still a dynamic branch of econometrics, we were not able to find fully comparable studies. We present here the literature as closest related to our research as possible.

2.2.1 Non-spatial models

Borowiecki (2009) looks closer at the situation in Switzerland between 1991 and 2007, using a vector autoregression (VAR) model with the dependent variable *CPI adjusted house price index*. The author finds that an increase in construction costs is fully transferred to the buyer. We expect a cost increase to have a lesser impact on Swedish house prices due to Sweden being more integrated within the European Union, thus allowing for stronger competition and giving the buyers greater bargaining power.

To incorporate the demographic influences in house prices, Borowiecki (2009) uses the changes in population in the interval between 20 and 64 years old. A

1% increase in this population segment is associated with a 2 % increase in house prices. This may, however, be something fairly unique to the Swiss setting during the study where the market adjusts slowly: Switzerland has a difficult topography to build upon, a fairly restrictive environment when it comes to foreign labour and contractors, and a heavily regulated construction sector. Comparing with Sweden, the similarities may lie in the bigger cities where there is a lack of space to build and in the overall country where the construction industry is a fairly regulated business. Assessing the impact of housing stock, Borowiecki (2009) finds negative relationship with the house prices.

Davis & Zhu (2011) analyse house prices in 17 countries between 1970 and 2003 with focus on five explanatory variables: real commercial property prices, real credit to the private sector, real GDP, real short term interest rate, and real private investment. Using a vector error-correction (VECM) specification, implemented with generalised least squares (GLS), the authors find significant impact for credit to the private sector. All of the results of credit for these estimations are displayed in table 2.1. They explain the negative estimates of credit being due to credit increasing housing supply and hence pushing prices downwards after some time. Davis & Zhu (2011) find that in Sweden the direction of causality is stronger from credit to the property prices rather than vice versa.

	Pooled	Non-G7	Bank-Dominated	Crisis Countries
Δ Credit	0.75	0.71	0.84	0.67
	(6.4)	(4.6)	(5.2)	(3.0)
Lagged Credit	−0.09	Insignificant	Insignificant	−0.14
	(2.4)			(2.3)

Table 2.1: Estimates of effects of credit on house prices (Davis & Zhu 2011, p.25); lagged variable is in logs and **t-values in parenthesis.**

Vogiazas & Alexiou (2017) also find credit to GDP having positive impact on house prices. They examine how residential property prices are affected by economic fundamentals and also the business cycle using generalised methods of moments (GMM) on panel data between 2002 and 2015 for seven OECD countries. The unemployment, however, is found to be positive and insignificant. They attribute it to the fact that booms and busts during the period under study give rise to sign switches, where booms may give one sign for the variable and busts another - as described by Agnello & Schuknecht (2009). Results for credit to GDP for their different models are presented in table 2.2.

With data stretching between 1939 to 1994, Holly & Jones (1997) find that real income is the most important determinant as a 1% increase of the current year real income is associated with an 0.77% increase in real house prices.

	OLS	Fixed Effects	TSLS	GMM
Δ Credit to GDP	0.001	0.21	0.01**	0.252**
	(0.10)	(1.71)	(2.19)	(4.05)

Table 2.2: Estimates of effects of credit on house prices (Vogiazas & Alexiou 2017, p.126); credit is estimated in changes (first differences).**T-values in parenthesis**, where (*), (**) and (***) indicates significance at 10%, 5% and 1% respectively.

2.2.2 Spatial models

Employing a combination of quantile and spatial regression, Liao & Wang (2012) estimate the effect of surrounding house prices. They look at 46 356 residential property sales during one year in the urbanized areas of Changsha, and the effect of variables such as floor area, number of bedrooms, distance to green areas, and distance to central business areas. There are few shortcomings of this analysis: the total effects are not presented[3] in the paper, and their choice of spatial model is the Spatial Lag (SAR) model - which may lead to wrong conclusion concerning indirect effects (Elhorst 2014). In table 2.3, we can see the point estimate for the pooled 2SLS and all of the deciles, except the insignificant ones. They look at one city, during one year - but this may also indicate spatial dependence that should be accounted for on a more aggregated level, and during a longer time period.

	2SLS	0.1	0.2	0.3	0.4	0.5	0.9
Wy	0.0986**	0.2593**	0.1637**	0.1652**	0.1088**	0.0536**	0.1407**
	(0.0074)	(0.0254)	(0.0104)	(0.0111)	(0.0144)	(0.0001)	(0.0170)

Table 2.3: Estimates of effects of spatially lagged house prices Wy on house prices (Liao & Wang 2012, p.23). (*), (**) and (***) indicates significance at 10%, 5% and 1% respectively. **Standard errors in parenthesis**.

Wilhelmsson (2008) examines housing prices in Sweden on municipality level between 1989 and 1998. The author finds that different regions have different speed of adjustment of house prices. The author models the effect of disequilibrium, (K_{diff}), housing stock, income, employment, vacancy rates, mortgage rates and tax rates on the ratio of house prices to income (K). Wilhelmsson (2008) finds income to have a negative impact on the ratio. Increasing employment is associated with an increase in the ratio, housing stock exerts a positive influence when it is significant.

Similarly to Liao & Wang (2012), Wilhelmsson (2008) employs SAR model and does not include estimates of total effects. The spatial autoregressive parameter for the dependent variable is found to be significant. In table 2.4, we see his estimates for our variables of concern with *column (2)* and without a spatial lag of the dependent variable *column (1).*

[3] Including estimates of total effects is recommended by LeSage & Pace (2009).

	(1) 2SLS	(2) 2SLS with WK
Δ Income	-1.435	-1.149
	(-14.1)	(-13.8)
Δ Employment	0.088	0.001
	(4.6)	(0.1)
Δ Stock	0.752	0.855
	(2.8)	(3.9)
WK		0.948
		(37.9)
K_{diff}	-0.396	-0.584
	(27.3)	(-39.6)

Table 2.4: (Wilhelmsson 2008, p.46, p.51). Δ indicates first difference, dependent variable are annual changes in K, values are in log form and **t-values in parenthesis**.

Zheng & Hui (2016) analyse the effect of liquidity on housing returns in period of 1995 - 2011. They are using a Spatial Durbin Model (SDM) and a row stochastic distance weight matrix to estimate direct and indirect effects of changes in liquidity (turnover rate) and liquidity shocks. They also present the point estimates of the spatial correlation. The indirect effects exceed the direct effects, indicating the importance of including spatial elements in the model.

Chapter 3

Methodology

The following chapter provides an overview of the data and variables used and introduces our model. It also describes methodology connected to spatial autocorrelation such as Moran's I.

3.1 Data

Our dataset - a balanced panel - consists of observation across all Swedish municipalities, $n = 290$, between years 2003 and 2016, $T = 14$. The variables used for our analysis are summarized in the tables 3.1, 3.2.

Our dependent variable is house market price. Explanatory variables are:

- construction costs - an index measuring costs that the investor pays for constructing a new building; natural variable that is expected to have a positive effect on house price
- unemployment rate - we expect that unemployment rate has a negative effect on prices, because we assume that unemployed have lower purchasing power
- population density - higher population density is expected to increase house prices due higher demand
- income/cpi - we use real income as it is a better proxy for available funds of a household that can be spend on purchasing new housing
- stock/population - We choose to use this ratio rather than using these variables separately due to possible stationarity issues. The ratio is expected to be more stable, while it retains economic meaning; higher ratio means that more houses are available per the same number of residents, thus house prices should be lower

- credit - credit may be a fundamental driver of the house prices; it also serves as a proxy and controls for economic cycle and bubble creation

Variable name	Description	Source
m_house_price	Market value for single-family houses in TSEK; measured on municipal level	Statistics Sweden (SCB)
cost_h	Construction-cost index for individual houses; measured on country level	Statistics Sweden (SCB)
unemp	Unemployment rate measured on municipal level	Arbetsförmedlingen
pop_den	Residents per square kilometer; measured on municipal level	Statistics Sweden (SCB)
income	The total income for age group 16+ in TSEK; measured on municipal level	Statistics Sweden (SCB)
cpi	Consumer price index; measured on country level	Riksbanken
stock_h	Housing stock - houses; measured on municipal level	Statistics Sweden (SCB)
pop	Population; measured on municipal level	Statistics Sweden (SCB)
credit_r	Outstanding amount of credit to households in percentage of GDP; measured on country level	Bank for International Settlements

Table 3.1: Description of the variables

Variable	N	Mean	St. Dev.	Min	Max
m_house_price	4 060	1 426.412	1 159.033	210.600	11 597.040
cost_h	4 060	1 651.971	180.034	1 345.700	1 889.600
unemp	4 060	0.035	0.013	0.006	0.100
pop_den	4 060	136.600	474.172	0.200	5 496.400
income	4 060	224.116	37.038	157.600	501.100
cpi	4 060	45.434	2.165	41.949	48.134
stock_h	4 060	6 951.886	6 256.226	506	54 840
pop	4 060	32 418.900	64 552.140	2 421	935 619
credit_r	4 060	71.659	10.541	53.100	84.550

Table 3.2: Summary statistics

3.2 Estimation Technique

As discussed in the section 2.1, researches are advised to use the SDM specification when aiming to capture spatial dependencies in the model. Thus our model is specified as an SDM extended from cross-sectional to panel specification:

$$\mathbf{Y}_t = \rho \mathbf{W} \mathbf{Y}_t + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{W} \mathbf{Z}_t \boldsymbol{\theta} + \mathbf{a} + \gamma_t + \boldsymbol{\epsilon}_t, \tag{3.1}$$

where Y is the dependent variable. ρ is the spatial autoregressive parameter of the dependent variable, W is the weight matrix. X is a set of independent variables and Z is a set of spatially lagged independent variables, where θ is spatial autoregressive parameter of the independent variables; ϵ is assumed to follow $N(0, \sigma^2)$. $\mathbf{a}$ is a constant for each unit - since we are using fixed effect model, and γ_t controls for time.

Having spatially correlated data, our focus is not primary on the estimated coefficients themselves, but rather on the total effects which take into account spatial feedback from the neighbouring regions. As Elhorst (2014) argues another advantage of SDM over SAR is that direct and indirect effects of variable k do not depend solely on ρ and W, but are also based on θ_k. SDM being not so restrictive leads Elhorst (2014) to the conclusion that SDM is "more attractive point of departure in an empirical study than other spatial regression specifications." (Elhorst 2014, p.22)

Since construction costs, and credit are available only on the country level, they are not spatially lagged i.e. not included in Z. We estimate log-log model; the only exception is $credit_r$ variable. We took differences of $credit_r$ - household debt measure - due to its high multicollinearity with our $cost_h$ variable - the construction cost index. VIF summary is provided in Appendix A table A.4.

3.3 Spatial Autocorrelation

To assess whether a spatial model should be applied there are various ways to start, where the easiest way to present is initially by only using cross-sectional data. Where, firstly, a visual inspection of the distribution of the variable we are examining can be useful, giving a quick, but a bit shallow, understanding of the situation.

Other ways to determine whether a spatial analysis is appropriate is by looking at the Moran's I and the Moran Scatter Plot. Moran's I (Global Moran's I) measures the spatial autocorrelation of a certain trait. It estimates whether the dispersion is random, dispersed or clustered (Michael D. Ward 2008). The formula is the following, for $i \neq j$:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \overline{y})(y_j - \overline{y})}{(y_i - \overline{y})^2} \tag{3.2}$$

Here the w_{ij} represents an element of the weight matrix and y some variable of interest. The formula gives a Global Moran's I index. This index is evaluated against the expected Moran's I, testing for significance with a z-score (indicating what kind of spatial correlation) and a p-value. The null hypothesis is that there is no spatial autocorrelation. To produce the test scores the following moments are used:

$$\begin{aligned} \mathbb{E}(I) &= \frac{-1}{(n-1)} \\ var(I) &= \mathbb{E}(I^2) - \mathbb{E}(I)^2 \end{aligned} \tag{3.3}$$

Inserted in the below formula:

$$z_I = \frac{I - \mathbb{E}(I)}{\sqrt{var(I)}} \tag{3.4}$$

What is of importance is how the weight matrix is specified. One of the features that is recommended to use is a row standardised matrix, making sure each row sums to one. This produces a Moran's I with an interval between -1 and 1. With

-1 representing the case where the distribution is dispersed: high values surrounded by low and vice-versa, 0 being the case where the distribution is spatially random and 1 indicates spatial clustering of similar values (Michael D. Ward 2008, Roger S. Bivand 2008)

After analysing the Global Moran's I, plotting all the Local Moran I's is also helpful in a Moran Plot. The Local Moran's I is calculated the following way Roger S. Bivand (2008):

$$I_i = \frac{(y_i - \overline{y}) \sum_{i=1}^{n} w_{ij}(y_j - \overline{y})}{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n}} \tag{3.5}$$

Plotting all of these individual I's, we visualize the clustering of the data, and by adding an OLS line to the plot we produce the Global Moran's I, which is the slope.

Here we have the variable of interest on the x-axis and the spatial lag of the variable on y-axis. What should be plotted is both the variable of interest and the residuals of OLS and spatial estimating techniques. The plot's upper right quadrant contains high values surrounded by high values and the lower left quadrant contains low values surrounded by other low values. These two quadrants represent positive spatial correlation, and the upper left (low/high) and the lower right (high/low) represent negative spatial clustering: values surrounded by dissimilar values. Figure 3.1 is an example of the plot from the R-package "spdep":

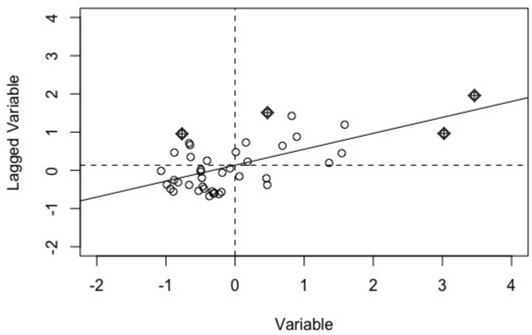


Figure 3.1: A Moran Scatter Plot

When working with panel data we can employ a randomization test of Pesaran's CD test (Millo 2017, Pesaran 2004), where the test estimates cross-sectional dependence and is defined in the following way:

$$CD = \sqrt{\frac{2T}{N(N-1)}}(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}); \quad \hat{\rho}_{ij} = \frac{\sum_t e_{it}e_{it}}{(\sum_t e_{it})^{1/2}(\sum_t e_{it}})^{1/2}, \tag{3.6}$$

where $\hat{\rho}_{ij}$ is the residuals' estimated pair-wise correlation. The CD formula captures averages of time for the pairwise correlation between cross-section entities. What we use is a modified version of the $CD(p)$ test (Millo 2017, Pesaran 2004), that takes the following form:

$$CD_p = \sqrt{\frac{T}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w(p)_{ij}}} (\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} [w(p)]_{ij} \hat{\rho}_{ij}) \tag{3.7}$$

This includes a proximity/weight matrix. Where p in $w(p)_{ij}$ determines the order of the matrix and ij the element. The null hypothesis for this test is that there is no spatial dependence. To arrive at a conclusion we use a permutation test where we reassign the weights randomly to the spatial weight matrix. If spatial dependence is present, then rearranging the connections should destroy such dependence. The alternative hypothesis is that the value of CD_p with its associated matrix is not random, and that it has more extreme values. There are different versions of the test, depending on whether you have a heterogeneous model or a homogeneous one, what order the matrix is of, if you look at the residuals or the "raw data"; what to chose depends on the data, tests and assumptions. This test is robust to global dependence and serial correlation (Millo 2017).

Chapter 4

Estimation

In this chapter, we discuss stationarity of our data, introduce the design of the weight matrix employed in the analysis, demonstrate that the data is spatially dependent, and perform basic residual analysis.

4.1 Stationarity

For sake of intrepretability, we transform the data only by logaritmization to reduce heteroskedasticity. Suspecting that we have unit root processes in the data, we run Pesaran's cross-sectionally augmented Dickey–Fuller (PCADF) test, which accounts for cross-sectional dependencies in the panel. Unlike recent literature, we choose not to run LLC or IPS unit root tests as their crucial assumption is cross-sectional independence (Baltagi 2008). As expected PCADF cannot reject the null of unit root being present. Nonetheless, all variables seem to follow I(1) processes, thus we utilize Kao, Pedroni, and Westerlund tests[1] for cointegration; since all 3 tests reject the null of no cointegration, it is safe to assume that there is no issue of spurious regression. We further decrease the issue of non-stationarity by utilizing a fixed effects specification; as Davidova (2015), Fidrmuc (2009) argue possible bias is small compared to the methods which directly control for non-stationary panels such as *Fully modified OLS* and *Dynamic OLS*. We also control for time to further capture any effects.

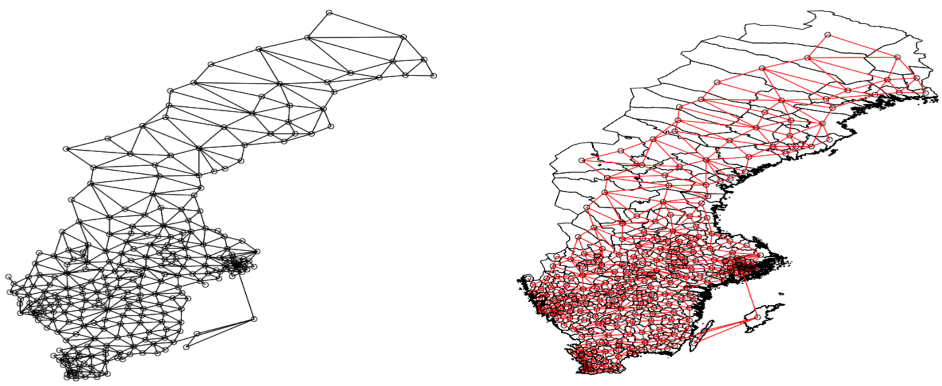
4.2 Weight Matrix

Our main weight matrix used is a matrix based on the Queen criterion. It specifies spatial association in a binary way: 1 if they share a border, zero otherwise. This matrix is then row standardised. We use this as our main matrix because if

[1] Cointegration tests as well as PCADF tests in Appendix A

we use distance we run into the risk of not giving any neighbours to the bigger municipalities in the north. This comes from the fact that we are using “point data” which means that there is one individual point within the municipality representing that area, as opposed to lattice/area data covering the entire area. Using a distance based matrix would yield very few or no neighbours for large municipalities. If we use the k-nearest neighbour we may over exclude for municipalities with many neighbours or over-include neighbours for municipalities with a long distance to the nearest neighbour. As discussed in the section 4.3 Sweden has a great variety in the amount of neighbours for the municipalities and the size of the municipalities.

Islands with a single municipality on them have to be configured manually. In our case this is Gotland, Lidingö and Öckerö. We base the configuration for Lidingö and Öckerö on k-nearest neighbour and for Gotland we chose Stockholm (due to the fact that it is a very popular summer residence for people from Stockholm) and the nearby municipalities on Öland (an island) and on the nearby mainland. A visual representation of our full weight matrix is depicted in figure 4.1. Nonetheless as a robustness check, we estimate our model with the two other approaches: k nearest neighbours and distance-based matrix. Since the matrix design is set exogenously, robust results in this regard are crucial.



(a) Connections without map (b) Connections with map

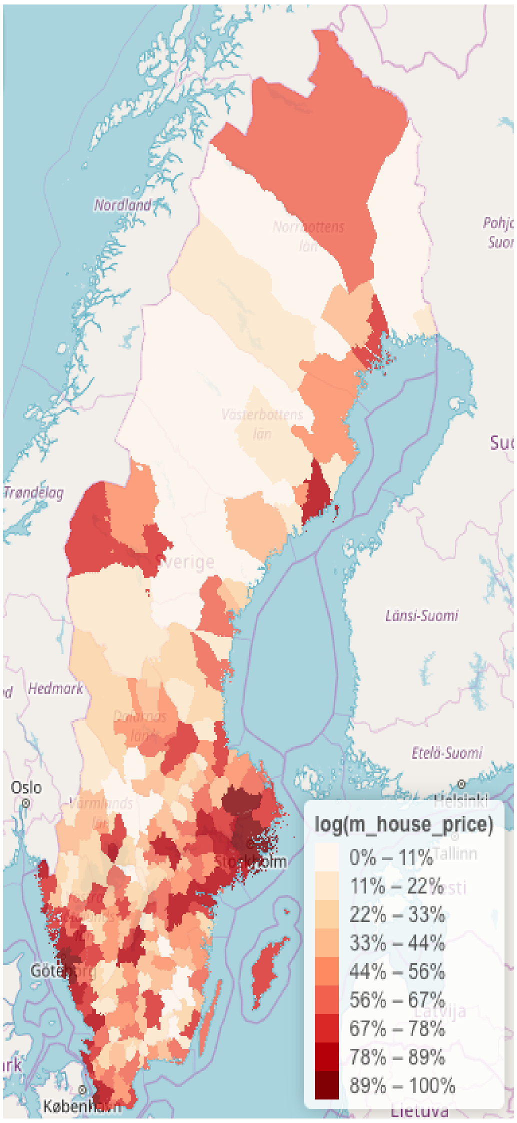
Figure 4.1: Full weight matrix

4.3 Spatial Autocorrelation

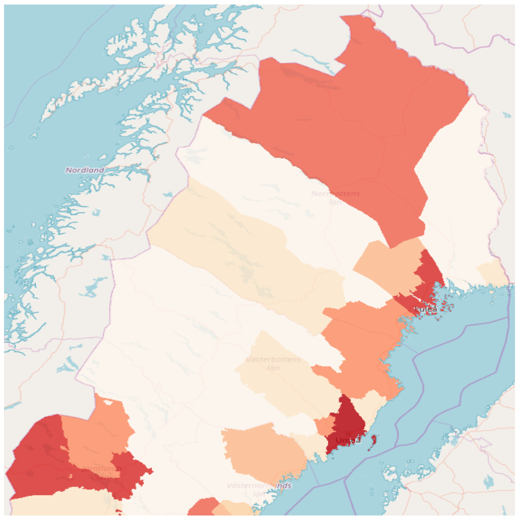
There are several methods to estimate Moran’s I, but we make use of two. The first one implements the formulas described in the section 3.3 (under the normality assumption and with the correcting randomization assumption), and the second way is to estimate the I using permutations: a Monte Carlo simulation made 1999 times, where the observed data is randomly reassigned to the spatial entities. Given enough simulations we can approximate the true distribution.

To implement Pesaran's $CD(p)$ test we use our final model as described in section 3.1 and residuals obtained from the regression. The full length of our dataset can be used and the test we run is for models that allow for heterogeneity among the individual entities (as suggested after a Chow test has been performed). The simulation makes use of 1999 permutations.

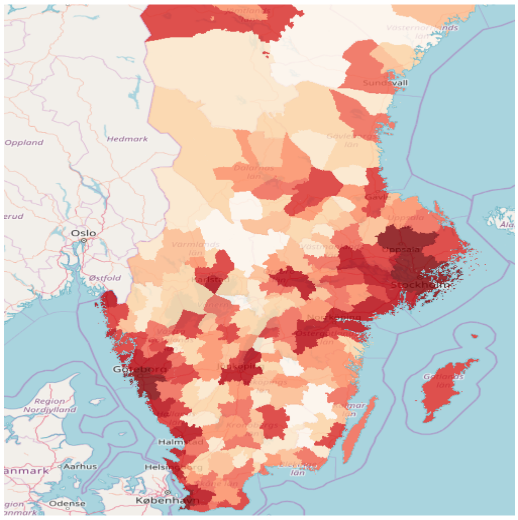
To get a sense for whether the case could be made that there is spatial autocorrelation in the data we plot the logged average market house prices for all the municipalities for year 2016 in figure 4.2.



(a) The entire country



(b) Zoom from Sundsvall to Kiruna



(c) Zoom from Sundsvall to Ystad

Figure 4.2: The distribution of logged average market price house prices for all municipalities, year 2016. The more intense red indicates higher value.

Looking at the level of house prices plotted in the figure 4.2, we can see that there is a difference between the north and the south. Moreover, for the southern part there is a concentration of higher house prices in east, west and south around the major cities: Stockholm, Göteborg and Malmö.

Here we present the results of the Global Moran's I for 2016, which provides an indication that there is spatial clustering.

In the table 4.1, we can see that when we are testing for global autocorrelation we get significant estimates and almost the exact same values for the observed Moran's I; for the ordinary Moran Test, both under randomization and under the less strict normality assumption. As well as for the Monte Carlo simulation.

	$Moran'sI$	$\mathbb{E}(I)$	$Var(I)$	$z-score$	$p-value$
Moran's I (r)	0.669	-0.003	18.059	0.001	<2.2e-16
Moran's I (n)	0.669	-0.003	0.001	18.069	<2.2e-16
Moran's I (MC)	0.669				5e-04

Table 4.1: Moran's I estimates for house market price variable

The result of Pesaran's $CD(p)$ test rejects the null hypothesis with a p-value of 0.0145 for a one sided test.

If we have a look at the Moran Plot 4.3, where we plot all the Local Moran estimations, we see that the we have positive spatial autocorrelation when we fit a line to the dot plot (the slope being equal to the Moran Test under normality assumption). The significant values are depicted in colour, and the insignificant ones in white.

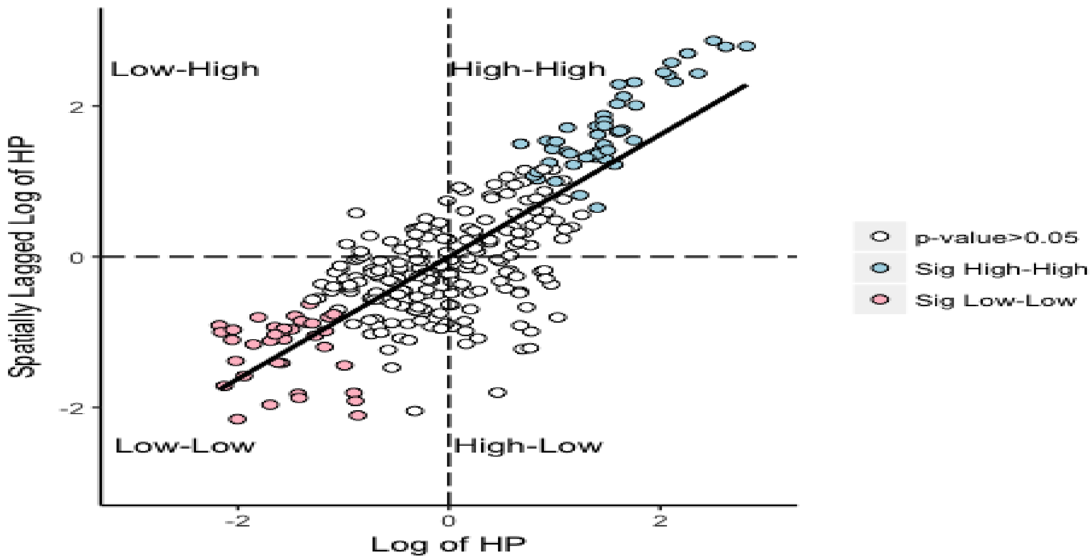


Figure 4.3: A Moran Scatter Plot

4.4 Model Specification

Before moving to spatial models, we run ordinary panel regressions - random and fixed effects model with HAC errors. The Hausman test suggest that the FE model is appropriate specification. This is in line with the economic reasoning. FE allows for correlation between explanatory variables and unobserved heterogeneity among individual units; while RE allows for no correlation. In our analysis, we include variables that are measured on municipality level such as income, unemployment, or population. Since it is very unlikely that our model captures all the heterogeneities, we tend to assume that they are correlated with explanatory variables. For example, share of skilled labour to total labour force affects income as well as unemployment levels in a given municipality. Further, the analysis is run on data spanning from 2003 to 2016, thus such heterogeneity among the municipalities is rather stable; under these settings the FE model seems to suit our analysis better. We run also F test for individual effects, which confirms the hypothesis that the FE model is more suitable than pooled OLS - see table 4.2. The result is also confirmed for spatial models.

F test for individual effects
F = 147.97, df_1 = 289, df_2 = 3474, p-value = 0.000
Hausman test
chisq = 137.75, df = 6, p-value = 0.000

Table 4.2: Tests to determine correct specification

Breusch-Pagan LM test for cross-sectional dependence in panels as well as Pesaran CD test reject the null of no cross-sectional dependence at 1% significance level, thus we have further evidence of spatial dependencies in our panel. Breusch-Godfrey/Wooldridge test for serial correlation in panel models cannot reject the null of no serial autocorrelation and Breusch-Pagan test cannot reject the null of homoskedasticity. Therefore, we decide to use standard errors introduced by Driscoll & Kraay (1998). The estimator is robust to "general forms of cross-sectional as well as temporal dependence." (Hoechle 2007, p.5).

Chapter 5

Results

5.1 Main Findings

In table 5.1, we report results obtained from three different estimation frameworks. Columns (1) and (2) provide estimates of non-spatial models, while column (3) gives us estimates obtained from Spatial Durbin Model. We include column (1) and (2) just for comparison and we mainly focus on the spatial model - column (3). We observe upward bias in construction cost as well as in credit estimate, while upward correction can be seen for real income and population density estimates when comparing the non-spatial models with the spatial one. Housing supply and unemployment appear to be insignificant. Nonetheless, the focus should be mainly on the total effects which are computed for the spatially dependent variables in SDM framework and capture the total impact of the variables. We can directly interpret construction cost and credit, which are available only on country-level, thus are not spatially dependent in our setting.

To assess the predictive power of the models, we run out-of-sample forecast for years 2013 - 2016 and compute RMSE (root mean square error). We observe that the spatial model outperforms the non-spatial ones - see 5.2. Thus confirming our hypothesis that allowing for spatial dependencies, when estimating the determinants of the house prices, is a preferable option.

After the SDM estimation, we tests two hypotheses $H_0 : \boldsymbol{\theta} = \mathbf{0}$ and $H_0 : \boldsymbol{\theta} + \rho\boldsymbol{\beta} = \mathbf{0}$ suggested by Elhorst (2014) to determine whether Spatial Lag Model or Spatial Error Model fit the data better than Spatial Durbin Model[1] (SDM). Both hypothesis are rejected - see 5.3, thus we conclude that SDM model should be used.

[1]For model taxonomy, plese refer to figure 2.1

	Dependent variable: log(m_house_price)		
	Non-spatial		Spatial
	Fixed	Random	SDM
	(1)	(2)	(3)
Main			
log(cost_h)	1.389***	1.347***	0.680***
	(0.031)	(0.031)	(0.114)
dl_credit_r	0.585***	0.589***	0.307***
	(0.077)	(0.070)	(0.050)
log(income/cpi)	0.793***	0.924***	1.213***
	(0.051)	(0.049)	(0.192)
log(pop_den)	0.552***	0.293***	0.652***
	(0.043)	(0.011)	(0.064)
log(stock_h/pop)	−0.060*	−0.073**	0.007
	(0.034)	(0.030)	(0.010)
log(unemp)	−0.071***	−0.072***	−0.021
	(0.006)	(0.006)	(0.015)
Constant		−5.755***	
		(0.188)	
ρ			0.527***
			(0.036)
Wx			
log(income/cpi)			−0.886***
			(0.257)
log(pop_den)			−0.396***
			(0.110)
log(stock_h/pop)			−0.101**
			(0.045)
log(unemp)			−0.020
			(0.012)
Observations	3,770	3,770	3,770
Adjusted R^2	0.820	0.828	0.835
Note:			*p<0.1; **p<0.05; ***p<0.01
			For SPML Pseudo R^2 computed
			standard errors in parenthesis

Table 5.1: Comparison of estimates obtained by FE, RE, and SPML

Model	RMSE
Non-spatial FE model	0.341
Non-spatial RE model	0.320
Spatial Durbin model	0.104

Table 5.2: Predictive power of the models

H_0 : the spatial lag model best describes the data
$\boldsymbol{\theta} = \mathbf{0}$
$\chi^2_4 = 109.082$
Prob $> \chi^2 = 0.000$
H_0 : the spatial error model best describes the data
$\boldsymbol{\theta} + \rho\boldsymbol{\beta} = \mathbf{0}$
$\chi^2_4 = 48.492$
Prob $> \chi^2 = 0.000$

Table 5.3: Testing the models

5.1.1 Construction Costs

Increase in construction costs by 1% leads to an increase of house prices by 0.68%. The estimate is decreasing in absolute terms, comparing with the non-spatial models, which according to Lesage (2008) is expected since the SDM model also attributes changes of the dependent variable to the spatial lag of the dependent variable.

The full transfer of costs present in Borowiecki (2009) is not observed here. This could be explained by better bargaining position of house buyers due to the supply being less rigid than in Switzerland. Borowiecki (2009) specifically points to Swiss foreign labour restrictions, difficult topography, and strict regulation in the construction sector.

5.1.2 Credit

A boost in credit is also connected with a boost in house prices. It decreases when going to the spatial model, possibly due to the reasons cited by Lesage (2008). The estimate exceeds that of Vogiazas & Alexiou (2017), their estimate is 0.252 for the sample where Sweden is included. This may be due the higher dependence on credit when it comes to house purchases in Sweden e.g., found by Davis & Zhu (2011).

Variable	Impact measures: Direct	Indirect	Total
log(income/cpi)	1.179***	-0.491	0.688**
log(pop_den)	0.644***	-0.086	0.557**
log(stock_h/pop)	-0.006	-0.198**	-0.205**
log(unemp)	-0.025	-0.062***	-0.087***
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5.4: Total effects

5.1.3 Real Income

We find real income to be the most important determinant of house prices - similarly to e.g., Holly & Jones (1997). The main coefficient estimate is 1.213 for the SDM model, and the direct effects estimate is 1.179. However the contribution of the spatial connectedness of the indirect effects does not seem to be significant. The negative coefficient estimate of $Wlog(income/cpi)$ and the decrease in magnitude when looking at the significant total effects compared with the direct effects may allude to a competitiveness. Where an area draws to it highly sought after, high income labour, causing people to move there from nearby areas which reduces prices in the place from which they leave.

Our direct effect estimate indicates that a 1% increase of real income would push prices up by around 1.2%. This estimate is not far from the coefficient estimate of -1.149 on the price to income ratio in Wilhelmsson (2008) spatial model. Our model, to some extent, contradicts usual findings where the exogenous variable's coefficient estimate usually decrease in absolute terms when introducing spatial interaction (Lesage 2008). For Wilhelmsson (2008), such is not the case: introducing spatial dependence decreases the magnitude of income. This may be due to the fact that Wilhelmsson (2008) uses a SAR model. As discussed in section 2, SAR may be too restrictive, since it does not allow for the estimation of spatial autoregressive parameter for the independent variable θ. In our model, θ for *Real Income* is negative - leading to higher main coefficient estimate and direct effect estimate, but a lower total effects estimate. So the total effect is 0.6% increase in house price for a municipality which experiences 1% increase in real income.

5.1.4 Population Density

Looking at the direct effect, a 1% increase would cause an 0.6% push upwards in house prices. 86% of total effects are made up of the direct effect. Perhaps then, it is as such that an increase population density does cause people to move out - there is no spillover push. Which, possibly, is logical, because if you increase the density of an area, perhaps caused by people in surrounding areas moving there, then you would not expect the density of the surrounding areas to increase and push prices upwards. This perhaps also explains the negative coefficient estimate of Wx for population density.

We do not find the same impact of population changes as Borowiecki (2009). This may be attributed to the different characteristics specific to Switzerland. Sweden seems to accommodate the demographic changes in a better way. Perhaps due to the more obvious differences in topography and openness to foreign labour.

5.1.5 Housing Stock

The surrounding areas' housing stock seem to be exerting the major influence of this variable. The total effects are made up of 97% indirect effects. Perhaps this is because including also the surrounding areas, we are able to capture a significantly larger change: altering the stock of all the surrounding areas would be sufficient to move prices, however changing the housing stock within a municipality would not be enough to alter prices. In the case of housing stock, the spatial component plays the major part, hence not accounting for the spatial dependencies seems to yield biased results. We find that an increase in house/capita of 1% in surrounding areas would lead to 0.2% decrease in the original municipality.

If we look at the impact of increases to housing stock in Wilhelmsson (2008) study, he finds it to have a positive impact. This may be because he looks at housing stock, and we look at housing stock per capita. Perhaps the stock is increased where it is most profitable, hence the positive estimate, whereas an increase in stock per capita exerts a negative influence. This is in line with economic theory, increasing the housing supply, having the number of people fixed i.e., demand is fixed, leads to decrease in house prices. We assume that there is no feedback effect from house prices to housing stock. Firstly, we use a ratio stock/population. Secondly, any possible sources of endogeneity are captured by fixed effects. We assume that these sources are time-invariant e.g., developers are motivated to build houses rather in Stockholm than in rural areas due to higher prices in Stockholm throughout the examined period; we have not discovered any reason, why this should not hold over the period of interest.

5.1.6 Unemployment

Unemployment derives its main influence from the surrounding areas, indicated by the magnitude and significance of the indirect effect. It makes up 71.3 % of the total effects. Increasing the unemployment by 1 % would be associated with an indirect effect where house prices would go down by 0.62 %. So increasing unemployment in the surrounding areas, which in our case means only unemployment - since we have included real income in our regression, may put downwards pressure on prices. The effect may stem from a decrease in future wage expectation as well as higher uncertainty. Comparing these results with Wilhelmsson (2008), we see that the non-spatial models produce similar estimates of unemployment, but his spatial model yield insignificant estimates. Wilhelmsson (2008) does not produce the indirect effect of employment estimate which looking at our results seems likely to have a significant impact. Another possible explanation for the insignificant result of the direct effect could be attributed to what Agnello & Schuknecht (2009) infer, that is: booms and busts in the economy can lead to sign switches and unclear estimates. Our study period contains both: from 2003 to 2016, with the financial crisis around 2009. This in combination with the fact that we also include income in our analysis, hence we isolate the effect of unemployment only, not a change in income, may result in lower estimates and lower significance.

5.2 Depiction of the Spillover Effects

Here, we provide illustration of the spillover effects for the 3 most populated municipalities: Stockholm, Gothenburg, and Malmö; 1 mid-size municipality: Linköping; and two smaller municipalities (in terms of population): Mora in the middle of Sweden, and Kiruna in the north. We simulate 5% increase in a given variable in 2015 and plot the change in house price in the same year. For sake of clarity, we plot the effects that exceed one standard deviation of the mean of the effects. The more solid red, the stronger the effect is.

According to our results discussed in the previous section, we can observe that the lowest spillover effect is associated with changes in population density; while largest spillovers are observed when housing stock or unemployment changes. Note that while having large spillover effects, the total effects are smaller than for real income and population density due to insignificant direct effects of housing stock and unemployment. We do not observe significant differences in spillover patterns for different municipality sizes (in terms of population).

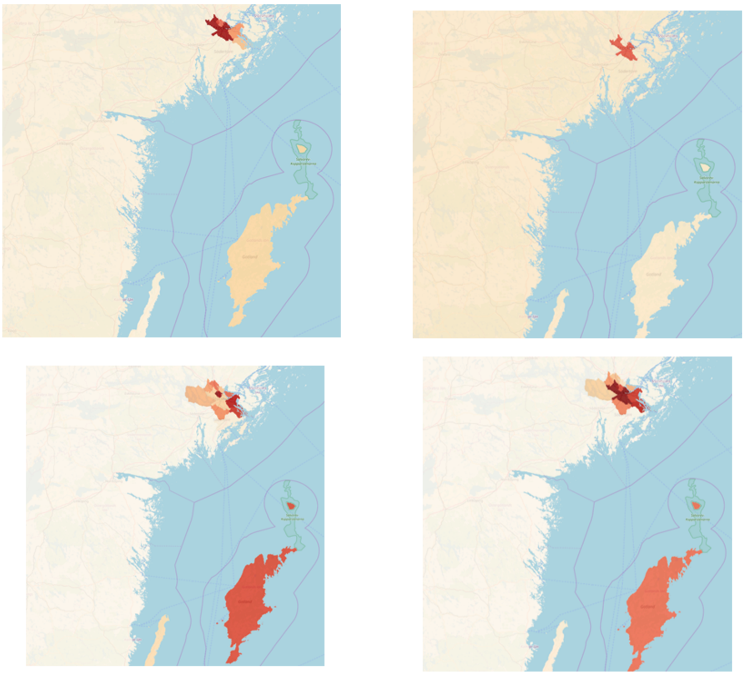


Figure 5.1: Spatial spillover for Stockholm municipality
top left: Real income top right: Population density
bottom left: Housing stock bottom right: Unemployment

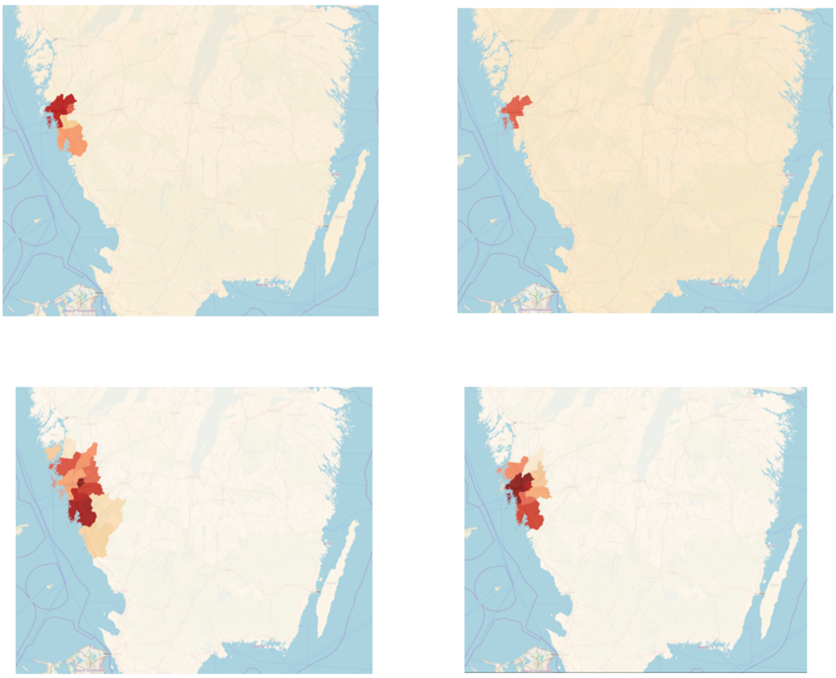


Figure 5.2: Spatial spillover for Gothenburg municipality
top left: Real income top right: Population density
bottom left: Housing stock bottom right: Unemployment

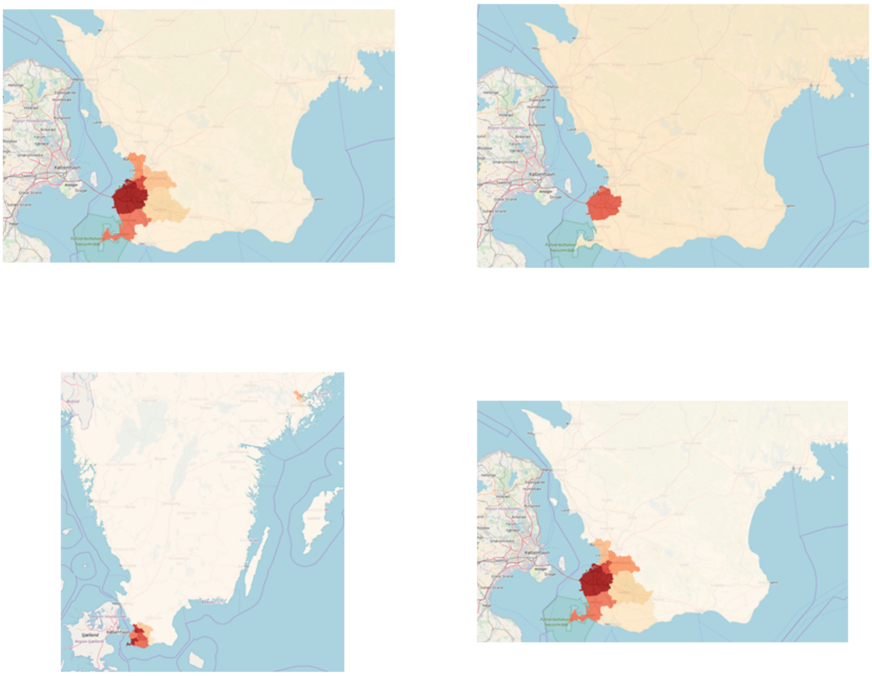


Figure 5.3: Spatial spillover for Malmö municipality
top left: Real income top right: Population density
bottom left: Housing stock bottom right: Unemployment

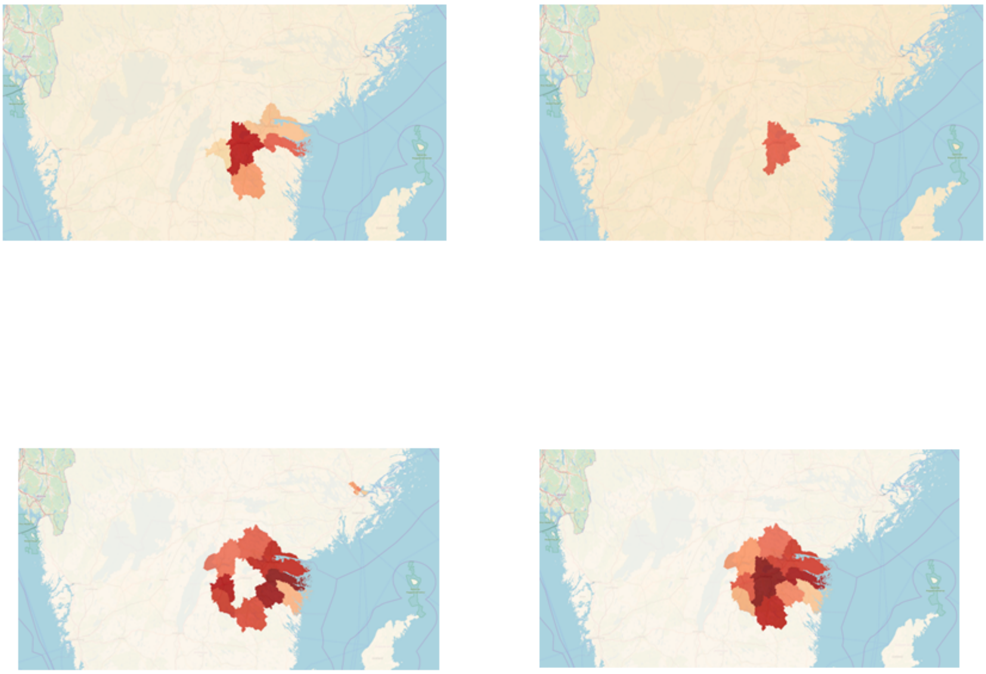


Figure 5.4: Spatial spillover for Linköping municipality
top left: Real income top right: Population density
bottom left: Housing stock bottom right: Unemployment

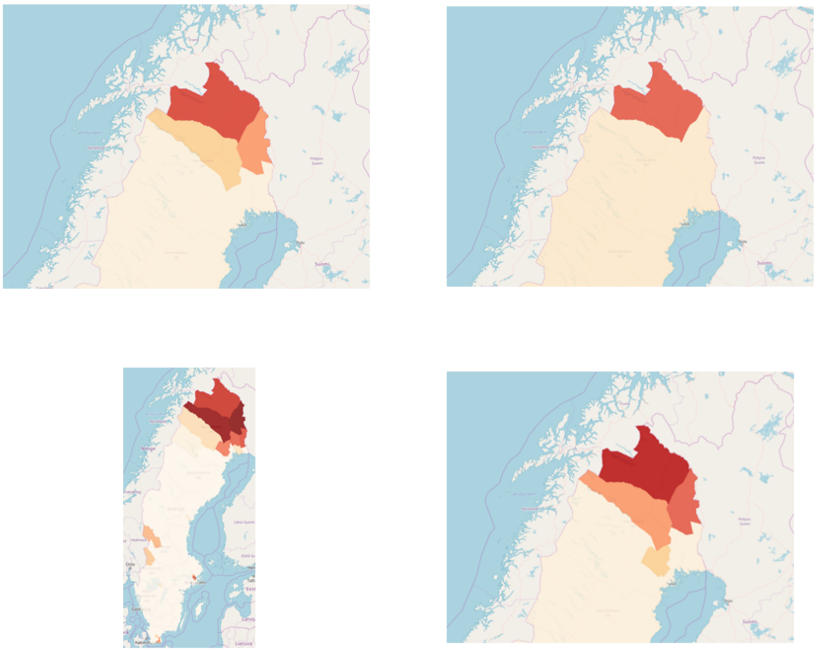


Figure 5.5: Spatial spillover for Kiruna municipality
top left: Real income top right: Population density
bottom left: Housing stock bottom right: Unemployment

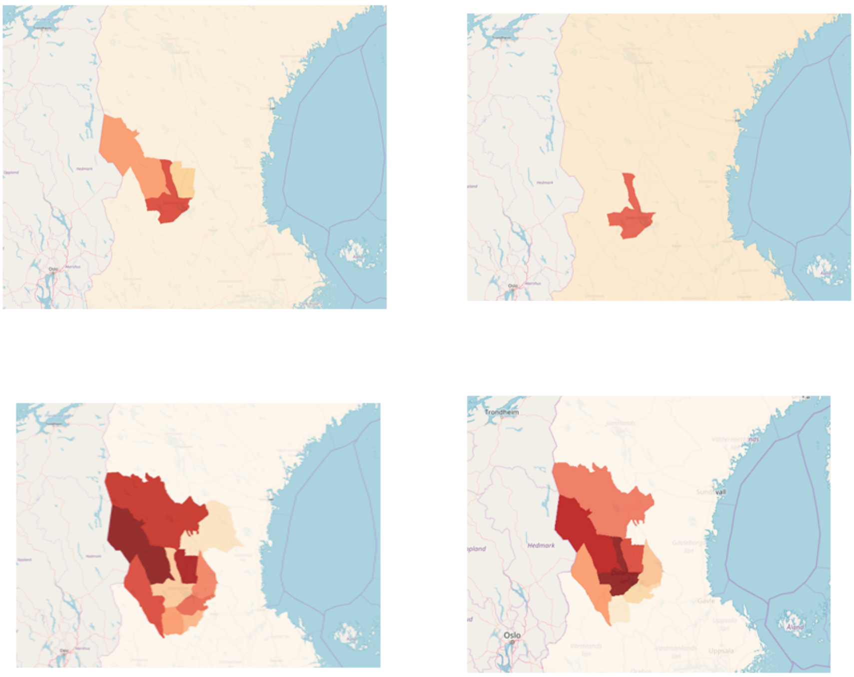


Figure 5.6: Spatial spillover for Mora municipality
top left: Real income top right: Population density
bottom left: Housing stock bottom right: Unemployment

5.3 Development of the Spatial Effects over Time

When looking only at the cross-section of the data, for each year, we eliminate some of the possible problems that may arise with time series, e.g., non-stationarity, spurious regression results and necessary transformation of data. Running the regression like this, and deriving all of the direct, indirect and total effects gives us a good sense of the development of the variables and their effect on house prices. On the other hand, we need to drop variables which do not vary over municipalities i.e., construction cost, and credit. We observe which variables were the most decisive for the house prices in a given year.

5.3.1 Real Income

In this setting, we see perhaps a more reasonable estimation of the effect of income: it carries with it also a positive indirect effect, however it is reduced during the financial crisis. A plausible reason for this could be that income increase did not induce people to relocate during the crisis period - due to uncertainty you would rather wait even with an increase in income. Compared to the other variables, this one is the most influential for the house prices. This is understandable, increasing real income has an immediate direct effect on prices.

5.3.2 Population Density

The effects of population density exhibit stable development and keeping a rather minor effect on prices. The impact of population density probably takes more time to become observable. Within a given year, prices do not react as quickly as it was the case for income. We cannot conclude that density has small effect on prices, we can only imply that the effect is smaller in the given year.

5.3.3 Housing Stock

The effects of increasing the housing supply seems to be continuously negative, with a larger magnitude for the indirect effect (corresponding with the panel estimation). But the impact of an increase in supply seems to be reduced during the financial crisis, especially the indirect effects. It implies that the relationship between housing supply and house prices became slightly weaker; it may be partly explained by e.g., low interest rate environment (BIS 2018), thus better availability of credit to households; this might push demand up, therefore the effect of supply was not that strong after 2010.

5.3.4 Unemployment

Here we see the sign switch mentioned by Agnello & Schuknecht (2009), where unemployment is slightly positive in 2003, and drops severely around the financial crisis. The direct and indirect effects appear to move in tandem, but the magnitude of the indirect effect is always greater - perhaps due to the fact already mentioned: we capture the effect of changing all of the surroundings. However, after the financial crisis, we see stabilization of the coefficient converging to the pre-crisis levels. Having coefficients close to zero makes sense, keeping in mind, we include real income in our regression; thus we observe the change in house prices due to the change in unemployment, while having real income fixed.

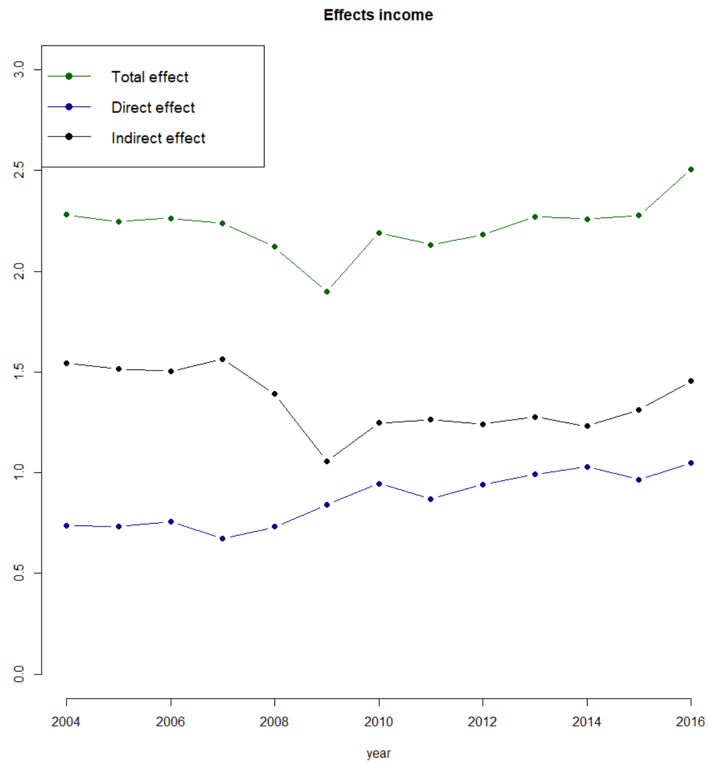


Figure 5.7: Development of the effects of income

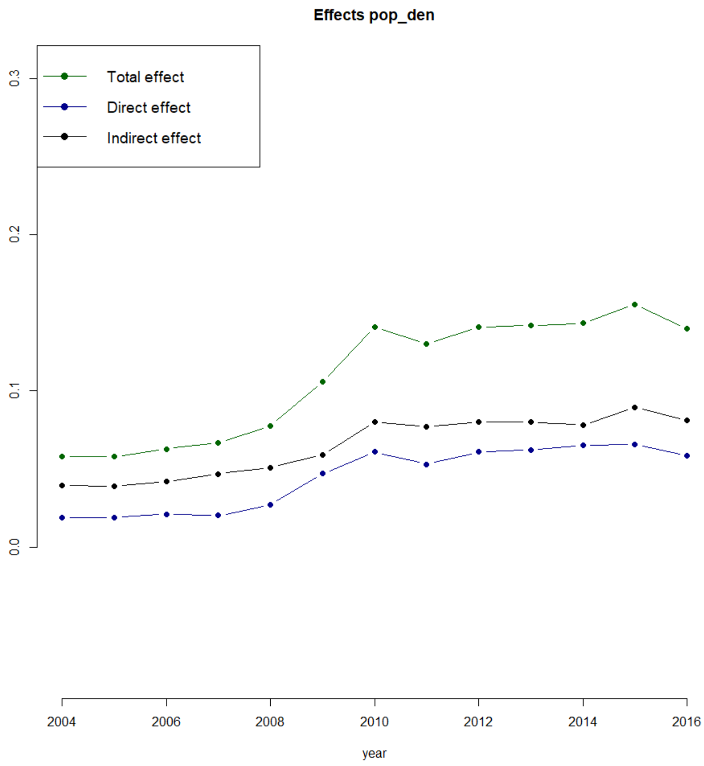


Figure 5.8: Development of the effects of population density

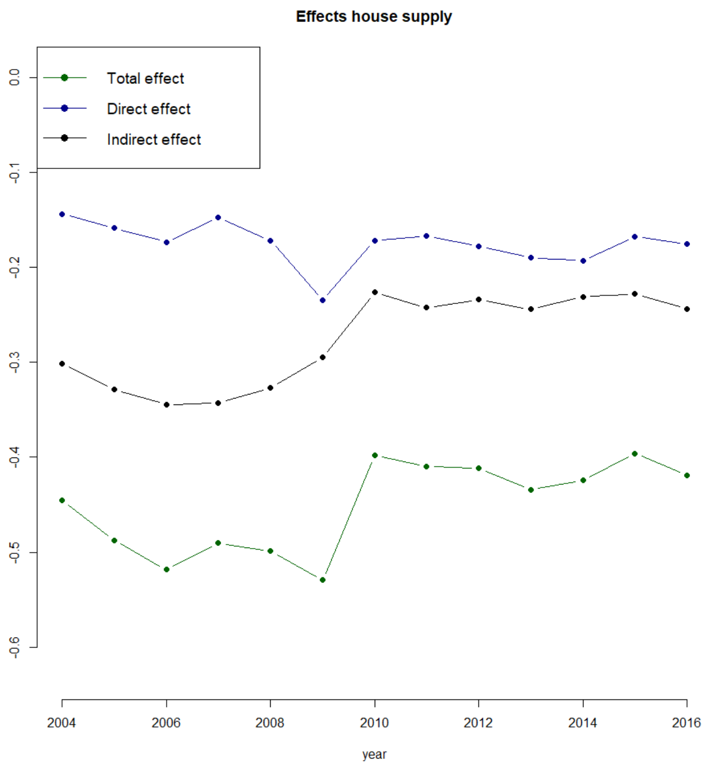


Figure 5.9: Development of the effects of stock

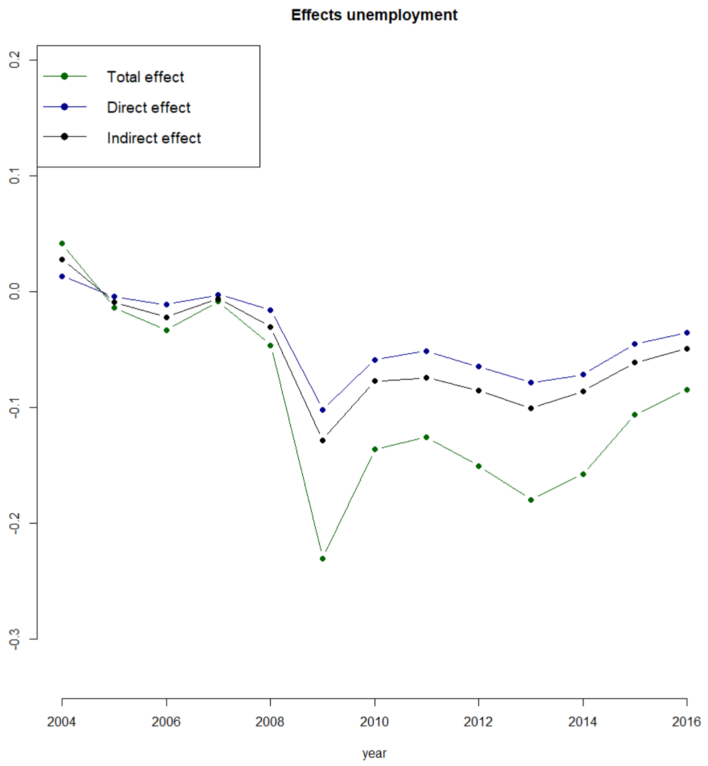


Figure 5.10: Development of the effects of unemployment

Chapter 6

Conclusion

The thesis analyses the determinants of the Swedish house prices. We utilize the panel of 290 municipalities between years 2003 - 2016. We employ the Spatial Durbin Model (SDM), which allows us to capture spatial dependencies in the data, thus obtaining unbiased estimates of the main determinants. Techniques not allowing for spatial relations in the data may yield biased estimates. We also run cross-sectional SDM for every year in the data set, so we can examine the behaviour of the variables over years. This proves to be especially valuable having the period of the financial crisis of 2008 in our data set. Spatial models can be peculiar since one needs to construct the weighting matrix exogenously and define relationships between the units; nonetheless our results are robust to various specifications of the matrix. We perform out-of-sample forecast to compare predictive power of non-spatial and spatial model - showing that spatial model performs better.

We find that the main determinants for the Swedish house prices are construction costs and real income. We show that the developers are not fully able to transfer the increase in construction prices on to the final buyers. For every 1% increase in the cost, there is only 0.68% increase in the house prices. We further show that non-spatial models give upward-biased estimates. Real income is another major determinant, having this variable available on municipality level, we compute direct, indirect, and total effects for the variable. While the direct effect is very high: 1.18% in house prices due to 1% increase in real income, it is reduced when total effect is computed, capturing the spatial relations in the data, having value of 0.69. We also observe the varying importance of real income across the years. Possibly due to high uncertainty in the markets during the financial crisis, people chose to wait with the house purchases, therefore the estimates exhibit smaller values in 2009; however they recovered quickly and are now at the comparable levels as during the pre-crisis years.

Another variable that is only available on country level, apart from construction cost, is outstanding credit to households as a % of GDP. Our estimate of 0.31%

increase in house prices for 1% increase in the variable, is slightly higher than others have estimated. This can be due to fact that Swedish households are prone to indebtedness and rely more on floating rates than some other countries e.g., Germany. Another determinant that has a positive effect on house prices in Sweden is population density, where we can observe more prominent direct effects than indirect ones. Especially for the cross-sectional point of view, the estimates are close to zero; our explanation is that it takes longer for the changes in population to be reflected in the prices and is not such an influencer within the given year.

Housing supply and unemployment affect the housing prices negatively, which is in line with the economic theory. We control for population changes, thus *ceteris paribus* an increase in supply leads to a decrease in price. Interestingly, indirect effects are stronger than the direct ones suggesting that increasing supply in the municipality itself is not sufficient. There has to be an increase also in the neighbouring regions to have an impact on prices. There is an upward jump in magnitude of the coefficient after the crisis implying that demand was higher than supply. Therefore increase in supply had milder effect on the prices compared to pre-crisis years.

The relatively high coefficients (in absolute terms) of the unemployment rate during the financial crisis were only temporal and the estimates are converging back to the pre-crisis levels. As for the unemployment, the indirect effects are dominant. This may be explained partly due to the fact that we also include income in the regression. This coefficient captures the effect of changes in unemployment having real income fixed, thus funds available remain the same. Only with unemployment increasing in more regions, we are able to see drop in prices, perhaps because expected wage increases are reduced.

Plotting the spillover effects, we see that while the total effects are higher (in absolute terms) for real income and population density, the spillover is more pronounced for changes in housing supply and unemployment. There is no visible difference for the patterns of different municipalities.

We believe that the analysis could yield more precise estimates if there was data available for construction cost and credit to household on municipality level. Suggestion for further research is to focus more closely on spillover effects and interactions among the municipalities.

Bibliography

Agnello, L. & Schuknecht, L. (2009), Booms and busts in housing markets: determinants and implications, Working Paper Series 1071, European Central Bank.
URL: *https://ideas.repec.org/p/ecb/ecbwps/20091071.html*

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Springer, Dordrecht.

Anselin, L., Gallo, J. L. & Jayet, H. (2008), *Spatial Panel Econometrics*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 625–660.

Baltagi, B. H. (2008), *Econometric analysis of panel data*, John Wiley & Sons, Chichester, UK.

BIS (2018), 'Credit-to-gdp gaps', [Online] Available from: https://www.bis.org/statistics/c_gaps.htm. [Accesed on 3.4.2018].

Borowiecki, K. J. (2009), 'The Determinants of House Prices and Construction: An Empirical Investigation of the Swiss Housing Economy', *International Real Estate Review* **12**(3), 193–220.

Davidova, L. (2015), Various estimation techniques of the gravity model of trade, Master's thesis, Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies.
URL: *https://is.cuni.cz/webapps/zzp/download/120177285/?lang=cs*

Davis, E. P. & Zhu, H. (2011), 'Bank lending and commercial property cycles: Some cross-country evidence', *Journal of International Money and Finance* **30**(1), 1–21.

Driscoll, J. C. & Kraay, A. C. (1998), 'Consistent covariance matrix estimation with spatially dependent panel data', *The Review of Economics and Statistics* **80**(4), 549–560.

Dubin, R. (1998), 'Spatial autocorrelation: A primer', *Journal of Housing Economics* **7**(4), 304–327.

Elhorst, J. P. (2014), *Spatial Econometrics From Cross-Sectional Data to Spatial Panels*, Springer.

Fidrmuc, J. (2009), 'Gravity models in integrated panels', *Empirical Economics* **37**(2), 435–446.

Finansinspektionen (2017), 'Fi analysis 9: Households' interest rate adjustment periods – an economic vulnerability?', [Online] Available from: https://www.fi.se/en/published/reports/fi-analysis/2017/fi-analysis-9-households-interest-rate-adjustment-periods--an-economic-vulnerability/. [Accesed on 1.4.2018].

Hoechle, D. (2007), 'Robust standard errors for panel regressions with cross-sectional dependence', *Stata Journal* **7**(3), 281–312.

Holly, S. & Jones, N. (1997), 'House prices since the 1940s: Cointegration, demography and asymmetries', *Economic Modelling* **14**(4), 549–565.
URL: *https://ideas.repec.org/a/eee/ecmode/v14y1997i4p549-565.html*

Holly, S., Pesaran, M. H. & Yamagata, T. (2010), 'A spatio-temporal model of house prices in the USA', *Journal of Econometrics* **158**(1), 160–173.

Hsiao, C. (2005), 'Why Panel Data?', *The Singapore Economic Review (SER)* **50**(02), 143–154.
URL: *https://ideas.repec.org/a/wsi/serxxx/v50y2005i02ns0217590805001937.html*

Julian Livingston-Booth, Jonathan Kownator, T. G.-O. C. (2018*a*), Castellum (cast.st) weaker tailwinds; downgrade to neutral, Technical report, Goldman Sachs.

Julian Livingston-Booth, Jonathan Kownator, T. G.-O. C. (2018*b*), Sweden: Downside risks dominate, Technical report, Goldman Sachs.

Kim, C. W., Phipps, T. T. & Anselin, L. (1998), Measuring The Benefits Of Air Quality Improvement: A Spatial Hedonic Approach, 1998 Annual meeting, August 2-5, Salt Lake City, UT 20959, American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association).

Las, O. (2018), 'Nordic outlook economic and financial trends', [Online] Available from: https://danskebank.com/-/media/danske-bank-com/pdf/news-insights/files-in-insights-pieces/2018/nordicoutlook_270318-.-la=en.pdf. [Accesed on 20.4.2018].

Lesage, J. P. (2008), 'An Introduction to Spatial Econometrics', *Revue d'économie industrielle* **0**(3), 19–44.

LeSage, J. P. & Pace, R. K. (2009), *An Introduction to Spatial Econometrics*, Taylor & Francis Group, LLC.

Liao, W.-C. & Wang, X. (2012), 'Hedonic house prices and spatial quantile regression', *Journal of Housing Economics* **21**(1), 16–27.

Michael D. Ward, K. S. G. (2008), *Spatial Regression Models*, SAGE Publications.

Millo, G. (2017), 'A simple randomization test for spatial correlation in the presence of common factors and serial correlation', *Regional Science and Urban Economics* **66**(C), 28–38.

Pesaran, M. (2004), 'General Diagnostic Tests for Cross Section Dependence in Panels', Cambridge Working Papers in Economics 0435, Faculty of Economics, University of Cambridge.

Qu, X. & Lee, L.-f. (2015), 'Estimating a spatial autoregressive model with an endogenous spatial weight matrix', *Journal of Econometrics* **184**(2), 209–232.

Riksbank, S. (2018), 'Search interest & exchange rates', [Online] Available from: https://www.riksbank.se/en-gb/statistics/search-interest--exchange-rates/?g2-SECBREPOEFF=on&from=13%2F03%2F2010&to=13%2F02%2F2018&f=Month&c=cAverage&s=Comma. [Accesed on 20.4.2018].

Roger S. Bivand, Edzer Pebesma, V. G.-R. (2008), *Applied Spatial Data Analysis with R*, Springer-Verlag New York.

SCB (2018), 'Expenditures per household, 2003–2012 in sek', [Online] Available from: https : / / www . scb . se / en / finding-statistics / statistics-by-subject-area / household-finances / household-expenditures / household-budget-survey-hbs / pong / tables-and-graphs / time-series-20032012 / expenditures-per-household-20032009-in-sek/. [Accesed on 20.4.2018].

Tsatsaronis, K. & Zhu, H. (2004), 'What drives housing price dynamics: cross-country evidence', *BIS Quarterly Review* .
URL: *https://ideas.repec.org/a/bis/bisqtr/0403f.html*

Vogiazas, S. & Alexiou, C. (2017), 'Determinants of Housing Prices and Bubble Detection: Evidence from Seven Advanced Economies', *Atlantic Economic Journal* **45**(1), 119–131.

Wilhelmsson, M. (2008), 'Regional house prices: An application of a two-equation model to the Swedish housing market', *International Journal of Housing Markets and Analysis* **1**(1), 33–51.

Zheng, X. & Hui, E. (2016), 'Does liquidity affect housing market performance? an empirical study with spatial panel approach', *Land Use Policy* **56**, 189 – 196.

Appendix A

Additional tables and graphs which are referred in the text can be found here.

A.1 Possibility of getting biased estimates

LeSage & Pace (2009) shows that by omitting spatial relations in the model, one risks getting biased and inefficient estimates. Biased when the spatially lagged independent or dependent variable influence the dependent variable, inefficient when the error term exhibits spatial autocorrelation. See illustrative extract from LeSage & Pace (2009):

We consider a dependent variable y, that is fully explained by the two uncorrelated variables x and z, with $N(0, I_n)$ distribution.

$$y = x\beta + z\theta \tag{A.1}$$

Where z is unobservable and *i.i.d.*, hence a disturbance term we redefine as ϵ in equation (A.2).

$$y = x\beta + \epsilon \tag{A.2}$$

In the case where the x is uncorrelated with ϵ, we would get an unbiased estimate of β using ordinary OLS. Looking at a situation where z is still uncorrelated with x but is spatially autocorrelated as in equation (A.3), where r is a vector of disturbances with mean zero and variance equal to σ_r^2, W is an $n \times n$ row stochastic (the rows sum to zero) weight matrix with zeros on the diagonal and positive values for neighbour elements. This could be the case where the unobservable variable could contain things such as quality of public transportation that are connected with neighbouring entities.

$$z = \rho W z + r \tag{A.3}$$

$$z = (I_n - \rho W z)^{-1} r \tag{A.4}$$

Introducing (A.4) into (A.1), and substituting θr with u yields equation (A.6). In this situation β estimation remains unbiased, but inefficient with θ increasing the variance.

$$y = x\beta + (I_n - \rho W z)^{-1}(\theta r) \tag{A.5}$$

$$y = x\beta + (I_n - \rho W z)^{-1} u \tag{A.6}$$

$$E(y) = x\beta \tag{A.7}$$

It is however not unlikely that x is correlated with u. Eg when population density, contained in x, is correlated with quality of public transportation. We can model this as u being linearly dependent on x together with a disturbance term. Where γ and the variance of the disturbance term determines how strong the relationship is.

$$u = x\gamma + v \tag{A.8}$$

$$v \sim N(0, \sigma_v^2 I_n \tag{A.9}$$

Then the data generating process would be the one represented in equation (A.11).

$$y = x\beta + (I_n - \rho W z)^{-1}(x\gamma + v) \tag{A.10}$$

$$y = x\beta + (I_n - \rho W z)^{-1}(x\gamma) + (I_n - \rho W z)^{-1} v \tag{A.11}$$

In this scenario ordinary OLS will not give an unbiased estimate of β.

$$(I_n - \rho W z)y = (I_n - \rho W z)x\beta + x\gamma + v$$

$$y = \rho W y + x(\beta + \gamma) + W x(-\rho\beta) + v$$

A.2 Equivalence of the effects

To show how the average total and indirect effects to and from an entity are numerically equivalent we redefine equation (A.12), presented in section 2.1, as (A.13), where the new term is defined as in equation (A.14). (A.14) can be expanded as in equation (A.15) (excluding $\mathbf{R}$). Here, the average direct effect is the average of $\mathbf{S}_k(\mathbf{W})_{ii}$, and the average total effect to an observation is the average of all the averaged row sums of $\mathbf{S}_k(\mathbf{W})$, and the average total effect from one observation to the other observations is the average of all the averaged column sums.

$$\mathbf{Y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta}) + \mathbf{R} \tag{A.12}$$

$$\mathbf{Y} = \sum_{k=1}^{r} \mathbf{S}_k(\mathbf{W})\mathbf{X}_k + \mathbf{R} \tag{A.13}$$

$$\mathbf{S}_k(\mathbf{W}) = (\mathbf{I} - \rho\mathbf{W})^{-1}(\boldsymbol{\beta}\mathbf{W}\boldsymbol{\theta}_k) \tag{A.14}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{bmatrix} = \sum_{k=1}^{r} \begin{bmatrix} S_k(W)_{11} & S_k(W)_{12} & \cdot & S_k(W)_{1n} \\ S_k(W)_{21} & S_k(W)_{22} & \cdot & S_k(W)_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ S_k(W)_{n1} & S_k(W)_{n2} & \cdot & S_k(W)_{nn} \end{bmatrix} \begin{bmatrix} x_{1k} \\ x_{2k} \\ \cdot \\ x_{nk} \end{bmatrix} \tag{A.15}$$

A.3 Stationarity

Here, we provide summary of unit root tests. Variables *cost_h*, *cpi*, and *credit_r* are reported on the country level only, thus are the same for the all municipalities. For such reason, we run ADF test for these variables. Assessing non-stationary behaviour has to be done also by graphical means, because time-series dimension is rather small $T = 14$ and the tests have low power Baltagi (2008).

Pesaran's CADF Unit-Root Test		
Variable	p-value before differencing	p-value after differencing
m_house_price	0.999	0.000
unemp	1.000	0.089
pop_den	1.000	0.000
income	1.000	0.005
stock_h/pop	1.000	0.000

Table A.1: PADF Unit Root Test

Even though p-values for the common variables lowered substantially, they failed to reach at least 10% significance level. Nonetheless, we take it as sufficient - knowing that for small T - the test has low power. Also, the graphs suggest that the data

ADF Test for common variables		
Variable	p-value before differencing	p-value after differencing
cost_h	0.987	0.193
cpi	0.945	0.283
credit_r	0.984	0.199

Table A.2: ADF Unit Root Test

are rather stationary - see A.1. Then summary for cointegration test results follows - see A.3.

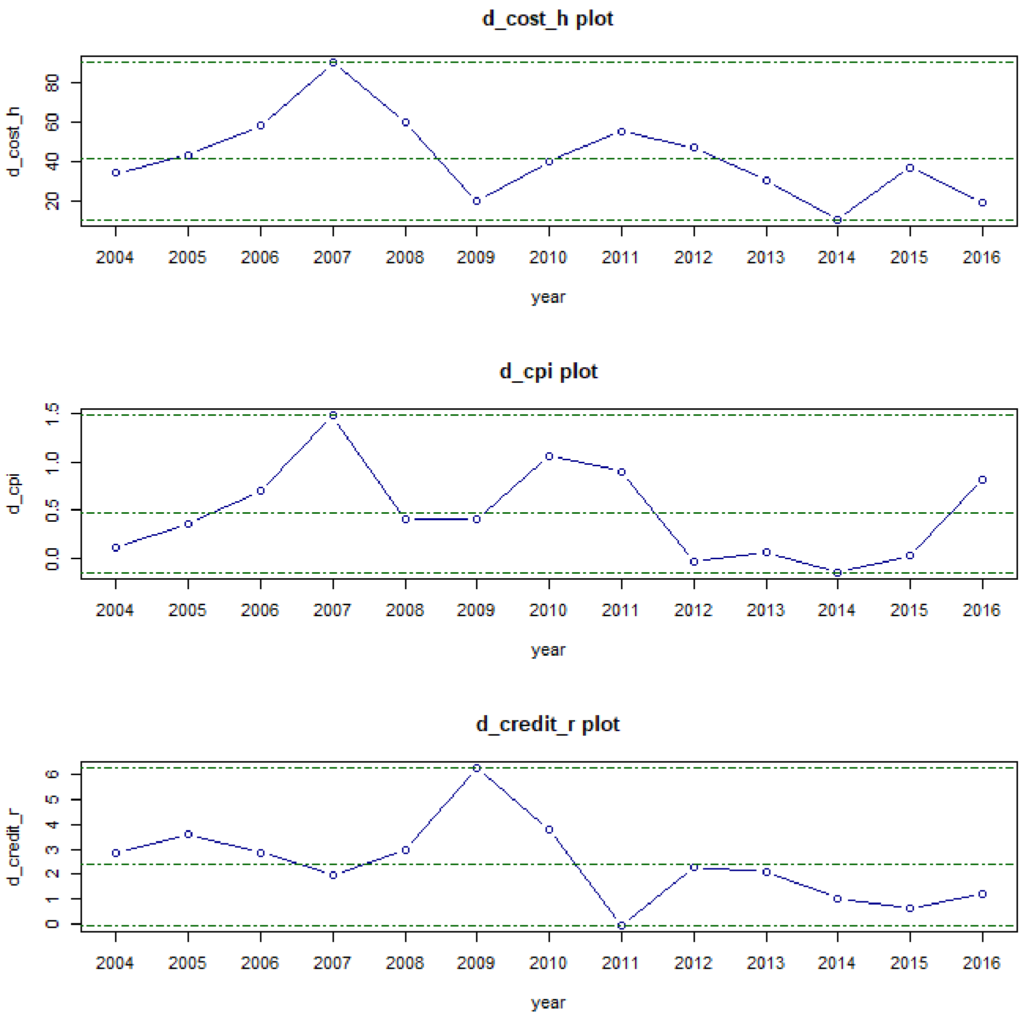


Figure A.1: Common variables development

Cointegration tests		
Kao test		
H_0: No cointegration		
H_A: All panels are cointegrated		
	Statistics	p-value
Modified Dickey-Fuller t	7.092	0.000
Dickey-Fuller t	10.133	0.000
Augmented Dickey-Fuller t	11.481	0.000
Unadjusted modified Dickey-Fuller t	1.149	0.125
Unadjusted Dickey-Fuller t	4.292	0.000
Pedroni test		
H_0: No cointegration		
H_A: All panels are cointegrated		
	Statistics	p-value
Modified Phillips-Perron t	25.902	0.000
Phillips-Perron t	-51.109	0.000
Augmented Dickey-Fuller t	-42.249	0.000
Westerlund test		
H_0: No cointegration		
H_A: Some panels are cointegrated		
	Statistics	p-value
Variance ratio	8.040	0.000

Table A.3: Cointegration tests

A.4 VIF analysis

VIF analysis					
log(cost_h)	log(unemp)	log(pop_den)	log(income/cpi)	log(stock_h/pop)	log(credit_r)
39.947	2.074	2.542	2.823	2.095	42.617
log(cost_h)	log(unemp)	log(pop_den)	log(income/cpi)	log(stock_h/pop)	dl_credit_r
2.328	1.626	2.536	2.695	2.064	1.713

Table A.4: VIF analysis