LUND UNIVERSITY
School of Economics and Management

Master Thesis, Spring 2009
School of Economics and Management
Lund University

“Gimme Shelter”

A Long- and Short-Run Analysis of Swedish Regional House Prices

Authors:
Sofia Bill
Malin Ivarsson

Supervisors:
Klas Fregert
Fredrik NG Andersson

“An economist is a man who states the obvious
in terms of the incomprehensible.”

Alfred A. Knopf

Abstract

The aim of this thesis is to examine the long-run relationship between regional house prices, how the house price model behaves in the short run and if aggregated national house price models may be misleading. To investigate whether house prices are related in the long run we use cointegration tests to look for common trends. The short-run behaviour of the model is estimated by an error correction model. To test the hypothesis that aggregated models are misleading, we compare the regional models and also use the same procedure to estimate the house price model with panel data. The results show that there are several long-run relationships between house prices of different regions. In the short run, CPI seems to have an important impact on house price changes in all regions whereas GDP and unemployment mainly influences house price movements in the metropolitan areas and the southern half of Sweden. The results from the panel regression lead to similar conclusions regarding the independent variables. Although the results are similar regarding the statistical significance of the variables, we cannot say anything regarding differences in economic significance between regions when using panel data and an aggregated model may thus be misleading.

Table of Contents

List of Figures and Tables

1 Introduction

This chapter aims to introduce the purpose of the thesis and its delimitations. A short summary of the employed methodology and the results is also included together with a description of the outline of the thesis.

1.1 Aim and Purpose

The current state of the world economy once again confirms how important the role of the housing market is. Changes in the housing finance systems over the past decades have often lead to easier access on the mortgage markets, enabling even less suitable borrowers to finance a home ownership and increase their consumption. These so called subprime loans have made a large contribution to the crisis in the US of today. Rising house prices allow homeowners to increase their household borrowing, using their home as collateral. In due time the price bubble is bound to burst and some borrowers suddenly face negative equity, i.e. the cost of the mortgage exceeds the value of the house used as collateral. Another interesting macroeconomic aspect of the housing market is that housing wealth forms a large part of private wealth and that house prices therefore may have a large impact on the household’s consumption and saving behaviour. This is why studies of the housing market’s interconnections with the macro economy are so frequent and in demand by policy makers. Local factors play an important role in determining supply and demand on the housing market which induces substantial regional differences.

The purpose of this thesis is to model regional house prices in Sweden and to look for long-run relationships in, firstly, a panel of the house price indices of eight different regions and regions divided into smaller groups. Secondly the different regional macroeconomic variables that form the house price model are tested for long-run relationships and thirdly the same tests are performed on a panel of macroeconomic variables. The housing price model can then be estimated using both regional error correction models (ECM) and a panel ECM. Our contribution to the already existing studies is the panel cointegration approach with the use of regional data instead of aggregated data on the national level.

The questions to be answered are: Is there a long-run relationship between regional house prices and are there long-run relationships between house prices in different groupings of regions? How do the components of the house price model behave in the short run and how does it respond to periods of disequilibrium?

1.3 Delimitations

To increase the number of observations, quarterly data is employed. This in turn imposes some restrictions on the availability of especially regional data. Whereas the interest rate is the same for all regions, measures of regional disposable income or GRP (gross regional product) are only available as yearly observations. Production levels are therefore measured by the aggregated GDP which naturally becomes the same for all regions. The lack of quarterly data leads to the only regional specific variables being house prices and unemployment, where unemployment is used as an indicator of regional production and uncertainty. Sweden is divided into eight regions (riksområden) which is a common used classification.

1.4 Disposition

The thesis is divided into three parts, where the first part is a cointegration analysis of the regional house price data. The aim of the first part is to investigate if there is a long-run common trend between the different regions' house prices. The second part investigates how the different variables affect the house prices in the different regions. Do the regions react differently and do they have differences in their speed of adjustment. Regional data is used as aggregated house prices models may overlook important regional factors and differences. The final part performs an ECM on panel data, testing for which variables that affect the house prices.

The first chapter is an introduction to the thesis that states its aim and purpose. The second chapter depicts the background of the thesis and investigates previous research. The model used is presented in chapter 3 along with the theory and the methodology. The fourth chapter describes the data. In chapter 5 and 6 the results are presented, followed by a discussion in chapter 7. The last chapter is a conclusion of the thesis.

2 Background and Previous Research

This section presents essential information to understand the importance of investigating the housing market and the contemporaneity of the subject together with part of the research previously conducted. The chapter also includes our hypothesis about regional housing markets.

2.1 Background

The housing market plays an important role in the economy and the importance keeps increasing. Over the last two decades the housing financing system has experienced a lot of innovation and changes. The dynamics of the housing sector depends a lot on the local factors, which determine the supply and demand, and the structure of the mortgage market. The Swedish housing financial market experienced a deregulation in the mid 1980's (IMF, 2008) and the housing market experienced a price boom in the late eighties, where the prices first increased substantially and then decreased by about the same amount some years later (ECB, 2000). A tax reform (91TR) was implemented in Sweden in 1991 with one of the aims being to decrease the distortion on the housing market. The tax reform meant that the deduction of interest payments was decreased, which resulted in higher user cost for private home owners (Barot & Yang, 2002). In recent years both the mortgage debts and the role of houses as collateral in Sweden have increased (Statistics Sweden, 2009). The mortgage market has experienced increased competiveness and flexibility which has made the Swedish mortgage market one of the most flexible and “complete” with easy access to mortgage credit.

These changes have led to increased spill-over effects from the housing market to the rest of the economy mainly due to the positive effect on consumption via increased household debt leverage (IMF, 2008). When houses are used as collaterals, an increase in house prices also increases the value of the collateral, which in turns leads to an increased consumption. When combining this with an expectation of higher future income, the effect gets extra strong as it allows the households to borrow against a higher expected income. Due to all these changes the role of the house sector on business cycles have also altered and become increasingly important. This has also led to an increased effect of monetary policy on house prices and thereby on consumption and output. The house prices have had a tendency to be closely correlated with the economic cycle, but with a lag of six quarters for Sweden (IMF, 2008). The increased role of the housing sector was for instance showed by the enlarged contribution to the downturn of GDP by the residential investment compared to previous decades. The modification of the housing financing system has also led to an increased sensitivity and exposure to shocks from the housing sector. It is not only the role of the housing sector that has changed, the housing cycles have also altered their appearance by lasting longer and being stronger.

As mentioned the house prices have an increasingly important effect on consumption. The previous research on the Swedish housing market mainly investigates how the housing market reacts to different shocks and how that in terms may affect the consumption. ECB (2000) conducted a study on which the main factors of fluctuations in the house prices are and concluded that monetary policy shocks have a great impact on house prices. House prices where also found to be more sensitive to changes in expectations than consumer prices are.

Housing markets have many special features resulting in that the structure of the housing market in some senses best is described as a collection of a lot of smaller regional markets. Cointegration studies between different regions have mainly been performed in Britain, where the researchers have tested for long-run relationships in house price changes between different regions on time series data, the so called “ripple effect” or “ripple down effect”. The “ripple effect” implies that house prices in one region are pulled up by the increase in house prices in some other region, the prominent region, and then pulled down by the decrease in prices in the same other region (MacDonald and Taylor 1993). In Britain the south-east region is said to be the prominent region of the ripple effect. As the ratio between the price in one region to the price of the prominent region is stationary it displays no long-run trend (Meen, 1999).

These studies can be implemented on the Swedish housing market in an interesting way. As shown by Barot and Yang (2002), there are many similarities between the housing markets in the two countries. In the British cointegration studies augmented Dickey Fuller tests (ADF test) are used to conclude whether there is any upward or downward trend in the series. The ADF tests search for unit roots and thereby tests for stationarity. If some stationary linear combination is found, the series are cointegrated. If house prices are non stationary and not cointegrated in two regions, the difference in house prices between the two regions will not return to an equilibrium level, it will increase over time, so there is no long-run stable pattern in the prices.

2.2 Previous Research

As mentioned in the background section a lot of the existing research on regional trends in the house prices has been performed in Britain due to the existence of the hypothesis of the “ripple effect”. MacDonald and Taylor (1993) analysed both the long-run and the short-run properties of regional house prices by testing for regional cointegration between house prices in Britain. The cointegration test showed that some interrelationships exist, indicating that there may be some segmentation of housing markets in Britain. When testing for the number of cointegrating vectors between the different regions, the Johansen maximum likelihood approach was used and up to nine cointegrating vectors was found, which increased the empirical support for earlier findings of a long-run interregional relationship in British residential house prices. In order to see how the different regions interact with each other, impulse response functions were computed and the results were in line with the “ripple effect”. As the “black box” technique was used, meaning that the regional interrelationships cannot be explained, the result cannot distinguish what caused the effect (MacDonald and Taylor, 1993). According to MacDonald and Taylor (1993) the result may have its cause either by arbitrage through house owners selling their houses in high price areas and moving to low price areas, or because there were some regional element to the business cycle in the data sample. MacDonald and Taylor can only conclude that there exists a “ripple effect”, not explain it.

Ashworth and Parker (1997) analysed the determinants of house prices in different regions in Britain by using the Johansen maximum likelihood (1991) technique, the same as MacDonald and Taylor (1993) but improved, on regional house price models. They investigate both the long-run structure of house prices and the short-run adjustments made by agents in the different regions. Ashworth and Parker used the equation derived by Drake[1] which states that the private sector house price in period t is determined by household income, an opportunity cost variable, which is the real interest rate defined as the mortgage rate minus the house price inflation, and personal sector housing starts, all in the same time period. When using the Johansen maximum likelihood technique, the test showed that there was at least one cointegrating vector for all regions except two. The results for the vectors were all consistent with the predictions of the model, that income per capita had a positive effect on house prices and the opportunity cost and housing starts had negative effects.

---

[1] The article that Ashworth and Parker (1997) refers to is ”Modelling UK House Prices Using Cointegration: An Application of the Johansen Technique” by Drake, L.M (1993)

Both substantial similarities and differences were found between the regions. When comparing the results for each region to the ones for the whole nation, the results were mixed. For the regions that have a prominent role in the British house market the difference was insignificant, as expected, but some other regions also experience insignificant results which could not be explained. Some differences were however found to be significant, indicating that the findings for the aggregated data on Britain could be misleading (Ashworth and Parker, 1997). For some regions the housing starts have a large impact on house prices while the effect is almost insignificant for other regions and the aggregated data shows that house starts are only significant at a ten per cent level. These differences support the earlier conclusions that when only looking at the aggregated data it may not consider the regional differences. There were considerable effects of the interest rate on house prices in the long run. Other variables that may affect the housing market in different ways regionally were changes in housing construction guidelines and changes in per capita income. Ashworth and Parker (1997) explain the lack of long-run relationships with the possibility of structural differences between the regions.

Ashworth and Parker (1997) use the Granger's (1983) Representation Theorem, which states that a short-run error correction mechanism is included in every long-run cointegration relationship, to further investigate the housing markets. The error correction mechanism describes how agents in the different regions respond to short-run changes in the economic environment. The results showed that there is a substantial similarity between different regions in the agents' short-run behaviour such as the negative effect of opportunity costs on house prices and positive secular growth. The test also showed that the adjustments conducted by the agents are in general small, suggesting that the return to equilibrium may be concealed by the short-run changes in economic conditions. According to Ashworth and Parker (1997) their results demonstrate some automatic correlations in the behaviour of the house prices. In all short-run analyses, housing starts are not included in the regressions as they were not found to be significant and as they are argued to have long-run supply effects. When comparing the short-run regional results with the aggregated, many similarities were found. Income had a positive effect on house prices, opportunity cost a negative and housing starts are not included in any of the short-run analyses. The error correction term had a low speed of adjustment, but that could be explained by the large cost of adjusting the house stock due to transaction and search costs. When performing spatial correlation tests, which show to what extent regions are linked to each other, the results contradicted the “ripple effect” and showed that the adjustment appeared at the same time in all the regions. Ashworth and Parker (1997) speculate that the disappearance of the “ripple effect” is due to their assumptions considering economic theory.

Meen (1999) proposes a regional house price ratio model which captures not only income differences and spatial lags but also structural differences between the regions, by testing for coefficient differences. An equation is estimated were house prices are a function of nominal interest rate, real income, which is measured by consumers’ expenditures, unemployment rate, which is a measurement of risk, and an error term. An ADF test is used to test for the possible consistency of a long-run ratio and the results show that there may have been a weak trend in relative prices. House prices move proportionately in the long run and in common for all regions is that they are sensitive to income changes. Income changes are found to have a smaller negative effect on house prices in the northern regions than in the southern regions of Britain and the national average. The same holds for the interest rate which also has a smaller short-run effect in the north. According to Meen (1999) the south would be more sensitive to changes in interest rates as there is higher leverage in the south than in the north, which also should explain the higher speed of adjustment and their higher sensitivity to unemployment.

Meen (1999) argues that even though there may be some variations in the short run in the difference between the prices in different regions, in the long run the relative prices return to their normal state. This is also shown when Meen (1999) tests how the different regions react to shocks in the national variables. Meen (1999), as Ashworth and Parker (1997), concludes that when adding economic effects to the analysis of the “ripple effect”, the result is not definitive. This is in contrast to MacDonald and Taylor (1993) who argues for the existence of the “ripple effect”, but cannot conclude what the cause of it is. They claim that the effect arises due to arbitrage or some regional components to the business cycle in the data (MacDonald and Taylor, 1993). According to Meen (1999) Britain’s housing market is best described as a collection of regional markets, which all have some structural differences which in turn affect national housing market models. These possible structural differences are another motivation of why distinguishing between the regional markets, instead of just looking at the aggregated models, is important. In contrast to Ashworth and Parker (1997), Meen (1999) claims that the “ripple effect” does not only occur because of spatial correlation, but that the same effect can be obtained by coefficient heterogeneity. The heterogeneity measures structural differences between regions such as differences in behaviour. Meen (1999) argues that changes in regional house prices can be decomposed into three smaller parts, common movements in all regions, movements due to different economic growth in different regions and structural differences in the regional markets. According to Meen (1999) there is spatial correlation between the coefficients causing changes in the housing market that we interpret as the “ripple effect”, when the true cause lies in the changes within the regions and not between different regions.

Hort (1998) investigates the determinants of urban real house price fluctuations in Sweden between the years 1968-1994. She estimated a restricted ECM of real house price changes on Swedish panel data. Changes in income, construction cost and user cost where found to affect the real house prices in the long run and also included a negative deterministic trend as a proxy for factors not included in the model. According to Hort, 80 % of the fluctuations in real house price changes can be explained by her short-run model and that periods of disequilibrium could explain the short-term prices.

Barot and Yang (2002) compare the Swedish and the British housing markets using an ECM by estimating housing demand and investment supply. They found that change in income, debt, real interest rate and Tobin’s Q affect the house prices in both Sweden and Britain, but that the change in debt affects the British house prices more than the Swedish and that the opposite holds for the interest rates. When testing for Granger causality, changes in financial wealth affected the house prices in Sweden. The house prices also have an impact on some of the variables in both countries, such as financial wealth, debt, real interest rate and Tobin’s Q in Sweden and on income, debt and real interest rate in Britain. When departing from the equilibrium level, twelve per cent of the shock is adjusted within a year in Sweden and twenty-three per cent in Britain. In both Sweden and Britain demographic changes are found to have strong effects on house prices (Barot and Yang, 2002).

2.3 Hypothesis

The authors regarded in the previous research section have proved that there are substantial regional differences in Britain. Whether these are a result of the “ripple effect” or structural differences between the regions may however be a subject for further discussion. As mentioned by Barot and Yang (2002) the house market in Britain and Sweden experience some similarities in the regional differences in the countries’ house prices. We expect that the eight Swedish regions will be cointegrated with each other, i.e. that there is a long-run relationship between the house prices in the regions. This hypothesis will be tested in section 5.1.2.1. Between the regions that in some way resemble each other a long-run relationship in house prices is expected. The resemblance could be due to geographical proximity or the existence of metropolitan areas in the region. Which features that are most important for the resemblance in the markets are tested in section 5.1.2.2 and in 5.1.2.3. Of the eight different markets in Sweden we expect that the most northern region and the regions containing a metropolitan area would differ most from the average housing market, in section 5.1.2.4 a test of this hypothesis will be performed.

The national housing market is often best described as a collection of many small markets (Meen, 1999). We use the classification of the regions as the definition of the smaller markets that the national market consists of. These markets may be similar in some cases and diverge from each other in other cases. The short-run behaviour of the housing markets in the metropolitan areas could be expected to be different compared to the markets in regions without any special features such as a large city or a sparsely population. As Meen (1999) concluded, changes in the interest rate have smaller effects in the northern regions than in the south which could be a result of lower leverage in the northern regions. This would also make the northern regions less sensitive to other variables such as unemployment. One may suggest that the regions in Sweden may show a similar pattern. As prices are lower in the northern regions in Sweden than in the south and in the metropolitan areas, households in the northern regions borrow less and are thereby less sensitive to changes in interest rates and unemployment. A decrease in income would have a negative affect in all regions; the effect might be less negative in the northern regions than in the south and metropolitan areas. This hypothesis will be tested in section 6.1.3.

In general, our hypothesis is that there will be a difference between the regional housing markets and studies of the aggregated national market may thus be misleading, since it does not account for regional features of the housing markets.

3 Theory and Methodology

The aim of this section is mainly to explain the methods used to test our stated hypotheses. The econometric theory behind the performed tests and the model estimations is described in detail. Since our theory and methodology are converging, they are included in the same chapter.

3.1 Modelling House Prices

The existing work on the housing market is immense as well as the number of models trying to explain house prices. That some measure of income and interest rates should form the basis of the model is more or less agreed upon unanimously. Higher income affects demand and in turn has a positive effect on house prices, whereas higher interest rates raise the cost of borrowing and thus decrease the demand for housing which affect prices negatively. One question, of which the answer is not as obvious as with income and interest rates, is whether housing stock should be included or not when wanting to model the housing market. Looking to the theories of Sørensen and Whitta-Jacobsen (2005, pp. 454-455), housing stock only really matters in the short run. An increase in the housing stock with demand held constant will lower the price until, in the long run, demand has had time to adjust. Since we are interested in the long-run relationships of house prices and macro variables, we have decided to not take the housing stock into account on the basis of the arguments presented by Sørensen and Whitta-Jacobsen (2005).

The house price model in this thesis is inspired by the work of Meen (1999) described in chapter 2. His test of stationarity in the ratios between regional house prices corresponds to testing for cointegration. The model developed in his article explains house prices as a function of consumers' expenditures, unemployment and nominal interest rates. Since the access to regional quarterly data is limited, we have substituted regional consumers' expenditures by aggregate measures of GDP and instead use unemployment rates as an approximation of regional production fluctuations. Nominal interest rates have been specified as the short- and long-term interest rates (see data section for more details). The reason for including both the short- and the long-term interest rates is to examine if the different maturities might affect house prices in different ways. As opposed to Meen, we will examine house prices both regionally *and* in the form of panel data. The initial model on which we base our studies could be described as:

$$p_t^H = f(GDP_t, CPI_t, U_t, r_t^s, r_t^l) \tag{3.1}$$

where p_t^H denotes the regional house price indices, GDP_t is the Swedish gross national product, CPI_t denotes the consumer price index, U_t indicates regional unemployment and r_t^s and r_t^l are the short- and long term interest rates. All variables except the interest rates are logarithmized. The log-linear specification means that the slope coefficients generated from estimating the model should be interpreted as elasticities. The coefficients of the interest rates are semi-elasticities[2].

To investigate the long-run relationships between regional house prices, we will perform a series of cointegration tests after having concluded that the variables are non-stationary through unit root testing. We will then move on to model the short-run behaviour of house prices by forming regional error correction models (ECM) with selected independent variables. This allows us to compare the short-run behaviour of house prices between regions. We will also estimate a panel ECM with the motivation that more observations may yield sharper results with increased efficiency. With panel data we will however not be able to draw any conclusions about possible differences between regions. The statistical programme used for testing and estimating the data is EViews version 6.

We will proceed as follows:

1) Testing for unit root in regional house prices and for cointegration between house prices in a panel consisting of eight regions and in smaller groupings.
2) Testing for unit root in the cross-sectional variables.
3) Testing for cointegration between the cross-sectional variables in the house price model defined in equation (3.1).
4) Estimating the regional house price models using an ECM.
5) Estimating the regional house price model with panel data using an ECM.

3.2 Theoretical Background to Methodology

3.2.1 Unit Root Testing

Since dealing with time series data always includes the risk of non-stationarity, it is important to consider this issue in order to not end up with spurious results. On the other hand, the

---

[2] That interest rates are semi-elasticities means that an absolute change in interest rates leads to a percentage change in the dependent variable, i.e. house prices.

presence of a unit root (i.e. non-stationarity) and cointegration can be of economic interest. What is done when testing for a unit root is simply that it is tested if the coefficient of the lagged value of a variable equals one. Consider for example a simple AR(1) model: $y_t = \rho y_{t-1} + u_t$. Taking first differences we get the Dickey-Fuller (DF) specification $\Delta y_t = (\rho - 1) y_{t-1} + u_t$. The hypotheses to be tested will then be:

$$\begin{aligned} &H_0 : \rho^* = (\rho - 1) = 0, \quad (i.e.\ \rho = 1) \\ &H_1 : \rho^* < 0, \quad (i.e.\ \rho < 1) \end{aligned} \tag{3.2}$$

where the rejection of the null hypothesis indicates stationarity, i.e. no unit root. For the case of higher-order autoregressive models the DF-test is extended to an augmented DF-test (ADF test) where the auxiliary regression will look as follows:

$$\Delta y_t = \rho^* y_{t-1} + \sum_{L=1}^{p_i} \rho_L \Delta y_{t-L} + u_t \tag{3.3}$$

Higher-order correlation is now controlled for by adding lags of the first differences of y_t. p_i denotes the maximum number of lags chosen and the hypotheses to be tested are the same as for the simple DF test.[3]

An alternative test has been developed by Kwiatkowski, Phillips, Schmidt and Shin (KPSS) where the null hypothesis of stationarity is tested against the alternative of a unit root. Their idea was that time series are decomposed into sums of a deterministic trend, a random walk and a stationary error term and under the null the variance of the random walk component is zero. The test statistic used to test the hypotheses is a Lagrange multiplier (LM) statistic and besides which deterministic components to include, it must be specified which spectral estimation method to apply.[4]

3.2.2 Cointegration

Cointegration between two or more non-stationary variables means that there exists a linear combination of these variables that is stationary. The implication of cointegration is the occurrence of a long-run relationship between the included variables, which in many cases is suggested by economic theory. Hence the economic interest of testing for cointegration. Testing for cointegration can be done using the two-step Engle-Granger (EG) approach. Consider the simple model:

$$y_t = \beta x_t + \varepsilon_t \tag{3.4}$$

---

[3] See for example Verbeek (2008)

[4] Ibid.

and suppose that y_t and x_t are $I(1)$ (i.e. non-stationary): y_t and x_t will be cointegrated if the residuals ε_t are $I(0)$ (i.e. stationary). If the error term is stationary around zero it means that it is mean reverting and will occasionally equal zero so that $y_t = \beta x_t$ in the long run. To test whether the residuals are stationary, the first step of the two-step EG approach is to estimate the simple model in equation (3.4) by ordinary least squares (OLS) and then save the estimated residuals and perform a unit root test on them by running the auxiliary ADF type regression:

$$\Delta\hat{\varepsilon}_t = \rho^* \hat{\varepsilon}_{t-1} + \sum_{L=1}^{p_i} \rho_L \Delta\hat{\varepsilon}_{t-L} + u_t \qquad (3.5)$$

where p_i denotes the chosen lag length. If deterministic components like a constant or a trend are to be included, they should be added to *one* of the test equations (3.4) or (3.5) and never to both. In multivariate models with more than two endogenous variables, there exists the possibility of several cointegrating relationships and the EG single equation approach may in this case be misleading and an alternative way of testing for cointegration has therefore been developed by Johansen.[5] This thesis will make use of both methods and since the Johansen procedure is rather complex, it will be explained in more detail below.

3.2.2.1 Cointegration Testing: The Johansen Approach

In the multivariate framework the possibility of more than one cointegrating relationship exists and an alternative approach to the Engle-Granger two-step procedure has been developed by Johansen who uses a maximum likelihood (ML) estimation method that allows the testing of the number of cointegrating relationships. Consider a vector autoregressive (VAR) specification with up to k lags of the vector z_t (a vector of n possibly endogenous variables):

$$z_t = A_1 z_{t-1} + \ldots + A_k z_{t-k} + u_t \qquad (3.6)$$

where z_t is $(n \times 1)$ and $u_t \sim IN(0, \Sigma)$.

Equation (3.6) can be transformed into a vector error correction model (VECM) which is a VAR model that considers the common history of a number of endogenous variables, in error correction form:

$$\Delta z_t = \Gamma_1 \Delta z_{t-1} + \ldots + \Gamma_{k-1} \Delta z_{t-k+1} + \Pi z_{t-k} + u_t \qquad (3.7)$$

---

[5] See for example Verbeek (2008)

If z_t is $I(1)$, Δz_t is $I(0)$ and this means that Πz_{t-k} must be $I(0)$ for u_t to be white noise. This holds for three occasions: (i) when all variables in z_t are stationary, (ii) when there is no cointegration so that there are no long-run relationships between the variables in z_t and (iii) the case of so called reduced rank when there are up to $n-1$ cointegration relationships r and Π can be written as $\alpha\beta'$ where α is a speed of adjustment term and β is a matrix of long-run coefficients. The cases (i) and (ii) are of little interest in this context, since (i) implies that the problem of spurious results due to autocorrelation does not exist whereas (ii) requests that estimation is done employing a model that does not involve any long-run components, e.g. a VAR.

The presence of $r \le (n-1)$ cointegrating relationships in β means that there are r columns that form r linear stationary combinations of the variables in the vector z_t and $(n-r)$ columns that form non-stationary relationships. For Πz_{t-k} to remain stationary, only the cointegration vectors may enter equation (3.7) and this in turn implies that the last $(n-r)$ columns in α are insignificant. Testing for cointegration thus means testing which columns in α that are zero or in other words how many r linearly independent columns that can be found in Π. For a given r the ML estimate for β is the same as the eigenvectors corresponding to the r largest eigenvalues of a $n \times n$ matrix. The estimated eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > ... > \hat{\lambda}_n$ can be used to test hypotheses about r (the rank of Π). If there are r cointegration relationships, $\log(1-\hat{\lambda}_i) = 0$ for the smallest $(n-r)$ eigenvalues (i.e. for $i = r+1, r+2, ..., n$):

$$\begin{aligned} &H_0 : r \le r_0 \\ &H_1 : r_0 < r \le n \end{aligned} \tag{3.8}$$

The null hypothesis is tested using the so-called trace statistic to see whether the smallest $(n-r_0)$ eigenvalues are significantly different from zero:

$$\lambda_{trace} = -T \sum_{i=r_0+1}^{n} \log(1-\hat{\lambda}_i) \tag{3.9}$$

Alternatively it can be tested whether the largest eigenvalue is significant using the maximum eigenvalue statistic:

$$\lambda_{\max} = -T \log(1-\hat{\lambda}_{r_0+1}) \tag{3.10}$$

where the null of r cointegration vectors is tested against the alternative of $r+1$ vectors. The tests are distributed under a multivariate DF distribution.[6]

3.2.3 Modelling Cointegrated Time Series

As mentioned in the beginning of section 3.2.1, estimating non-stationary time series may lead to spurious results. If this issue is not considered and one estimates a regression trying to explain y_t from x_t by OLS, one would probably end up with significant but misleading results. These so-called spurious regressions are characterized by high explanatory power (R^2) and high t-statistics indicating significant values of the slope coefficients. The reason why the OLS estimator tends to indicate significant correlation between the variables even though they are uncorrelated is due to that they are both trending. If however the variables are cointegrated and share a common trend, consistent slope coefficients can be estimated by OLS and these estimated coefficients will even be super consistent, meaning that they converge towards their true value at a faster rate than usual OLS estimators. A spurious regression will have non-stationary error terms leading to inconsistent estimators of the slope coefficients. By testing the error terms as the EG approach prescribes one could conclude if OLS may be used to generate consistent estimators.[7]

A long-run cointegrating relationship suggests that conclusions can be drawn regarding the short-run adjustment behaviour of the non-stationary variables. If there exists a long-run relationship $y_t = \beta x_t$, there must be some force that drives the equilibrium error $\varepsilon_t = y_t - \beta x_t$, i.e. the term that measures the distance away from equilibrium during periods of disequilibrium, back towards zero. This is modelled using an error correction model (ECM) where the equilibrium error is represented by the error correction term. The ECM could be said to be a continuation of the EG two-step procedure, since we need to affirm that the variables are cointegrated before we move on to estimating the model. Consider the simple model:

$$y_t = \phi_0 x_t + \phi_1 x_{t-1} + \alpha_1 y_{t-1} + u_t \tag{3.11}$$

where ϕ_0 denotes the reaction of y_t to a change in x_t in the short run. Taking first differences gives:

---

[6] See for example Harris and Sollis (2003)

[7] See for example Verbeek (2008)

$$\begin{aligned}
\Delta y_t &= \phi_0 x_t + \phi_1 x_{t-1} + (\alpha_1 - 1) y_{t-1} + u_t \\
\Delta y_t &= \phi_0 (x_t - x_{t-1}) + (\phi_0 + \phi_1) x_{t-1} - (1 - \alpha_1) y_{t-1} + u_t \\
\Delta y_t &= \phi_0 \Delta x_t - (1 - \alpha_1) \left[y_{t-1} - \left(\frac{\phi_0 + \phi_1}{1 - \alpha_1} \right) x_{t-1} \right] + u_t \\
\Rightarrow \Delta y_t &= \phi_0 \Delta x_t - \gamma (y_{t-1} - \beta x_{t-1}) + u_t
\end{aligned} \tag{3.12}$$

which represents the ECM where $\gamma = (1 - \alpha_1)$ and $\beta = \left(\frac{\phi_0 + \phi_1}{1 - \alpha_1} \right)$. If the error correction term $-\gamma(y_{t-1} - \beta x_{t-1}) = 0$ we are in equilibrium and the model becomes $\Delta y_t = \phi_0 \Delta x_t + u_t$. γ denotes the speed of adjustment towards equilibrium. Since it has already been concluded by the EG approach that we have cointegrated variables, equation (3.12) can be estimated by OLS despite the non-stationary properties of the data.[8]

3.3 Panel Data

The major advantage of working with panel data is the possibility to allow for heterogeneous individuals, in this case regions. In addition we get more variability and, since we combine cross-sectional observations we get a larger number of observations yielding more degrees of freedom, leading to greater efficiency when estimating the model.

A simple panel data model may look as follows:

$$y_{it} = X'_{it} \beta + z'_{it} \gamma + e_{it} \tag{3.13}$$

where $i = 1, ..., N$ denotes individuals and $t = 1, ..., T$ denotes time, $X'_{it} = \{x_{1t}, x_{2t}, .., x_{Nt}\}$ is a vector of N independent variables and z is a deterministic component which can take on several forms. Allowing for heterogeneity in individuals, i.e. to have separated intercepts, implies that z takes on the form of the fixed effect α_i. Another alternative would be to assume random effects which mean that α_i is homoskedastic across individuals. This assumption is however more restrictive than the fixed effects.[9]

Applying this to our house price model inspired by Meen (1999) y_{it} will denote p^H_{it} and the vector X'_{it} can be defined as:

$$X'_{it} = \{ GDP_{it}, CPI_{it}, U_{it}, r^s_{it}, r^l_{it} \} \tag{3.14}$$

---

[8] See for example Harris and Sollis (2003)

[9] See for example Verbeek (2008)

Working with panel data also gives rise to some limitations. The assumption of independent observations will be hard to fulfil, since the same individuals are observed several times, which means that OLS estimators might be less efficient. This can however be resolved by using other estimates that take correlation of the error terms into account. When used correctly, panel data sets will actually often yield more accurate estimates than examinations of individual cross-sections.[10]

3.3.1 Panel Unit Root Testing

Although there are several advantages from working with panel data, additional issues arise when testing for unit roots in panels compared to the single time series case. Assumptions about cross-sectional independence have to be made, which contradicts the expectations of finding long-run relationships when testing for cointegration. It has been shown that many of the tests perform poorly when the error terms are cross-sectionally correlated or under cross-sectional cointegration.[11]

There are several tests available for computing panel unit root tests and they can in turn be divided into two groups: tests with a common unit root process and tests with individual unit root processes.

3.3.1.1 Common Unit Root Process

Levin, Lin and Chu (2002) have developed a panel unit root test (LLC test) based on the ADF specification:

$$\Delta y_{it} = \rho^* y_{it-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + z_{it}' \gamma + u_{it} \qquad (3.15)$$

where $\rho^* = (\rho - 1)$ is common for all cross-sections, but the lag order, p_i, for the differences is allowed to vary. The hypotheses to be tested are:

$$\begin{aligned} &H_0 : \rho^* = (\rho - 1) = 0 \\ &H_1 : \rho^* < 0 \end{aligned} \qquad (3.16)$$

If the null hypothesis cannot be rejected all individuals contain a unit root and if the null can be rejected all individual series are stationary. Under the null the modified t-statistic for the estimated $\hat{\rho}^*$, adjusted for serial correlation, is asymptotically normally distributed:

---

[10] See for example Verbeek (2008)

[11] See for example Harris and Sollis (2003)

$$t_{LLC} = \frac{t_{\rho^*} - (N\tilde{T})S_N \hat{\sigma}^{-2} se(\hat{\rho})\mu_{m\tilde{T}}{}^*}{\sigma_{m\tilde{T}}{}^*} \rightarrow N(0,1) \qquad (3.17)$$

where t_{ρ^*} is the standard t-statistic for $H_0 : \rho^* = 0$, S_N is the average standard deviation ratio, $\hat{\sigma}^2$ is the estimated variance of u_{it}, $se(\hat{\rho})$ is the standard error of $\hat{\rho}$, $\mu_{m\tilde{T}}{}^*$ and $\sigma_{m\tilde{T}}{}^*$ are adjustment terms for the mean and standard deviation and:

$$\tilde{T} = T - \left(\sum_i \frac{p_i}{N} \right) - 1 \qquad (3.18)$$

Thus, inference can be performed using critical values based on the normal distribution as opposed to the first developed Levin, Lin (LL) test where new critical values adjusted for serial correlation had to be computed.

Another panel unit root test that accounts for a common unit root process is the Hadri test that is similar to the KPSS test described in short in 3.2.1. Unlike the LLC test, this test employs the null hypothesis of no unit root in any of the series against the alternative of a common unit root in the panel data. The test is based on the residuals generated from regressing y_{it} on a constant or a constant and a trend and estimating the equation by OLS:

$$y_{it} = \delta_i + \eta_i t + \varepsilon_{it} \qquad (3.19)$$

The estimated residuals $\hat{\varepsilon}_{it}$ are then used to form a Lagrange multiplier (LM) statistic that is standard normally distributed.[12]

3.3.1.2 Individual Unit Root Process

In common for the different tests in this category is that they all relax the homogeneity constraint imposed in the tests for a common unit root process by allowing for ρ to vary across the individual series. The Im, Pesaran and Shin (IPS) test is based on ADF specifications that are separate for each cross-section (as with the LLC test the lag length is allowed to vary between individuals):

$$\Delta y_{it} = \rho_i^* y_{it-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + z_{it}' \gamma + u_{it} \qquad (3.20)$$

with the following hypotheses:

$$\begin{aligned} &H_0 : \rho_i^* = (\rho_i - 1) = 0 \quad \forall i \\ &H_1 : \rho_i^* < 0 \quad \textit{for at least one i} \end{aligned} \qquad (3.21)$$

---

[12] See for example Harris and Sollis (2003)

The null hypothesis, if not rejected, states that each series contains a unit root for all cross-sections against the alternative that at least one of the individual series is stationary. To get the t-statistic, the IPS test averages the individual t-statistics obtained for ρ_i^* and generates a standard normally distributed t-bar statistic:

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_{\rho^*} \tag{3.22}$$

where t_{ρ^*} is the individual t-statistic for testing $H_0 : \rho_i^* = 0$ for all i. One drawback with the IPS test is that, if the null is rejected, we cannot tell which of the series that is stationary or non-stationary.[13]

An alternative approach to test for individual unit root processes is the Fisher-type ADF test proposed by Maddala and Wu (1999). The Fisher-ADF test combines the p-values generated from estimating an ADF test for each cross-section so that we get:

$$P = -2\sum_{i=1}^{N} \ln p_i \rightarrow \chi^2_{2N} \tag{3.23}$$

which has a χ^2-distribution with 2N degrees of freedom. The null and alternative hypotheses are the same as for the IPS test. The Fisher-ADF test thus has some similarities to the IPS test, but instead of averaging the t-statistics obtained from individual ADF tests, it averages the p-values.

3.3.2 Panel Cointegration Testing

As with the panel unit root tests, there are different types of panel cointegration tests and different complexities that might arise when they are applied. Additional to the problems already described under panel unit root testing, one needs to regard the possibility of heterogeneity in the parameters and the number of cointegrating relationships and potential cointegration between series from different cross-sections.[14]

The Pedroni test and the Kao test are based on an Engle-Granger two-step approach, i.e. they test for a unit root in the residuals of a panel cointegrating model, since when regressing non-stationary series the residuals need to be stationary in order for the series to be cointegrated.

---

[13] See for example Harris and Sollis (2003)

[14] Ibid.

The Pedroni test allows for heterogeneous intercepts and trend coefficients and is based on the following regression:

$$y_{it} = \alpha_i + \delta_t t + \beta_{1i} x_{1it} + \beta_{2i} x_{2it} + \ldots + \beta_{Ki} x_{Kit} + e_{it} \qquad (3.24)$$

from which the residuals are saved to be examined for unit roots:

$$\hat{e}_{it} = \rho_i \hat{e}_{it-1} + v_{it} \qquad (3.25)$$

The hypotheses to be tested are the null of no cointegration against the alternative that (in this example) x and y are cointegrated:

$$\begin{aligned} &H_0 : \rho_i = 1 \quad \forall i \\ &H_1 : \rho_i = \rho < 1 \quad \forall i \quad or \quad \rho_i < 1 \quad \forall i \end{aligned} \qquad (3.26)$$

There are actually two alternative hypotheses depending on whether it is a within-dimension test that pools ρ_i across the individuals so that $\rho_i = \rho$ or a between-dimension (also called group-mean approach) test which averages $\hat{\rho}_i$ for each individual allowing for heterogeneity. Since the between-dimension does not presume that ρ_i is the same for all individuals, it is less restrictive than the within-group approach. Pedroni has developed several tests with the null of no cointegration and different alternative hypotheses depending on the way serial correlation is taken into account and whether the within- or between approach is employed. When performing the Pedroni test in EViews, seven test statistics are generated according to Pedroni's initial instructions. Four of them belong to the within-group approach and three of them to the between-dimension. There are three non-parametric tests controlling for serial correlation: the variance ratio statistic (v-statistic), the Phillips and Perron (PP) ρ-statistic and the PP t-statistic.[15] And then there is a fourth parametric ADF-type test. The test statistics are asymptotically normally distributed. The within-dimension presents all four of the statistics described here, whereas the v-statistic is excluded for the between-dimension tests.[16]

The Kao test uses the same basic approach as Pedroni, but imposes homogeneous slope coefficients. The intercepts are however allowed to vary between cross-sections. Consider the panel regression:

$$y_{it} = x'_{it}\beta + z'_{it}\gamma + e_{it} \qquad (3.27)$$

---

[15] As opposed to the ADF type unit root tests that account for higher-order autoregressive models by adding lagged first differences, the Phillips-Perron tests correct the t -statistic non-parametrically to account for autocorrelation.

[16] See for example Harris and Sollis (2003)

where x and y are non-stationary and $z_{it} = \alpha_i$ (fixed effects). The pooled auxiliary regression used to test for unit root in the residuals then becomes:

$$\hat{e}_{it} = \rho \hat{e}_{it-1} + v_{it} \tag{3.28}$$

and the hypotheses to be tested are the null of no cointegration against the alternative that x and y are cointegrated:

$$\begin{aligned} &H_0 : \rho = 1 \\ &H_1 : \rho < 1 \end{aligned} \tag{3.29}$$

The Kao test differs from the Pedroni test in the sense that it is assumed that $\rho_i = \rho$ for all individuals so that if the null is rejected it means that all units are cointegrated. Kao has developed four DF-type tests of which two assume strong exogeneity of the regressors and errors in equation (3.27) whereas the other two control for endogenous relationships. An augmented version of the Kao test includes lagged changes in the residuals and gives an ADF-type test. All of the Kao tests are asymptotically distributed under the standard normal distribution.[17]

The Fisher or combined Johansen test is a third panel cointegration test, adapted to the multi-equation framework. The methodology of the Fisher-type unit root test described in the previous section is, according to Maddala and Wu (1999), also applicable to cointegration tests with panel data when we test for unit roots in the residuals. The aim is again to compute the significance levels for the individual cross-sections and to combine these to get a test statistic for the full panel:

$$P = -2\sum_{i=1}^{N} \ln p_i \rightarrow \chi^2_{2N} \tag{3.30}$$

The statistic is based on the p-values obtained by MacKinnon et al. (1999), originally developed for more accurate results when employing the trace- and maximum eigenvalue tests by Johansen. The null hypotheses are the same as when performing an ordinary Johansen test, but instead of the usual trace and maximum eigenvalue statistics, they are tested against the Fisher trace and maximum eigenvalue statistics developed by Maddala and Wu (1999).

---

[17] See for example Harris and Sollis (2003)

4 Data

Chapter 4 includes a description of the data used to test our hypotheses. The data is also presented in graphs.

The data used was collected from Statistics Sweden (SCB, 2009). The data presented in the figures are raw data, meaning that no modifications have been made. Before the data was applied in the model all variables were seasonally adjusted, in order to remove potential seasonal components to affect the tests for non-seasonally trends. They were then turned into logarithms, except the short and long-term interest rate. House prices are measured by the real estate price index for one- or two-dwelling buildings for permanent living by region with 1981 as the base year. As the house price index is a nominal measurement, the consumer price index (CPI) is added in the model to take the inflation into account. CPI measures the price level of a basket of consumer goods. The CPI series base year is 1980 and measured monthly, so an average was calculated for each quarter. For the interest rate variable both the short-term and long-term interest rates were used, where the short-term is a three month treasury bill and the long-term is a ten year treasury bond.

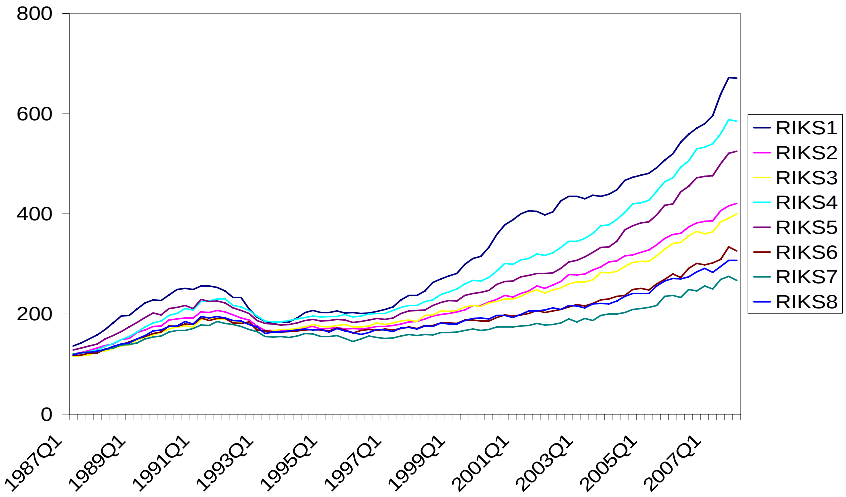
The income variable is measured by real GDP in millions SEK with 2000 as the reference year. As the GDP is a national measurement of production, regional unemployment is used as an indicator of regional production. When unemployment is low the production is high and vice versa. The definition of unemployment was changed in April 2005 in order for it to be in line with the definition used by the EU. The difference between the old Swedish definition and the new is that in the old definition full-time students that are searching for work are excluded. The unemployment measurement used in the unemployment variable here is the old definition, due to the problem of finding data on unemployment according to the new definition before 2005. All unemployment is measured in thousands and by the same definition.

Sweden is divided into eight regions named as RIKS1 to RIKS8. The counties included in each region are defined by Statistics Sweden (SCB, 2009).

Table 4.1 Definition of regions

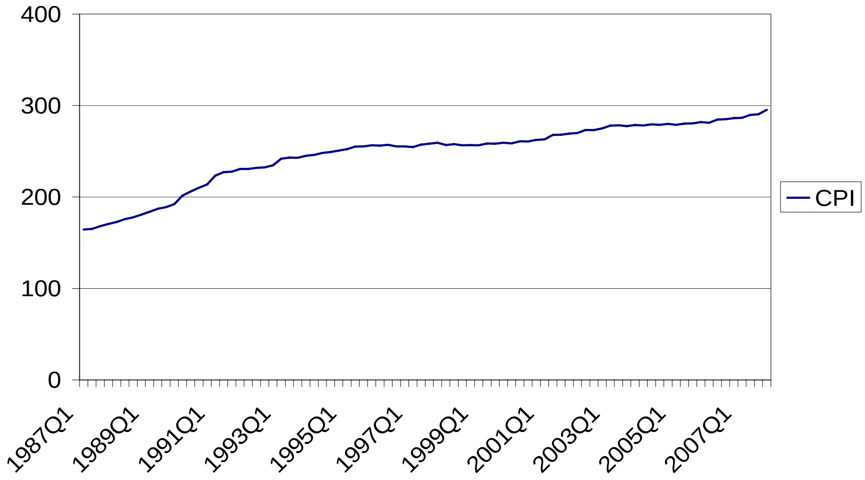
Regions	*Counties*
RIKS1 Stockholm	Stockholms
RIKS2 Eastern Middle-Sweden	Uppsala, Södermanlands, Östergötlands, Örebro, Västmanlands
RIKS3 Småland and the Islands	Jönköpings, Kronbergs, Kalmar, Gotlands
RIKS4 Southern Sweden	Blekinge, Skåne
RIKS5 Western Sweden	Hallands, Västra Götalands
RIKS6 Northern Middle-Sweden	Värmlands, Dalarnas, Gävleborgs
RIKS7 Middle Northern Sweden	Västernorrlands, Jämtlands
RIKS8 Upper Northern Sweden	Västerbottens, Norrbottens

Figure 4.1 House Price Index: All regions (1981=100)



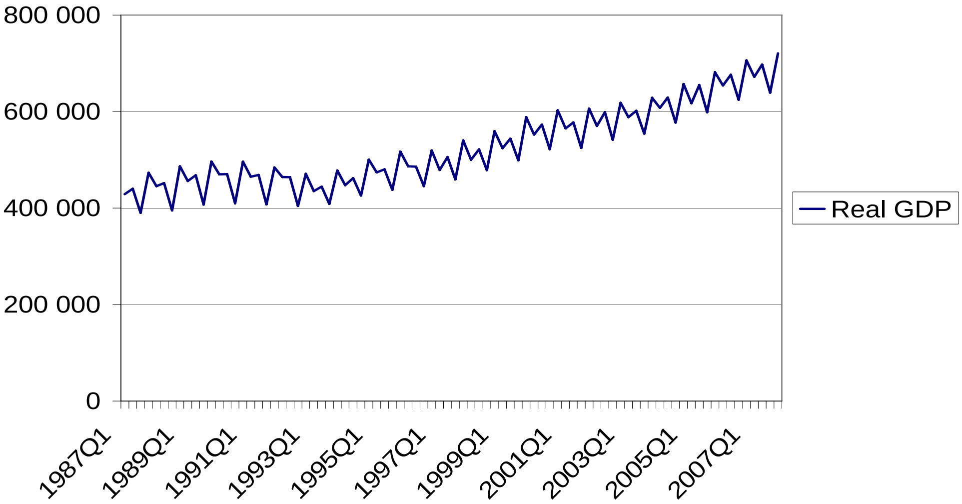
As the figure shows all regions' house prices move in the same direction, but the regional prices has differed more over the last years and the absolute difference is increasing.

Figure 4.2 Consumer Price Index (1980=100)



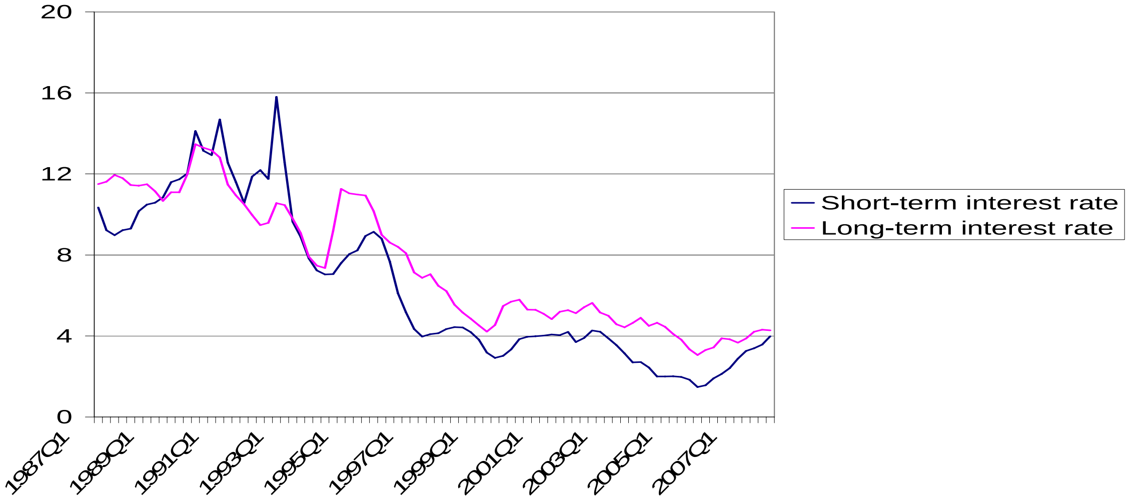
The consumer price index is increasing, as expected, due to inflation. Around year 1993 there is a clear stabilization of the CPI which coincides with the introduction of the inflation target in Sweden.

Figure 4.3 Gross Domestic Product



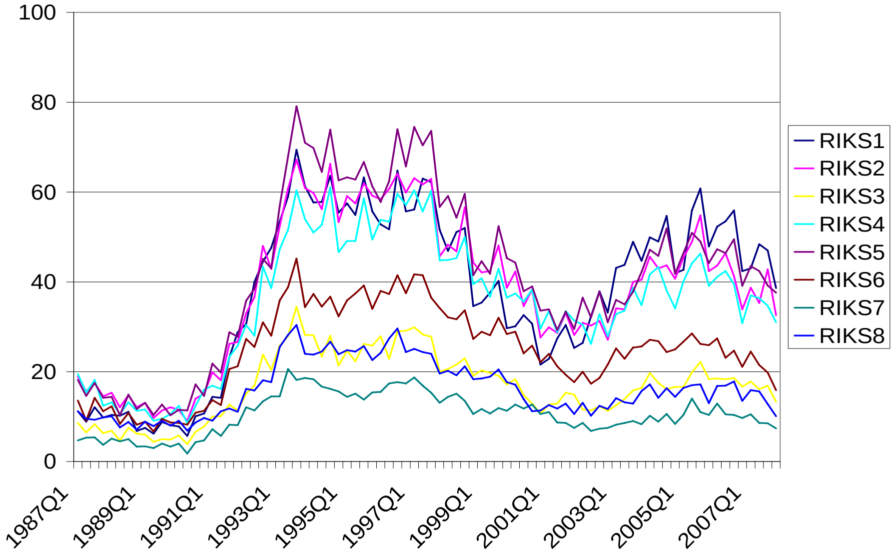
Graph 4.3 shows how the GDP fluctuates continuously. The decline resulting from the crisis in the beginning of the nineties can be distinguished around 1993. It can also be seen how GDP growth takes off after that as Sweden recovers from the crisis.

Figure 4.4 Short-Term and Long-Term Interest Rate



The interest rate is a national variable so it is the same for all regions. Both the short-term and the long-term interest rates have declined substantially over the last years and follow each other closely as expected.

Figure 4.5 Unemployment



There is a clear relationship between the unemployment in the different regions, but since the beginning of the nineties the levels of unemployment have differed a lot between the regions. Around year 1993-1998 Western Sweden and Småland and the Islands experienced the highest unemployment levels, but since the beginning of the twenty-first century Upper Northern Sweden has had the highest levels.

5 Long-Run Results

In this section, the results from the cointegration tests are presented and analyzed in regard to our hypotheses. We begin by investigating the house prices in all regions. This is then followed by investigating the long-run relationships of house prices in different groups of regions. Finally, all regional house prices are compared to the national average.

5.1 Cointegration Testing of Regional House Prices

5.1.1 Group Unit Root Testing on Regional House Prices

The results from the unit root tests on a group consisting of house price indices for the eight different regions are presented in the table below. All tests assume individual fixed effects by means of the inclusion of an individual intercept. The lag length used to correct for autocorrelation was automatically chosen using the Schwarz Bayesian Information Criterion (BIC). The BIC was chosen over Akaike's Information Criterion (AIC), since the penalty added for increasing the number of regressors is larger for the BIC and over parameterized models thus can be avoided.[18] For the Hadri test, where the autocorrelation correction is controlled by spectral estimation, the Bartlett kernel method was chosen with the Newey-West automatic bandwidth selection.

Table 5.1 Results from group unit root tests on regional house price indices

Test	Null hypothesis	Statistic	p-value
Levin, Lin & Chu	*Common unit root process*	2.083[a]	0.981
Hadri	*Stationarity*	16.998[b]	0.000
Im, Pesaran & Shin	*Individual unit root process*	4.789[c]	1.000
Fisher-ADF	*Individual unit root process*	0.997[d]	1.000

The statistics reported are a) LLC t-statistic b) Hadri heteroskedasticity consistent Z-statistic c) IPS $\bar{t}$ -statistic d) Fisher-ADF χ^2 -statistic

As table 5.1 shows, it can be concluded that the regional house price indices are non-stationary. All four tests suggest strong evidence of the presence of a unit root.

The next step is to test if the regional house prices are cointegrated and how many cointegrating relationships that exist if they are cointegrated. This is done employing the Johansen approach applied on cross-sectional data.

---

[18] See for example Verbeek (2008)

5.1.2 Cointegration between Regional House Prices

To determine the appropriate lag length, an AR model was constructed to test for the number of significant lags: $y_t = \delta + \sum_{k=1}^{5} \rho y_{t-k}$, where δ denotes a constant. The lag length was set to 5 and the results show significance for 1 lag in all cases except RIKS1. Since the lags specified in the Johansen test in EViews are for the first differenced terms in the auxiliary regression and not for levels, a test with one lag in the levels of the data requires the lag interval specification 0-0. Since the vast majority only indicate one significant lag, we will satisfy with that lag interval specification. For the inclusion of deterministic variables, it was assumed that our level data experience linear trends, but that the cointegrating equations only have intercepts.

As mentioned earlier, the housing market may deviate from region to region and it is thus problematic to describe the housing market on a national level. For the following investigation, the national housing market has been divided into eight separate markets where each region represents one housing market. Although there are eight separate markets there may be some similarities between them and some may be more related than others.

5.1.2.1 Cointegration between Regional House Prices: All Regions

The first cointegration test looks for a common trend in the house price indices of all eight regions. The Johansen approach will in this case test for up to seven cointegrating equations and the result is presented below. Rejecting the null of up to r cointegration relationships means that there may be $r+1$ cointegration vectors.

Table 5.2 Results from cointegration test on all regional house price indices

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.642	283.488**	85.367**
$r \leq 1$	0.508	198.121**	58.812**
$r \leq 2$	0.501	139.309**	57.694**
$r \leq 3$	0.424	81.615**	45.829**
$r \leq 4$	0.202	35.786	18.770
$r \leq 5$	0.129	17.016	11.485
$r \leq 6$	0.063	5.531	5.359
$r \leq 7$	0.002	0.172	0.172

** reject the null hypothesis at the 5 % significance level

According to the results in table 5.2, the trace test indicates four cointegrating equations as does the maximum eigenvalue statistic. It can thus be concluded that the regional house prices of Sweden are tied together by four long-run relationships, which is in line with our hypothesis stated in section 2.2. Since several of the factors affecting house prices in our model are aggregate measures, some long-run relationships between the regional house prices were expected. Some cautions in the interpretation of the results should however be considered. The Johansen test is very sensitive to the choice of lag length and which deterministic components that should be included. The economic implications and their plausibility should therefore be taken into account.[19]

One drawback with the Johansen approach is that we cannot tell which regions that are cointegrated. The test results in table 5.2 only tell us that four long-run relationships between some regions exist. For further investigation of which regions that are related, we have in the next section divided the regions into three groups: Northern Sweden, Middle Sweden including Stockholm and South-West of Sweden.

5.1.2.2 Cointegration between Regional House Prices: North, Middle and South

A long-run relationship is expected between the house prices in the regions which are most alike. The factor determining the resemblance tested for here and in the next section are geographical proximity and the existence of a metropolitan area in the regions. For this test, the eight regions were divided into three groups where Northern Sweden consists of RIKS6-8, Middle Sweden examines RIKS1-2 and South-Western Sweden contains the regions RIKS3-5. The results from the three Johansen tests performed are to be found in the tables below.

Table 5.3 Results from cointegration testing of regions in Northern Sweden

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.350	52.246**	35.800**
$r \leq 1$	0.180	16.446**	16.446**
$r \leq 2$	0.000	0.000	0.000

** reject the null hypothesis at the 5 % significance level

The results from the regions belonging to Northern Sweden indicate that there are two cointegrating equations. This is the maximum number of cointegrating equations that three regions could experience without it being a full rank which means that the variables are

---

[19] See for example Verbeek (2008)

stationary. The presence of two long-run relationships is in line with our hypothesis and this may also be an indication that these three housing markets have many common features.

Table 5.4 Results from cointegration testing of regions in Middle Sweden

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.230	22.698**	21.643**
$r \leq 1$	0.013	1.055	1.055

** reject the null hypothesis at the 5 % significance level

Since Middle Sweden only consists of two regions, there can only be one cointegrating relationship and the test results confirm that the two regions share a common trend. The regions accounted for under Middle Sweden belong to the most populous areas of the country with several of the largest cities. These cities are also allocated over a geographically small area which could lead to similarities in the housing markets. The close distance allows commuting to a larger extent which could lead to the markets converging to become more like a single market.

Table 5.5 Results from cointegration testing of regions in South-West of Sweden

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.220	34.530**	20.672
$r \leq 1$	0.151	13.857	13.608
$r \leq 2$	0.003	0.249	0.249

** reject the null hypothesis at the 5 % significance level

For South-West of Sweden the results are not as clear as for the two previous groups. The trace statistic indicates one cointegration equation whereas the maximum eigenvalue statistic implies that no long-run relationships exist. That the trace statistic and the maximum eigenvalue statistic leads to different conclusions regarding the number of cointegrating vectors is not uncommon and the interpretation in general of results obtained from the Johansen type cointegration tests can be difficult. One could for example use graphical examination of the cointegration relationships to help decide which statistic to adopt (Johansen and Juselius, 1992). To further refine the results for this regional grouping, cointegration tests were performed for all possible combinations of the three regions. The outcome of these tests shows that a common trend exists between RIKS3 (Småland and the Islands) and RIKS4 (Southern Sweden) which are the two most southern regions of Sweden.

The test statistics from Southern Sweden are listed in appendix A.1 (together with the results from the other combinations) with both the trace and the maximum eigenvalue statistic indicating cointegration.

5.1.2.3 Cointegration between Regional House Prices: Metropolitan areas

As some regions seem to be more similar than others, we would also like to test if there exists some correlations between the metropolitan areas. The cointegration tests performed in this section thus include the regions containing the three largest Swedish cities: Stockholm, Gothenburg and Malmö (RIKS1, RIKS4 and RIKS5).

Table 5.6 Results from cointegration testing in metropolitan areas

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.300	33.275**	29.637**
$r \leq 1$	0.030	3.638	2.513
$r \leq 2$	0.013	1.125	1.125

** reject the null hypothesis at the 5 % significance level

According to table 5.7 there is one long-run relationship that connects the metropolitan areas. Since we cannot tell between which regions this cointegrating equation lies, we performed further tests on the possible combinations of the three metropolitan regions. The tests performed led to the conclusion that Stockholm and Southern Sweden, including Malmö, share a common trend (see appendix A.1 for details). That the Stockholm and Malmö regions are cointegrated may be explained by the population growth rates of these two regions being higher than in the Gothenburg region (see Discussion in section 7).

5.1.3 Cointegration between Regional House Prices and the National Average

This section aims to examine the possible long-run connections between the regional house prices with the national average. As opposed to the previous cointegration tests, the Engle-Granger two-step procedure is employed here. The reason is that all these cointegration tests only include two variables at the time, whereas the tests of section 5.1.2 mostly include more than two variables. As described in the methodology section, with more than two variables there may be more than one long-run relationship, which the Johansen test was developed to control for. In bivariate cases however, the EG approach may generate more efficient results and this motivates the choice of method for this section. Altogether eight tests were performed. The first step of the test is to estimate a single equation including the two variables to be tested (one regional price index together with the national average) by OLS. The residuals generated are then saved to be tested for a unit root using the ADF test. Stationary residuals indicate that the regional house price index shares a common trend with the national average. The results from the ADF test are listed below. The null hypothesis is the presence of a unit root.

Table 5.7 Results from cointegration testing between regions and the national average

	t-statistic
RIKS1	-1.164
RIKS2	-2.120**
RIKS3	-2.116**
RIKS4	-1.722*
RIKS5	-1.760*
RIKS6	-2.630***
RIKS7	-1.387
RIKS8	-1.869*

* reject the null hypothesis at the 10 % significance level
** reject at the 5 % significance level
*** reject at the 1 % significance level

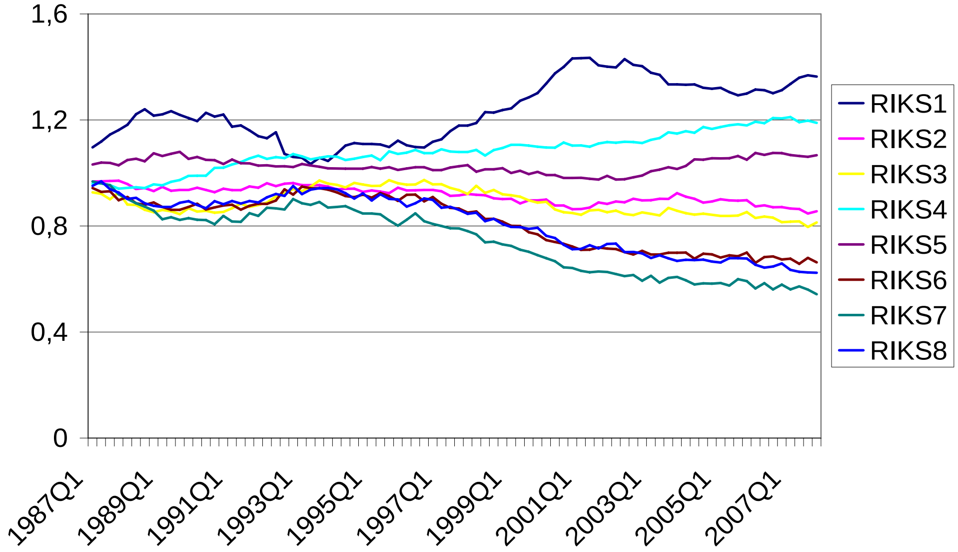
Table 5.7 shows that house prices in RIKS1 (Stockholm) and RIKS7 (Middle Northern Sweden) do not have any significant long-run relationship with the national average. At the ten per cent level it can be concluded that the house price indices in RIKS4 (Southern Sweden), RIKS5 (Western Sweden) and RIKS8 (Upper Northern Sweden) have significant cointegrating relationships with the average Swedish house price index. In regions RIKS2 (Eastern Middle-Sweden) and RIKS3 (Småland and the Islands), the null of a unit root can be rejected at the five per cent level and these regions thus share a common trend with the average. In RIKS6 (Northern Middle-Sweden) the null hypothesis can even be rejected at the one per cent level indicating stationary residuals and hence cointegration.

When performing the tests, it was expected that Stockholm and the northern regions would not experience common trends with the national average, since these regions are the two extremes regarding house prices on the Swedish housing market. The results confirm our expectations to some degree. The most northern parts belonging to RIKS8 (Upper Northern Sweden) do however yield significant results in support of cointegration, although only at the ten per cent level. RIKS1 (Stockholm) and RIKS7 (Middle Northern Sweden) on the other hand underpin our theory. Weaker support for cointegration is found in RIKS4 (Southern Sweden) and RIKS5 (Western Sweden). These are regions containing the second and third largest cities of Sweden and since the housing market in metropolitan areas often differs from the market in other areas they may be further away from the national average. The regions who indicate significant long-run relationships at the lowest levels (RIKS2, RIKS3 and

RIKS6) are regions without special features such as metropolitan areas and large unpopulated areal. These regions may thus provide an appropriate representation of the average Swedish housing market.

To show the relationships between regional house prices and the national average graphically, the ratios were calculated. These ratios are displayed in figure 5.1.

Figure 5.1 House Price Ratios to National Average



The graphical analysis of the ratios of regional house prices to the national average clearly shows how RIKS1 (Stockholm) deviates the most. The Stockholm price ratio is situated well above the other regions and also moves differently. The peaks of the housing market in Stockholm are not reflected in the house price movements of the other regions. Since the beginning of the nineties, the house prices in RIKS4 (Southern Sweden) have overhauled the prices in RIKS5 (Western Sweden). This may be a result of the increased efforts of integrating the Öresund region. Whereas house prices experience a positive trend relative to the national average in RIKS1, RIKS4 and RIKS5, the opposite holds for the most northern regions RIKS6-8. According to the graph, RIKS2 and RIKS3 seem to be the most accurate representations of the national average.

6 Short-Run Results

This chapter aims at estimating regional house price models that take the short-run behaviour into account. After having tested the properties of the data we will use an error correction model. This is then followed by applying the same tests and estimation method on panel data.

6.1 Testing the Regional House Price Model

6.1.1 Unit Root Testing on the Variables of the House Price Model

The purpose of testing for a unit root is to confirm that the variables are non-stationary to allow for cointegration testing. The results from the unit root tests are presented in the tables below. In the ADF test an intercept was included and the lag length was chosen using BIC, with a maximum number of lags set to eleven. The KPSS test was performed using the spectral estimation method of Bartlett Kernel and the Newey-West automatic bandwidth selection. As CPI, GDP, short-term interest rate and long-term interest rate are national variables there were no reason for performing more than one unit root test. However, as the house price index and the unemployment are regional variables, sixteen unit root tests were performed, two for each region. In the ADF test t-statistics are used to test the null hypothesis whereas in the KPSS test the test statistics are Lagrange multiplier (LM) statistics.

Table 6.1 Unit root test results, national variables

	ADF test	KPSS test
Null hypothesis	*Unit root*	*Stationarity*
CPI	-3.512**	0.266***
GDP	1.103	1.109***
Short-term interest rate	-1.190	0.974***
Long-term interest rate	-0.885	1.069***

** reject the null hypothesis at the 5 % significance level
*** reject at the 1 % significance level

As the results in table 6.1 show, it can be concluded that the variables at the national level are non-stationary. There may be some doubts regarding the non stationarity of CPI, since we can reject the null of a unit root at the five per cent significance level using the ADF test. On the other hand, the KPSS test indicates a unit root as the LM statistic allows us to reject the null of stationarity at the one per cent level. Since economic implications suggest that CPI would be non-stationary as it (usually) increases over time we choose to accept the result proposed by the KPSS test.

Table 6.2 Unit root test results, regional variables

	House price index		Unemployment	
	ADF test	KPSS test	ADF test	KPSS test
Null hypothesis	*Unit root*	*Stationarity*	*Unit root*	*Stationarity*
RIKS1	-0.522	1.039***	-2.124	0.524**
RIKS2	-0.221	1.023***	-2.317	0.422
RIKS3	0.171	1.085***	-1.643	0.360
RIKS4	-0.505	1.071***	-2.662	0.459
RIKS5	-0.106	1.011***	-2.000	0.429
RIKS6	-0.435	1.093***	-1.773	0.370
RIKS7	0.060	0.982***	-1.796	0.329
RIKS8	-0.423	1.077***	-1.453	0.244

** reject the null hypothesis at the 5 % significance level
*** reject at the 1 % significance level

Looking at the regional specific variables, the results indicate that the house price indices are clearly non-stationary. This confirms the conclusion drawn about house prices being non-stationary in chapter 5.1.1. According to the ADF test, we cannot reject the null hypothesis of a unit root and the test statistics generated from the KPSS test lead to the rejection of the null of stationarity at the one per cent level. Regarding unemployment the results are somewhat contradictive. The ADF test shows that unemployment is non-stationary whereas the null of stationarity in the KPSS test only can be rejected at the five per cent level for RIKS1 (Stockholm). Since the assumption of non-stationarity is necessary for testing for cointegration, the results from the ADF test will be the ones we base our assumption regarding unit roots in unemployment on.

6.1.2 Cointegration Testing between the Variables of the House Price Model

Testing for cointegration is a way of testing for a long-run relationship between the variables included in our house price model. The reason for testing for such a long-run relationship is that it enables us to estimate a model that regards both the long-run and the short-run behaviour of the variables. The method used for performing cointegration tests in this part of the thesis is the Engle-Granger (EG) two-step procedure. This is because we need estimates of the residuals to include in the ECM later on. We started by estimating the following equation by OLS:

$$\ln p_t^H = \delta + \varphi t + \beta_1 \ln CPI_t + \beta_2 \ln GDP_t + \beta_3 \ln U_t + \beta_4 r_t^s + \beta_5 r_t^l + \varepsilon_t \quad (6.1)$$

where δ and φt denote a constant and time trend respectively.[20] Since we chose to include the deterministic components in the initial equation, we must exclude these in the auxiliary regression used to test for unit roots. This in turn leads to the KPSS test being inappropriate since the EViews settings will not allow us to exclude these components in that test. Consequently the ADF test will be used here. The residuals generated from the estimation of (6.1) were saved and tested for unit roots. Stationary residuals indicate that the variables are cointegrated. The results from the unit root tests are presented in the table below, where the t-statistic from the ADF test is listed.

Table 6.3 Unit root test results, residuals

Null hypothesis	ADF test *Unit root*
RIKS1	-4.325***
RIKS2	-3.493***
RIKS3	-4.206***
RIKS4	-3.569***
RIKS5	-3.718***
RIKS6	-3.497***
RIKS7	-3.853***
RIKS8	-4.300***

*** reject at the 1 % significance level

If the variables are cointegrated, the test results should indicate stationarity. This means that the null of a unit root is rejected. As table 6.3 shows, the results indicate that we can reject the null hypothesis at the one per cent level. It can thus be concluded that the residuals are stationary for all regions. This indicates that the variables are cointegrated and thus have a long-run relationship that is stationary.

6.1.3 Estimating the House Price Model: ECM

After having concluded that the variables included in our house price model are cointegrated, it is possible to move on to estimating the regional models in order to see how they behave in the short and the long run. This will be accomplished by specifying our house price model as an error correction model (ECM). The ECM is estimated by including the residuals generated from the EG procedure used when testing for cointegration in an equation of the variables in first differences:

---

[20] The reason why we choose to include a trend is to get more significant results when we test for a unit root in the residuals. That adding a trend leads to better results is a sign of omitted variables, i.e. there is some factor affecting house prices that we have not accounted for.

$$\Delta \ln p_t^H = \delta + \phi_1 \Delta \ln CPI_t + \phi_2 \Delta \ln GDP_t + \phi_3 \Delta \ln U_t + \phi_4 \Delta r_t^s + \phi_5 \Delta r_t^l + \gamma \varepsilon_{t-1} + u_t \qquad (6.2)$$

Again δ denotes a constant, γ is the speed of adjustment term, ε_{t-1} are the lagged residuals from the EG procedure, i.e. the cointegrating relationship, and u_t is an error term. Together, $\gamma\varepsilon_{t-1}$ forms the error correction term. The results from the estimation of the ECM are shown in table 6.4 below. Inference is performed using the Newey-West standard errors compensating for autocorrelation in the residuals. The null hypothesis states that the slope coefficients are equal to zero.

Table 6.4 Estimation output from the regional ECM:s

	RIKS1		**RIKS2**		**RIKS3**		**RIKS4**	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
δ	0.009	1.133	0.009	1.600	0.007	1.380	0.011	1.942*
$\Delta \ln CPI_{t-1}$	0.719	1.880*	0.583	1.742*	0.807	2.087**	0.828	2.031**
$\Delta \ln GDP_{t-1}$	1.186	2.158**	0.463	1.219	0.494	1.641	0.552	1.505
$\Delta \ln U_{t-1}$	-0.070	-4.091***	-0.040	-1.579	-0.042	-2.198**	-0.078	-2.817***
Δr_{t-1}^s	0.003	0.571	0.005	1.165	0.004	1.121	0.005	1.083
Δr_{t-1}^l	-0.001	-0.146	0.000	0.012	-0.003	-0.693	-0.003	-0.653
ε_{t-1}	-0.182	-1.846*	-0.071	-1.838*	-0.087	-1.513	-0.068	-1.438

	RIKS5		**RIKS6**		**RIKS7**		**RIKS8**	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
δ	0.007	1.227	0.005	1.173	0.004	0.782	0.005	1.065
$\Delta \ln CPI_{t-1}$	0.858	2.185**	0.854	2.806***	0.502	1.628	0.544	1.442
$\Delta \ln GDP_{t-1}$	0.718	1.930*	0.309	1.030	0.462	1.271	0.425	1.110
$\Delta \ln U_{t-1}$	-0.073	-1.568	-0.065	-2.238**	-0.014	-1.047	-0.043	-1.394
Δr_{t-1}^s	0.002	0.500	0.007	1.701*	0.002	0.481	0.003	0.688
Δr_{t-1}^l	-0.003	-0.564	-0.003	-0.537	0.003	0.600	-0.004	-0.879
ε_{t-1}	-0.036	-0.683	-0.047	-1.340	-0.084	-1.618	-0.086	-1.273

* reject the null hypothesis at the 10 % significance level
** reject at the 5 % significance level
*** reject at the 1 % significance level

The table above presents the results from the eight regions considered in this thesis. The error correction term is only significant at the ten per cent level in RIKS1 (Stockholm) and RIKS2 (Eastern Middle-Sweden). The slope coefficient marks the speed of adjustment which in RIKS2 is very small meaning that it takes a while for house prices to respond and return to the equilibrium growth path. In RIKS1, the speed of adjustment is slightly higher but still fairly slow. Regarding the other regions we cannot exclude that the speed of adjustment term equals zero and there may be no error correction mechanism. This implies that the variables in these regions are not cointegrated, which contradicts our results from the Engle-Granger cointegration tests that all found evidence of long-run relationships. We thus cannot draw any conclusions regarding how the variables are pulled back towards their long-run equilibrium relationship. This could be due to too few observations to investigate the long-run connections of the model.

Considering the short-run behaviour of the model, CPI is significant at the ten per cent level in six regions of which RIKS3 (Småland and the Islands), RIKS4 (Southern Sweden) and RIKS5 (Western Sweden) show significance at the five per cent level and RIKS6 (Northern Middle-Sweden) at the one per cent level. This is thus the only variable that is significant throughout more or less all the regions. The only regions where GDP seems to have an impact on house prices are RIKS5 (at the ten per cent level) and RIKS1 (at the five per cent level). Unemployment is significant at the five per cent level in two regions and at the one per cent level in two regions out of eight. The short-term interest rate shows significance at the ten per cent level in RIKS6. As the significance of the short-term interest rate is observed only once and in a region without any special features (as defined here), it could be explained by properties of the data that have little to do with the economic implications as opposed to CPI and unemployment that have a clearer pattern of significance. The long-term interest rate is insignificant in all regions and at every level (1-10 per cent). The signs of the slope coefficients are consistent with economic theory. Changes in unemployment will have a negative effect on house price growth whereas CPI and GDP changes have a positive effect.

Our hypothesis stated that the northern regions would be less sensitive to changes in interest rates and unemployment than the southern and metropolitan areas, as the prices are lower resulting in the leverage being lower. The results of the ECM show that the effect of interest rates is not significant, except for RIKS6 (Northern Middle-Sweden), which results in that no conclusions can be made to accept or reject our hypothesis about the difference in sensitiveness to interest rates between the regions.

Changes in the unemployment should according to the hypothesis presented in section 2.3 have higher impact on the house prices in the south and metropolitan areas than in the north, due to higher leverage. The impact of unemployment on house prices was found to be insignificant in the most northern regions, so a comparison between the northern and the southern regions is difficult. Since the coefficients are insignificant in these regions, it is suggested that unemployment does not have any effect on house prices, contradicting the results from some of the southern regions. Of the regions where unemployment was found to be significant, the most sensitive to changes in unemployment were RIKS1 (Stockholm) and RIKS4 (Southern Sweden). The region that seems least affected by changes in unemployment is RIKS3 (Småland and the Islands). The regions where unemployment is insignificant are in part located in the more unpopulated regions in the north of Sweden, with the exception of RIKS2 (Eastern Middle-Sweden) which contains several fairly large cities and RIKS5 (Western Sweden) where Gothenburg is situated. The results are otherwise in compliance with the hypothesis that house prices would be more receptive to changes in unemployment in metropolitan areas and the southern regions. The continuing urbanization due to better work opportunities in larger cities creates higher pressure on the housing market, making it more sensitive to changes in unemployment. In less populated areas, the only unemployment affecting house prices is the local one whereas in urban areas there is in part the unemployment of the residents as well as of the new incoming city-dwellers. Low unemployment rates in the cities induce urbanization and thus increases demand for housing, pressuring the house prices upwards.

The hypothesis also stated that income changes would have smaller effects on the northern regions than the southern. The GDP variable was insignificant for almost all regions except RIKS1 (Stockholm) and RIKS5 (Western Sweden), which contains the two largest cities Stockholm and Gothenburg. The results indicate that the house prices in RIKS1 prove to exhibit a substantially larger positive effect from a change in income. The coefficient for RIKS5 also indicates larger effects than the coefficients in the regions where GDP proved to be insignificant. This suggests that GDP would have a greater impact on house prices in the metropolitan areas. This result is partly in line with the hypothesis, since the insignificant results imply that GDP would not have any effects on house prices. That house price changes are more sensitive to changes in GDP in the Stockholm area may be due to the urban housing market being already more burdened by excess demand and thereby becoming more responsive to business cycle fluctuations.

The motivation for including CPI in the model is that the house price index denoting the dependent variable is a nominal value. That CPI, which was included to control for inflation, is significant in more or less all regions (CPI turned out to be insignificant in RIKS7 and

RIKS8, the two most northern regions) was thus expected, since it would be curious if inflation did not have an impact on nominal prices.

The results for the two most northern regions (RIKS7 and RIKS8) cannot tell us much about the dynamics of the house price model since there are no significant variables. This suggests that the model used here is inappropriate for explaining house price changes in northern Sweden.

6.2 Testing the House Price Model with Panel Data

To be able to estimate the house price model with panel data by specifying an ECM, we need to go through the same procedure as in section 6.1, however applied on panel data. The variables in this section are not seasonally adjusted as EViews cannot perform seasonally adjustment on data with more than 600 observations and there are 672 observations in the panel data.

6.2.1 Panel Unit Root Testing on the Variables of the House Price Model

In order to look for cointegration between our macroeconomic variables, we begin by testing whether they are non-stationary or not. Several available panel unit root tests will be performed where it is tested both for a common unit root process and individual unit root processes. The null hypothesis will in general be the presence of a unit root except for the Hadri test that tests the null of stationarity (compare with the KPSS test). The results from the different unit root tests are presented in a table below. For the LLC test, the IPS test and the Fisher-ADF test the lag length was chosen by BIC. For the LLC test and the Hadri test a spectral estimation method also has to be specified (the Hadri test does not require lag length specification) and this was done using the method of Bartlett Kernel with the Newey-West automatic bandwidth selection. All tests include an individual intercept to the test regression.

Table 6.5 Panel unit root tests of the variables included in the house price model

Test	Null hypothesis	House prices	CPI	Short-term interest rate
Levin, Lin & Chu	*Common unit root process*	2.083	-18.914***	-0.176
Hadri	*Stationarity*	16.998***	15.507***	15.334***
Im, Pesaran & Shin	*Individual unit root process*	4.789	-13.849***	1.083
Fisher-ADF	*Individual unit root process*	0.997	195.189***	6.369

Test	Null hypothesis	Unemployment	GDP	Long-term interest rate
Levin, Lin & Chu	*Common unit root process*	-3.709***	7.315	0.467
Hadri	*Stationarity*	4.357***	20.522***	17.107***
Im, Pesaran & Shin	*Individual unit root process*	-2.926***	7.521	1.763
Fisher-ADF	*Individual unit root process*	34.075***	0.082	4.054

*** reject the null hypothesis at the 1 % significance level

According to the results in table 6.5 house prices, GDP and the short- and long-term interest rates all contain a unit root and are thus clearly non-stationary. Looking at the test results for CPI and unemployment, the only test indicating non-stationarity is the Hadri test where the null of stationarity can be rejected at the one per cent level. As mentioned under 6.1.1, the economic implications of the properties of CPI are that it could be non-stationary, since it usually increases over time. We will therefore accept the results generated through the Hadri test and assume that CPI is non-stationary. Regarding unemployment, the Hadri test indicates non-stationarity whereas the other tests suggest stationarity. Unemployment fluctuates around its mean, but the fluctuations are not as frequent as with clearly stationary data. This suggests non-stationarity since periods of increasing or decreasing unemployment stretch over several observations. We will thus accept the results from the Hadri test.

6.2.2 Panel Cointegration Testing of the Variables of the House Price Model

Since we have concluded that the variables entering the house price model are non-stationary we can proceed to the next step and test if they are cointegrated and thus share a common trend. For this purpose we will apply the panel cointegration tests described in the methodology section: the Pedroni test, the Kao test and the combined Johansen test. The results are presented in three tables, one for each cointegration test.

The Pedroni test was performed by assuming that the variables have both an individual intercept and an individual trend. The maximum number of lags was chosen by BIC and the spectral estimation method used is the Bartlett Kernel method with the Newey-West automatic bandwidth selection.

Table 6.6 Results from Pedroni panel cointegration test

Null hypothesis	*No cointegration*		
Within-dimension		Between-dimension	
	Statistic		Statistic
v-statistic	-1.391		
ρ-statistic	-0.156	ρ-statistic	0.797
PP t-statistic	-3.062***	PP t-statistic	-2.751***
ADF statistic	0.051	ADF statistic	0.778

*** reject the null hypothesis at the 1 % significance level

According to the results from the Pedroni test in table 6.6, the null of no cointegration can only be rejected if we look at the PP t-statistic, where the null can be rejected at the one per cent level for both the within- and the between-dimension. The results are thus somewhat conflicting regarding whether there is a long-run relationship between the variables or not.

The specifications needed to be able to perform the Kao test are the assumption of an individual intercept, maximum number of lags chosen by BIC and the Bartlett Kernel spectral estimation method with the Newey-West automatic bandwidth selection.

Table 6.7 Results from Kao panel cointegration test

Null hypothesis	*No cointegration*
	Statistic
ADF t-statistic	-6.070***

*** reject the null hypothesis at the 1 % significance level

If we instead use the Kao test to see if the variables share a common trend, the result is clearer than the ones from the Pedroni test. According to the ADF t-statistic in table 6.7, the null of no cointegration can be rejected at the one per cent significance level. The Kao test thus suggests that the variables are cointegrated.

When performing the combined Johansen test we assume that our level data experiences linear trends whereas the cointegrating equations only have intercepts. The lag length was set to one lag in the level data based on results from testing for significant lags as we did in part 5.1.2. Again, one lag in the level data requires the specification of the lag interval 0-0 in EViews, since the lag specification in EViews is for differenced data.

Table 6.8 Results from combined Johansen panel cointegration test

Null hypothesis	Fisher trace statistic	Fisher max. eigenvalue statistic
$r = 0$	471.249***	231.872***
$r \leq 1$	293.583***	238.387***
$r \leq 2$	107.566***	116.105***
$r \leq 3$	19.111	14.734
$r \leq 4$	14.041	14.997
$r \leq 5$	12.322	12.322

** reject the null hypothesis at the 5 % significance level
*** reject at the 1 % significance level

The combined Johansen test does not only test for cointegration but also for the number of cointegration relationships that exists. Recall from the methodology section that with more than two variables there may be more than one long-run relationship. As the results in table 6.8 show, there may be four significant cointegration vectors according to the two Fisher statistics. Recall that rejecting the null of up to r cointegration relationships means that there may be $r+1$ cointegration vectors. The null of up to three cointegrating vectors can be rejected at the one per cent level for both statistics. In our case however, deciding the number of cointegrating vectors is not relevant, since we only want to investigate whether they are cointegrated or not to be able to continue with estimating the model.

Based on the results from the Kao test, the combined Johansen test and the ρ-statistic and the PP t-statistic from the Pedroni test we can assume that there exists some long-run relationship between the variables that our house price model consists of. This allows us to continue with the next step which is to estimate an ECM with our panel data to see how the variables behave in the short run.

6.2.3 Estimating the House Price Model with Panel Data

Previously, we estimated regional house price models using the error correction method to see if any conclusions regarding the short-run behaviour of the included variables could be drawn considering each region. The results were to some extent unclear, since the error correction term was insignificant most of the time. To increase the number of observations in the hope of more efficient results, we thus put all the regions together to form a panel of individuals observed over time. When working with panel data there are several assumptions that need to be made about the properties of the data and how the individuals are connected. In our case we choose to work with fixed individual effects that capture differences between the regions that are independent of time by including individual heterogeneous intercepts. The choice of fixed effects is justified by the fact that it is a less restrictive assumption, compared to random effects, in that it allows the observed regressors to be correlated with unobservable features of these effects. Fixed effects are also considered more appropriate when the individuals to be examined are unique, i.e. countries, companies or regions and not some random sample drawn from a larger population.[21] By allowing for contemporaneous correlation between the regions, EViews will generate a feasible generalized least squares (FGLS) estimator that controls for heteroskedasticity and correlation between regions. To get the cointegrated long-run equation, a simple model was first estimated to yield residuals that can be tested for unit roots. The results from the EG type cointegration test are enlisted below.

Table 6.9 Panel unit root tests, residuals

Test	Null hypothesis	Statistic	p -value
Levin, Lin & Chu	*Common unit root process*	-5.222^a	0.000
Fisher-ADF	*Individual unit root process*	55.517^b	0.000

The statistics reported are a) LLC t-statistic b) Fisher-ADF χ^2 -statistic

As both tests indicate, the p-values are small enough to conclude that the residuals are stationary which implies that the variables are cointegrated. Only two unit root tests are presented here. Since we chose to include the deterministic components in the initial regression, we could only perform unit root tests that allowed us to exclude the intercept and trend in the auxiliary regression.

The next step is thus to estimate the ECM where the lagged residuals are included in a regression together with the first differences of the regressors with the dependent variable being differentiated house prices:

$$\Delta \ln p_{it}^{H} = \alpha_i + \phi_1 \Delta \ln CPI_{it} + \phi_2 \Delta \ln GDP_{it} + \phi_3 \Delta \ln U_{it} + \phi_4 \Delta r_{it}^{s} + \phi_5 \Delta r_i^{l} - \gamma \varepsilon_{it-1} + u_{it} \quad (6.3)$$

The fixed effects are denoted by α_i, $\gamma \varepsilon_{it-1}$ is the error correction term and γ is the speed of adjustment. The slope coefficients of the macro variables describe the short-run reactions of the model, consistent with their common trend, whereas the error correction term shows how the model is pulled back towards its long-run equilibrium. The residuals saved from the EG procedure thus measure the distance away from equilibrium. The regression output, with the assumptions stated in the beginning of section 6.2.3 being applied, is presented in table 6.10. The null hypothesis to be tested is whether the slope coefficients are zero.

---

[21] See for example Verbeek (2008)

Table 6.10 Estimation output from the panel ECM

	Coefficient	t-statistic
δ	0.012	4.521***
$\Delta \ln CPI_{t-1}$	0.632	2.891***
$\Delta \ln GDP_{t-1}$	-0.095	-4.500***
$\Delta \ln U_{t-1}$	-0.014	-2.486**
Δr^{s}_{t-1}	0.004	1.730*
Δr^{l}_{t-1}	0.002	0.518
ε_{t-1}	-0.019	-2.472**

* reject the null hypothesis at the 10 % level
** reject at the 5 % significance level
*** reject at the 1 % significance level

According to these results, the error correction term is significant at the five per cent level. The coefficient of the EC term measures how fast house prices are pulled back towards their equilibrium growth path. Here, the speed of adjustment term is very low, implying that house prices return slowly to their long-run steady state.

The variable included to control for inflation, CPI, is significant which was expected, since the house prices in our model are nominal. As in some of the regional models, the t-statistic of GDP indicates a significant impact on house price changes. The sign of the slope coefficient is however negative which contradicts the conclusions from the regional ECM:s. These results suggest that an increase in GDP would have a negative effect on house price changes which is somewhat confusing considering economic theory. As in the regional models, unemployment is significant. The short-term interest rate is also significant at the ten per cent level. We can thus come to the conclusion that regarding the short-run behaviour of the house price model, the panel ECM and the regional ECM:s yield similar results. The main difference is the significance of the cointegrating equation describing the long-run relationship between the variables, i.e. the error correction term.

Comparing the panel ECM with the regional ECM:s one might declare that a panel, which generates results for the aggregated level, is somewhat misleading. The major motivation for this is that the panel generates a negative GDP coefficient as opposed to the regional models who produce positive slope coefficients for GDP and that the short-term interest rate, that only affects house prices in one out of eight regions, is significant in the panel. The panel results should however be considered cautiously, since these types of regressions are very sensitive to the assumptions made. The long-term interest rate is the only variable that is insignificant at every level (1-10 per cent).

7 Discussion

In this section, a discussion of possible causes and explanations for the empirical findings of the thesis is conducted.

Although the regional housing markets of Sweden may differ there are also some resemblances between them, since they face the same macroeconomic conditions. Regional house prices are found to experience several long-run relationships which indicate that some regions are cointegrated while some are not. That the regional house prices are cointegrated means that they share a common trend which in turn means that they move in the same direction, but may be situated at different levels.

Sweden is a much diversified country with large differences between the northern and the southern parts, where the north of Sweden consists of large unpopulated areal and long distances from the continent and the major trade centres. Attempts to stop the depopulation of the north by placing public authorities in smaller towns of the northern regions have not led to desired results. This may in turn affect the regional housing markets and to account for this the house prices were arranged into three parts: North, Middle and South. The classification of the northern and the middle parts of Sweden generated results in line with our hypothesis that the geographical proximity may lead to common features of the housing markets.

On the contrary, our classification of southern Sweden could be inaccurate, since the tests indicate that the western region (with Gothenburg) differs from the other regions included in the south. This leads on to the cointegration tests between the metropolitan areas where a common trend was found between Stockholm and the region including Malmö. That Stockholm and Malmö have a long-run relationship but not Stockholm and Gothenburg or Malmö and Gothenburg could be explained by population growth. When comparing the population growth between these three cities for the years 1987-2007, both Malmö and Stockholm have experienced higher yearly average growth rates than Gothenburg. The population growth in Stockholm is however by far much higher. Until 1997, Malmö and Gothenburg had similar growth patterns, but after that the population growth of Malmö escalated (see appendix A.2). This could be due to the large efforts to promote the Öresund region after the Öresund Bridge was completed in 2000.

When comparing the regional house prices to the national average, it is reasonable to expect that the regions that do not have any extreme features such as metropolitan areas or geographical isolation would yield results indicating cointegration. These regions are in general a better representation of the national average than the other regions. This supports our hypothesis stating that the markets in Stockholm and the northern parts of Sweden would not be well represented by an aggregated national housing market.

To further test our hypothesis that the housing markets differ between regions, we estimate a house price model including selected variables. Changes in GDP turn out to be insignificant for house price changes in almost all regions, which contradict the results from the panel regression. In the regions where GDP turns out to be of importance, it seems to have a higher impact on house price changes in Stockholm. GDP could be considered as an approximation for national income levels and increasing income will increase demand and thus put higher pressure on the housing market. It is thus curious that national income does not have a more frequent impact on house prices throughout the regions. Unemployment proves to have significant effects in particularly the southern more populated regions and it appears to be of more economic significance in the metropolitan regions (RIKS1, RIKS4), with the exception of RIKS5. Periods of low unemployment rates in metropolitan regions may attract people from other parts of the country who seek new work opportunities. The housing markets in these regions thus need to cope with the demand from incoming workers on top of the initial demand. The panel regression confirms the significance of unemployment, however on the aggregate level. The short-term interest rate turns out to be significant in the panel and for one out of eight regions when estimating the regional models. The coefficients for the significant short-term interest rates are however so small that their economic significance could be considered very negligible. There are no indications of long-term interest rates having an impact on house price changes. According to the Global Property Guide (2008), this could be explained by the fact that the majority of Swedish borrowers that fix their interest rates do that for less than five years. The interest rate used to explain long-term interest rates here is however the ten year treasury bond.

The most interesting feature of the panel estimation is the significant error correction term. This means that the variables of the model are moving towards a common equilibrium: when all variables reach their steady state where they experience constant growth rates. The slope coefficient marks the speed of adjustment which tells us how fast house prices will return to their steady state after a shock to the other variables. According to the speed of adjustment in this case, house prices are very slow to return to their equilibrium growth path.

The two extremes regarding regional housing markets, Stockholm and the northern regions, are very different compared to each other. The market in Stockholm could be said to be rather stressed with high activity, which is the opposite of the markets of the northern regions. Northern Sweden has long had to cope with depopulation issues which lead to the regions ending up in a vicious circle. The urbanization has forced local businesses, schools and care units to close down due to falling demand, which makes it harder for the remaining residents and further spurs urbanization. The housing demand in Stockholm and the north thus becomes very biased. Increasing demand in metropolitan regions push prices upwards whereas the decreasing demand in the northern regions pull down the house prices.

If interpreting the "ripple effect" as the prominent region being the one most sensitive to shocks in the independent variables, Stockholm may in one way be the prominent region in Sweden. We can however only tell that changes in the macro variables will have a larger impact on house price changes in Stockholm. Since the error correction terms are insignificant in most of the regions, we cannot really tell if it is the region to respond the fastest to possible shocks. Comparing the speed of adjustment in Stockholm with the coefficients of the error correction term in the insignificant regions could suggest that Stockholm reacts faster to changes in the macro variables. The insignificant cointegration relationships in the other regions however limit the possibility to draw any conclusions.

8 Conclusion

This chapter describes the conclusions to be drawn from our study with very brief presentations of the overall results. The section ends with suggestions for further research and some words of caution for policy makers.

The aim of the thesis is to test whether aggregated national house price models are misleading as they do not consider regional differences in housing markets. Our hypothesis is that there exists a long-run relationship between house prices in different regions and that the regions with the most resemblance will be cointegrated. When performing a cointegration test including all regions, the results indicate that there are several cointegration relationships, but we are not able to tell between which regions these relationships lie. To determine where the cointegration is, the regions were divided into three groups: North, Middle and South. The regions included in North and Middle share a common trend respectively as opposed to South where the long-run relationship exists between the two most southern regions whereas the western region included in the group does not indicate such a relationship. When testing for cointegration between the metropolitan areas, the Gothenburg region (i.e. the western region) once again shows no indication of being cointegrated with the other regions.

Estimating the house price model in error correction form generates results indicating that CPI has a significant impact on house price movements, whereas unemployment and GDP mainly influences the house price changes in the metropolitan areas and the southern half of Sweden. When applying the hypothesis proposed by Meen (1999) on Sweden, that the southern and metropolitan regions would be more sensitive to changes in the interest rate because of higher leverage, our results indicate no such effect. Even though the same variables may be statistically significant in several regions their economic significance is of different proportions. GDP seems to have a larger impact in Stockholm as does unemployment. The panel estimation leads to similar conclusions with the exception that the error correction term is significant. We are thus able to see how house prices return to their steady state growth path through the speed of adjustment term, which in this case however is very low, indicating that house prices are pulled back towards equilibrium very slowly. The estimation output for the two most northern regions proved to have no significant macro variables. This could indicate that the model specification used in this thesis does not adequately cover the specific features of these markets.

Some aspects not accounted for in this thesis which would be interesting for future research are how demographic changes, such as urbanization and population growth, could affect regional house prices. It could also be interesting to investigate how the housing market in southern Sweden has changed, since the efforts of integrating the Öresund region have increased. We have only looked at the immediate reaction of house prices to changes in macro variables. The housing market is however sometimes referred to as slow and it could thus make sense to include time-lag effects in the house price model. Another suggestion is to look for cointegration in other classifications of the regions than the ones used in this thesis.

To conclude this thesis, aggregated national housing models may be very misleading, especially in a country as diversified as Sweden. As the results showed, the model specification used in this thesis proved to be inappropriate for the two most northern regions, which further supports that regional differences affect the housing markets differently. This should also be accounted for when investigating the housing market for policy purposes.

References

Published references

Alexander, C. & Barrow, M. (1994), “Seasonality and Cointegration of Regional House Prices in the UK”, *Urban Studies*, Vol.31, No.10, pp. 1667-89

Ashworth, J & Parker, S. (1997), “Modelling Regional House Prices in the UK, *Scottish Journal of Politics Economy*, Vol. 44, No.3, pp. 225-46

Barot, B. & Yang, Z. (2002), “House Prices and Housing Investment in Sweden and the UK: Econometric analysis for the Period 1970- 1998”, *Review of Urban & Regional Development Studies*, Vol. 14, No 2, July 2002

Gallin, J. (2006), “The long-run relationship between house prices and income: evidence from local housing markets”, *Real Estate Economics*, Vol. 34, No. 3, pp. 417-438

Harris, R. and Sollis, R., (2003), *Applied Time Series Modelling and Forecasting*, John Wiley & Sons Ltd.

Hort, K. (1998), “The Determinants of Urban House Price Fluctuations in Sweden 1968-1994”, *Journal of Housing Economics*, Vol. 7, pp. 93-120

Iacoviello, M. (2000), “Housing Prices and the Macroeconomy in Europe: Results from a Structural VAR Analysis”, European Central Bank, Working Paper series 18.

IMF, (2008), “World Economic Outlook, April 2008, Chapter 3: The Changing Housing Cycle and the Implications for Monetary Policy”

Johansen, S. (1988), “Statistical Analysis of Cointegration Vectors”, *Journal of Economic Dynamics and Control*, Vol. 12, pp. 231-54

Johansen, S. (1989), “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”, Miemeo, Institute of Mathematical Statistics, Copenhagen, *Econometrica*

Johansen, S. (1991), “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”, *Econometrica*, Vol. 59, pp. 1551-80

Johansen, S. and Juselius, K., (1992), “Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK”, *Journal of Econometrics*, Vol. 54, No. 1-3, pp.211-245

Levin, A., Lin, C-F. and Chu, C-S. J., (2002), “Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties”, *Journal of Econometrics*, Vol. 108, No. 1, pp. 1-24

MacDonald, R & Taylor, M.P. (1993), “Regional House Price in Britain: Long-Run Relationships and Short-Run Dynamics”, *Scottish Journal of Political Economy*, Vol. 40, No.1, pp. 43-55

MacKinnon, J. G., Haug, A. A. and Michelis, L., (1999), “Numerical Distribution Functions of Likelihood Ratio Tests for Cointegration”, *Journal of Applied Econometrics*, Vol. 14, No. 5, pp. 563-577

Maddala, G. S. and Wu, S. (1999), “A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test”, *Oxford Bulletin of Economics & Statistics*, Vol. 61, No. S1, pp. 631-652

Meen, G. (1999), “Regional House Prices and the Ripple Effect: A new Interpretation”, *Housing Studies*, Vol.14, No.6, pp.733-53

Sørensen, P.B. and Whitta-Jacobsen, H.J., (2005), *Introducing Advanced Macroeconomics: Growth & Business Cycles*, McGraw-Hill Education

Verbeek, M., (2008), *A Guide to Modern Econometrics*, John Wiley & Sons Ltd.

Electronic references

Global property Guide, “Credit crunch grounds Sweden’s housing market”, (Aug 27, 2008)
http://www.globalpropertyguide.com/Europe/Sweden/Price-History, (May 23, 2009)

Riksbanken, (2009), ”Historia, viktiga årtal 1900-talet”,
http://www.riksbank.se/templates/Page.aspx?id=9139, (May 5, 2009)

Statistics Sweden, http://www.ssd.scb.se/databaser/makro/SaveShow.asp, (May 5, 2009)

Appendix

A.1

Cointegration tests for submarkets

RIKS3 & RIKS4

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.216	20.289**	20.238**
$r \leq 1$	0.001	0.051	0.051

RIKS3 & RIKS5

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.162	14.852	14.651
$r \leq 1$	0.002	0.201	0.201

RIKS4 & RIKS5

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.153	14.075	13.754
$r \leq 1$	0.004	0.321	0.321

RIKS1 & RIKS4

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.250	24.498**	23.892**
$r \leq 1$	0.007	0.606	0.606

RIKS1 & RIKS5

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.096	9.502	8.408
$r \leq 1$	0.013	1.094	1.094

RIKS4 & RIKS5

Null hypothesis	Eigenvalue $\hat{\lambda}$	$\hat{\lambda}_{trace}$	$\hat{\lambda}_{\max}$
$r = 0$	0.153	14.075	13.754
$r \leq 1$	0.004	0.321	0.321

** reject the null hypothesis at the 5 % level

A. 2

Regional population growth 1987-2007 (in %).

The growth rates have been calculated using data collected from Statistics Sweden (SCB, 2009).

	RIKS1	**RIKS2**	**RIKS3**	**RIKS4**
Average yearly growth	0.973	0.364	0.121	0.641
Growth 1987-1997	9.760	4.795	2.404	6.656
Growth 1997-2007	10.584	2.616	0.039	6.541
Growth 1987-2007	21.378	7.536	2.443	13.632

	RIKS5	**RIKS6**	**RIKS7**	**RIKS8**
Average yearly growth	0.525	-0.147	-0.305	0.016
Growth 1987-1997	6.111	0.048	-1.581	2.950
Growth 1997-2007	4.653	-2.946	-4.416	-2.548
Growth 1987-2007	11.049	-2.900	-5.927	0.328