A Panel Cointegration Analysis of the Real Estate Market: Theory and Evidence from Spanish Provinces

Pedro Álvarez-Lois[†], Galo Nuño-Barrau[‡]
BBVA Economics Research Department

16 March 2007

Abstract

In this paper we analyze the dynamic adjustment of real house prices using data at the level of Spanish regions (*provincias*). We develop the micro-foundations of the model by using a dynamic general equilibrium approach where a house plays the dual role of a good that produces housing services and an asset that can be resold in a future date. We examine the extent to which real house prices at the regional level are driven by fundamentals by applying Panel Cointegration methods such as Common Correlated Effects or Dynamic Ordinary Least Squares to find a valid cointegration relationship. Then we try to analyze the role of the Real estate market in the economy by estimating a Vector Error Correction (VECM) Panel model. Results underline the importance of long-run adjustment process and persistence processes to explain the short-run dynamics of the price.

JEL Classification: C23, R21, R31

Keywords: Panel Cointegration, housing prices, adjustment dynamics, macroeconomic model

[†] E-mail: pedro.a_lois@grupobbva.com
[‡] E-mail:galo.nuno@grupobbva.com

1. Introduction

Housing is a key component of the wealth of the families and therefore it is an essential factor to understand expenditure patterns. Actually, according to Case, Quigley and Shiller (2001) and BIS (2002), wealth effects associated with real estate are more significant than those linked to financial asset holdings in most economies. Notwithstanding, wealth effects are not the only channel through which falls in real house prices affect activity. A reduction on private residential investment and loan losses in the banking sector due to inadequate provisions (Dobson and Hufbauer 2001) might also severely affect the economy, underlying the importance of understanding the dynamics beneath the movements of house prices in real terms.

One of the main challenges when analyzing the housing price is the dual nature of houses. The fact that a house is both a *good* that renders valuable housing services and an *asset*, in the sense that it is a durable good that may be sold in the future, implies that the housing market is composed by the whole range of individuals that go from the pure speculator, who buys the house only as an investment, to the pure owner, who is only concerned about the utility that he obtains from the possession of the house. Most of the literature[1] departs from this fact to construct asset pricing models *à la* Campbell and Shiller (1988) that exploit some intertemporal links between house prices, rents (if a sufficiently liquid rental market does exist) and interest rates. The final aim of these models is to shed some light on the relative importance of 'expectations' and 'fundamentals' in understanding a number of stylized facts observed in episodes of surge in house prices, including the positive correlation of the price with the volume of housing purchases for investment motives or the negative correlation with real interest rates.

There are a growing number of empirical studies that have tried to analyze the role in the price formation of fundamentals such as family incomes, demographic trends, unemployment rates, restrictions on building and reduction on interest rate[2]. In order to do that, they 'traditionally' employ three different families of econometric models. The Error Correction Model (ECM) approach is probably the most extended way to cope with the problem (Capozza et al. 2002, Jacobsen and Naug 2005, McCarthy and Peach 2002), as it is simple to estimate and it may offer additional information about the

[1] Some recent examples include Ayuso and Restoy (2006), Arce and López-Salido (2006), Brunnermeier and Juillard (2005) or Nagahata et al. (2004).

[2] A survey of the literature in OECD countries can be found in Girouard et al. (2006).

underlying dynamics and about evidences of 'euphoria' or backward-looking expectations (Capozza and Seguin, 1996). A second approach is to apply a Vector Autoregression (VAR) o Structural VAR (SVAR) framework to model the dynamic interaction between the housing price and some of its fundamentals, such as inflation, growth rate of GDP or real interest rates (Tsatsaronis and Zhu 2004). The third approach is less extended, and it consists of applying some kind of dynamic factor model to capture the co-movement between a group of observable time series (the price and its fundamentals) and a small number of unobservable variables, which try to capture global and local environment in the housing market and in the economy as a whole (IMF 2004).

Notwithstanding, there is a fourth family of econometric methods that may be used to estimate the relationship between price and fundamentals. They are panel cointegration methods, which apply cointegration techniques to find log-run relationships in panel data with a substantial number of cross-section units (Pedroni 1995). They pose the challenge of capture heterogeneity due to region-specific characteristics, which is usually solved by considering that the mean trend and the short-run dynamics may differ across regions, whereas the long-run relationship is the same for all. The reason for assuming a homogeneous long-run (or equilibrium) relationship is that the underlying economic principles that are employed to establish the equilibrium should apply similarly in all economies, whereas the adjustment process may differ due to behavioural and institutional features. In consequence it is convenient to have a well micro-founded economic relationship that may be tested as the equilibrium relationship.

Pedroni (1995, 2000) and Philips and Moon (1999) proposed an asymptotically efficient Panel Cointegration estimation that is based on the "fully modified OLS" (FMOLS) that accounts for the possible endogeneity of the regressors and serial correlations of the errors. Another possibility is the Dynamic OLS (DOLS) of Kao and Chiang (2000), which performs somewhat better. The problem with both estimators (FMOLS and DOLS) is that they may be severely biased in data with reduced number of time samples. Another shortcoming of these methods is that they do not fully capture another important feature of the panel data model: the possibility of contemporaneous correlation among cross section units, which cannot be captured by a time-specific random effect (Breitung, 2005).

In an attempt to overcome these shortcomings, Pesaran (2006a) has proposed the Common Correlated Effects (CCE) estimator, conceived to work in panel data models with a multifactor error structure where the unobserved common factors are correlated with exogenously given individual-specific regressors, and the factor loadings differ over the cross section units. This estimator offers a new approach to cope with the cross-section correlation, but it may also have problems when working in panels with small *T* (the number of time series observations).

When working with ‘short *T*’ panels, as it is the case presented in this paper, the parametric Vector Error Correction Model (VECM) proposed by Breitung (2005) seems to outperform DOLS and FMOLS. Results presented in this paper also indicate that it works better than the CCE for the sample under consideration. The estimator is based on a two-step cointegrated VAR model that may account for contemporaneously correlated errors.

The aim of this paper is to analyze the relationship between house prices and their fundamentals, especially in cases when the rental market is underdeveloped, so approaches *à la* Campbell and Shiller (1988) may not be appropriated. In order to do so, in Section 2 we develop a theoretical model that departs from the literature in Dynamic General Stochastic Equilibrium (DGSE) models such as Heathcote and Davis (2005) to derive a supply-demand equation for the housing market. This equation allows for multiple (‘sunspot’) equilibrium and takes into account both the expectation about future housing prices and fundamentals such as interest rates, consumption or stock of houses per capita. Then in Section 3 we briefly introduce panel cointegration methods[3] and discuss which one may be more suitable for the problem under consideration. As commented above, the advantage of panel cointegration methods is that they are a useful tool to exploit all the information in the diverse regions to obtain a long-run or cointegration relationship while allowing them to have different short-run responses.

In Section 4 we present the database we have constructed to analyze the Spanish housing market. The reason to choose Spain is double. On the one hand, Spain has a very rich regionalised information due to its fairly decentralised political structure (especially since 1995) and consequently it is relatively easy to gather data about regional GDP, unemployment, population or stock of houses for its 50 regions (“*provincias*”). On the other hand, Spain presents a fairly illiquid rental market as most of the families decide to purchase rather than to rent (Ruiz and San Martín 2004); therefore Spain is a good candidate to empirically test the results of our model. Additionally, Spain has been one of the industrialised economies that have experienced the housing market boom since the late 1990s so it may be considered as good benchmark to analyse this important phenomenon[4].

[3] A complete review of the State-of-the-Art in panel cointegration may be found in Breitung and Pesaran (2007)

[4] The literature of panel cointegration analysis of the housing also includes the US market (Holly, Pesaran and Yamagata 2006) and the Japanese one (Nagahata et al. 2004)

Maza and Villaverde (2007) have analyzed the shocks affecting the Spanish regions between 1975 and 2005. To do that, they have applied a relatively new methodology which has made it possible to offer new insights on this issue. The most relevant conclusion is that, during this period of ever increasing globalization, the Spanish regions have been mainly affected by symmetric shocks. As a result, one should expect a great deal of cross-section dependence in the data. The econometric methods applied in the analysis should take into account this feature of the data.

In Section 5 we apply two single equation estimators, the DOLS and the CCE, to obtain the vector of cointegration. The CCE was better a priori for this problem, as it was expected that the housing price series were contemporaneously correlated. However, results show that it is not able to find any significant relationship, probably due to the short time length of the series, which range 1995:2005. The DOLS offer some interesting results both in the long and short runs and it validates the proposed theoretical equation as a valid cointegration relation.

In Section 6 we apply a modification of the two-step VECM system estimator presented by Breitung (2005). This parametric approach has the advantage of accounting for cross-section correlation, outperforming other estimators in 'short *T*' panels, and modelling the whole real estate market (supply, demand and price). Moreover, feed-back effect amongst the variables can be accounted for in this set-up Results for the long-run elasticities are in line with the theoretical model and the previous literature. An analysis of the contributions of the different variables show that the "disequilibrium" term has played an important role in explaining the short-run dynamics of the price of housing and interestingly the growth of per-capital GDP.

In Section 7 we finally conclude and suggest some future lines of research in this interesting topic.

2. A Model of Housing and the Macro-Economy

In this section, we introduce a dynamic macro-economic model that takes explicit account of the housing market. The objective is two fold: firstly, to show the conditions under which the real price of housing is non-stationary and, secondly to analyse in a coherent way the determinants of the real price of housing and its interaction with the rest of variables in the economy.

In the model, the economy is populated by identical, infinitely-lived households. The population grows at a constant gross rate h so that in what follows all variables are in per-capita terms. The economy is made up of two productive sectors. One sector produces final goods that can either be consumed or invested in productive capital. The other sector produces houses. Both sectors differ in the level technology as well as in the quantity of inputs used for production. For simplicity we do not consider the government in the model. Moreover, we assume perfect competition in all markets and absence of real as well as nominal frictions.

2.1 Firms

We consider two types of inputs, namely labour and physical capital. Let P_t^h be the price of houses in terms of the final good. Let w_t and r_t^k be, respectively, the competitive wage and rental rate on capital measured in the same units of the final good. The representative housing-producing firm maximises profits according to

$$\max P_t^h Y_t^h - w_t N_t^h - r_t^k K_t^h \quad \text{s.t.} \tag{2.1}$$

$$Y_t^h = A_t^h \left(K_t^h\right)^{a_h} \left(N_t^h\right)^{1-a_h} \tag{2.2}$$

where A_t^h is the level of total factor productivity in the construction sector. We assume that this variable follows a non-stationary unit-root process

$$A_t^h = A_{t-1}^h \exp\left(\Phi^h + e_t^h\right) \tag{2.3}$$

where e_t^h is a white noise shock and Φ^h is the long-run rate of growth of the increase in the technology factor. The first order conditions associated with (2.1) imply that the optimal demand of inputs equalises their price to the marginal productivities:

$$a_h P_t^h A_t^h \left(K_t^h\right)^{a_h - 1} \left(N_t^h\right)^{1-a_h} = r_t^k \tag{2.4}$$

and

$$\left(1 - a_h\right) P_t^h A_t^h \left(K_t^h\right)^{a_h} \left(N_t^h\right)^{-a_h} = W_t \tag{2.5}$$

Similarly, firms in the final-goods sector choose optimal demand of production factors in order to maximise their profits, that is,

$$\begin{aligned} &\max Y_t^c - w_t N_t^c - r_t^k K_t^c \\ &s.t. \end{aligned} \tag{2.6}$$

$$Y_t^c = A_t^c \left(K_t^c\right)^{a_c} \left(N_t^c\right)^{1-a_c} \tag{2.7}$$

where A_t^c is the level of total factor productivity in the final-goods sector. As before, we assume that this variable follows a non-stationary unit-root process

$$A_t^c = A_{t-1}^c \exp\left(\Phi^c + e_t^c\right) \tag{2.8}$$

where e_t^c is a white noise shock and Φ^c is the long-run rate of growth of the increase in the technology factor. The first order conditions associated with (2.6) are

$$a_c A_t^c \left(K_t^c\right)^{a_c - 1} \left(N_t^c\right)^{1-a_c} = r_t^k \tag{2.9}$$

and

$$\left(1 - a_c\right) A_t^c \left(K_t^c\right)^{a_c} \left(N_t^c\right)^{-a_c} = W_t \tag{2.10}$$

Now, notice that (2.5) combined with (2.2) can be expressed as

$$\left(1 - a_h\right) P_t^h \frac{Y_t^h}{N_t^h} = W_t \tag{2.11}$$

Similarly, combining (2.10) with (2.7) we get

$$\left(1 - a_c\right) \frac{Y_t^c}{N_t^c} = W_t \tag{2.12}$$

Now, from (2.11) and (2.12) we obtain a relationship for the real price of houses:

$$P_t^h = \frac{(1-a_c)}{(1-a_h)} \frac{Y_t^c}{Y_t^h} \frac{N_t^h}{N_t^c} \qquad (2.13)$$

It can be shown that the production of Y_t^h has a trend growth rate of $\Phi^h / (1-a_h)$ and that trend growth of final goods production is $\Phi^c / (1-a_c)$. Moreover, employment has no trend growth in this model. Hence, taking into account (2.13), the real price of housing would be non-stationary whenever the trend growth rates of the two productive sectors differ. For instance, if the final good sector in more capital intensive than the construction sector, that is, $a_c > a_h$ and that technology grows more slowly in the latter sector, that is, $\Phi^h < \Phi^c$ then the trend growth rate of the real price of housing would be positive. The intuition behind this result is that the supply of housing would be insufficient to meet the demand.

2.2 Households

Next, we introduce households in the model. Specifically, a representative household supplies homogeneous labour and rents capital to the two productive sectors in the economy. The representative household derives utility each period from per-capita consumption C_t, from per-capita housing owned H_t and from leisure. The amount of per-household member labour supplied plus leisure cannot exceed the period endowment of time, which is normalised to 1. Period utility per household member at date t is assumed to be given by

$$U(C_t, H_t, (1-N_t)) = \log C_t + g \log H_t + r \log(1-N_t) \qquad (2.14)$$

where $\gamma > 0$ is the utility weight of housing and $\rho > 0$ is that of leisure. The functional form of the instantaneous utility is motivated by theoretical results in Ngai and Pissarides (2004). These authors show that necessary and sufficient conditions for the existence of an aggregate balanced growth path in a multi-sector economy are logarithmic preferences and a non-unit price elasticity of demand.

At date 0, the expected discounted sum of future period utilities for the representative household is given by

$$E_0 \sum_{t=0}^{\infty} b^t h^t U(C_t, H_t, (1-N_t)) \qquad (2.15)$$

where $b < 1$ is the discount factor.[5] Households receive income from the supply of labour and capital. They also receive income from selling houses. Income is divided between consumption, spending on new capital that will be rented out next period, and spending on new housing that will be occupied next period. The depreciation rate for capital is given by d_k and d_h represents that of housing. Thus, the household's budget constraint is:

$$C_t + K_t + P_t^h H_t = W_t N_t + \frac{(1-d_h)}{h} P_t^h H_{t-1} + \frac{(1-d_K)}{h} K_{t-1} \tag{2.16}$$

The representative household chooses state-contingent values for consumption, hours, capital and housing for all t = 0 to maximize expected discounted utility (2.15) subject to a sequence of budget constraints (2.16) and a set of inequality constraints C_t, N_t, H_t, $K_t = 0$ and $N_t = 1$. The household takes as given prices, a probability distribution over future possible states, and the initial stocks of capital and housing. The first order conditions are the following:

$$U_{1,t} = b E_t \left[U_{1,t+1} \left(1 + r_{t+1}^k - d_k\right)\right] \tag{2.17}$$

where $U_{1,t}$ is the partial derivative of the instantaneous utility unction $U(\cdot)$ with respect to its first argument, that is, consumption. Equation (2.17) is the standard Euler condition for capital accumulation. The first order with respect to housing accumulation is given by

$$P_t^h U_{1,t} = U_{2,t} + b E_t \left[U_{1,t+1} P_{t+1}^h \left(1 - d_k\right)\right] \tag{2.18}$$

The intuition behind this equation is the following: investing in one additional unit of housing has a cost equal to the price of housing, P_t^h, times the number of consumption goods that cannot be consumed. On the other hand, investing in one unit of housing yields a direct utility to the household in the period the investment is made. The next period, the household can sell a portion $\left(1-d_k\right)$ of housing at a price P_{t+1}^h and buying with this consumption goods, that would report each a utility of $U_{1,t+1}$. The optimal amount of housing is the one that equalises the costs and benefits. Finally, the supply of labour is governed by the condition that the real wage rate is equal to the marginal rate of substitution between leisure and consumption, that is,

[5] Notice that the flow of utility that households receive from occupying housing they own will constitute an implicit rent that is untaxed.

$$W_t = \frac{U_{3,t}}{U_{1,t}} \tag{2.19}$$

2.3 Equilibrium

Equality between supply and demand of production inputs lead to equilibrium in these two markets, that is,

$$N_t^h + N_t^c = N_t \tag{2.20}$$

and

$$K_t^h + K_t^c = K_{t-1} \tag{2.21}$$

Moreover, in equilibrium, the total stock of housing that may be enjoyed by households evolves according to

$$H_t = Y_t^h + \frac{(1-d_h)}{h} H_{t-1} \tag{2.22}$$

Next, equilibrium in the final goods market is

$$Y_t^c = C_t + K_t + \frac{(1-d_k)}{h} K_{t-1} \tag{2.23}$$

Finally, gross domestic product in this economy is given by

$$Y_t = Y_t^h + Y_t^c \tag{2.24}$$

2.4 Qualitative Analysis

Once we have presented the model and stated the equilibrium of the economy, the natural next step would be to provide numerical values to the parameters and proceed to solve and simulate the model. However, given that the empirical analysis offered in Section 5 is based on a single equation approach, it might be interesting to first present some partial equilibrium results. To that end, we shall focus on equation (2.18), which relates the real price of housing to fundamental variables in the economy and which for convenience is restated here:

$$P_t^h U_{1,t} = U_{2,t} + b\, E_t\left[U_{1,t+1} P_{t+1}^h (1-d_k)\right]$$

Next, given the assumption that the instantaneous utility adopts a logarithmic form, we have that

$$P_t^h = g \frac{C_t}{H_t} + b E_t \left[\frac{C_t}{C_{t+1}} P_{t+1}^h \left(1 - d_k\right) \right] \tag{2.25}$$

The next step is to log-linearise (2.25) around a steady state. Given that there is growth in the model due to the non-stationary technology factors, we have to de-trend the variables. Hence, we define the de-trended real price of hosing as $p_t^h \equiv P_t^h Z_t^h / Z_t^c$, where $Z_t^c \equiv \left(A_t^c\right)^{1/(1-a_c)}$ and $Z_t^h \equiv \left(A_t^h\right)^{1/(1-a_h)}$; similarly, de-trended consumption $c_t \equiv C_t / Z_t^c$ and the de-trended stock of houses is $h_t \equiv H_t / Z_t^h$. Equation (2.25) thus becomes:

$$p_t^h = g \frac{c_t}{h_t} + b E_t \left[\frac{c_t}{c_{t+1}} \frac{p_{t+1}^h}{z_{t+1}^h} \left(1 - d_k\right) \right] \tag{2.26}$$

where $z_{t+1}^h \equiv Z_{t+1}^h / Z_t^h$, which by the definition of Z_t^h given in (2.3) is equal to $z_{t+1}^h \equiv \left[\exp\left(\Phi^h + e_{t+1}^h\right) \right]^{1/(1-a_h)}$. Next, we denote a variable in log deviation from its steady state value with a "^" and proceed to log-linearise (2.26), leading to

$$\hat{p}_t^h \approx g \frac{c}{p} \hat{c}_t - g \frac{h}{p} \hat{h}_t + \left(1 - d_k\right) b E_t \left(\hat{p}_{t+1}^h \right) + \frac{\left(1 - d_k\right) b}{p} E_t \left(\hat{c}_t - \hat{c}_{t+1} \right) \tag{2.27}$$

In order to gain some intuition on equation (2.27), notice that in this economy we can price any asset using the Stochastic Discount Factor, that is, the Euler equation associated with the Lagrange multiplier corresponding to the Household's budget constraint. It is thus straightforward to show that

$$\hat{r}_t \approx E_t \left(\hat{c}_{t+1} - \hat{c}_t \right) \tag{2.28}$$

where $\hat{r}_t$ is the net return on a risk-less one-period discount bond. Hence, after combining (2.28) with (2.27) we arrive to

$$\hat{p}_t^h \approx g \frac{c}{p} \hat{c}_t - g \frac{h}{p} \hat{h}_t - \frac{\left(1 - d_k\right) b}{p} r_t + \left(1 - d_k\right) b E_t \left(\hat{p}_{t+1}^h \right) \tag{2.29}$$

This expression shows that, ceteris paribus, the real de-trended price of housing depends positively on consumption and on its expected future value; whereas there is a negative relationship with the stock of houses and with the real interest rate. It would e interesting to analyse the role of expectations in the dynamic of the real price of housing. To that end, we keep with the partial equilibrium analysis and focus the attention on Equation (2.29), which we reformulate as

$$\hat{p}_t^h = (1 - d_k)\, b\, E_t\left(\hat{p}_{t+1}^h\right) + \Gamma u_t \tag{2.30}$$

where u_t is a vector of fundamental variables. For simplicity, we assume that the fundamental variables follow a non-persistent exogenous process. Next, we re-write (2.30) as a system of first order matrix difference equations in the two endogenous variables $\hat{p}_t^h$ and $E_t\left(\hat{p}_{t+1}^h\right)$:

$$\begin{bmatrix} 1 & -b\,(1-d_h) \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{p}_t^h \\ E_t\left(\hat{p}_{t+1}^h\right) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{p}_{t-1}^h \\ E_{t-1}\left(\hat{p}_t^h\right) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Gamma u_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_t \tag{2.31}$$

where w_t is a non-fundamental term related to expectations errors and it is defined as the difference between the observed price in period *t* and the expected price in *t-1*. Pre-multiplying (2.31) by the inverse of the left-hand side matrix, we obtain:

$$\begin{bmatrix} \hat{p}_t^h \\ E_t\left(\hat{p}_{t+1}^h\right) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1/a \end{bmatrix} \begin{bmatrix} \hat{p}_{t-1}^h \\ E_{t-1}\left(\hat{p}_t^h\right) \end{bmatrix} + \begin{bmatrix} 0 \\ -1/a \end{bmatrix} u_t + \begin{bmatrix} 1 \\ 1/a \end{bmatrix} w_t \tag{2.32}$$

where $a \equiv b\,(1-d_h)$. The dynamics of the system would depend on the roots of the right hand side matrix. The characteristic polynomial is

$$F^2 - \frac{1}{a} F = 0 \tag{2.33}$$

In this case, there are two roots which we call q and l. One of this roots, let's say q, is equal to zero and the other one, namely l is equal to $1/a$. Notice that if the equilibrium is unique, there must be one un-stable root that allows one to pin-down the non-predetermined variable $E_t\left(\hat{p}_{t+1}^h\right)$ as a function of the lagged state variable $\hat{p}_{t-1}^h$ and the fundamental variable u_t. For this condition to be satisfied $|l| > 1$, that is, $\left|1/b\,(1-d_h)\right| > 1$ which is satisfied in this model since we have assumed that $b > 1$ and

$0 \leq d_h \leq 1$. Finally, applying the methods developed by Beyer and Farmer (2006), it is possible to show that the solution to (2.31) is

$$\tilde{p}_t^h = \Gamma u_t \tag{2.34}$$

and

$$E_t\left(\tilde{p}_{t+1}^h\right) = 0 \tag{2.35}$$

Hence, the real de-trended price of housing is a linear function of the fundamentals. In this analysis we have implicitly assumed for simplicity that the fundamentals are not persistent. Hence, the economy is expected to return to its steady state immediately after a shock. That is why the expectation in (2.35) is equal to zero. Notice that the analysis so far has been of a partial character. One should analyse the full system dynamics. In this case, there might be combinations of the parameters such that the solution to the model would be indeterminate and thus non-fundamental or expectations-driven dynamics could not be precluded. We left the analysis of possible "sunspot" dynamics for future research.

3. Introduction to Panel Cointegration Models

A review of the panel counter part of the classical literature on cointegration techniques developed by Engle and Granger (1987), Johansen (1995) and Philips (1991) has been recently presented in Breitung and Pesaran (2007). Here we simple present a brief overview of the models that will be used in the next sections.

Consider the m time series $\mathbf{z}_{it} = (z_{i1t}, z_{i2t}, z_{i3t},\ldots, z_{imt})'$ where $z_{i1t} = p_{it}$ is the housing price and $\mathbf{x}_{it} = (z_{i2t}, z_{i3t},\ldots, z_{imt})'$ is a vector of fundamentals observed on the i^{th} region $i = 1, 2,\ldots, N$, over the period $t = 1, 2,\ldots, T$. Suppose that for each I

$$z_{ijt} : I(1), \quad j = 1,2,\ldots,m \tag{3.1}$$

Then $\mathbf{z}_{it}$ is said to form $r_i = 1$ cointegration relations if there are linear combinations of z_{ijt} for $j = 1, 2,\ldots, m$ that are $I(0)$ i.e. if there exists an $m \times r_i$ such that

$$\underset{r_i \times m}{b_i'} \quad \underset{m \times 1}{\mathbf{z}_{it}} = \underset{r_i \times 1}{e_{it}} : I(0) \tag{3.2}$$

3.1 Residual-Based Approaches

The residual-based approaches, such as FMOLS, DOLS and CCE are appropriate when $r_i = 1$, and $\mathbf{z}_{it}$ can be partitioned such that $\mathbf{z}_{it} = (p_{it}, \mathbf{x}'_{it})'$ with no cointegration among the m - 1 variables of $\mathbf{x}'_{it}$. The system cointegration methods are much more generally applicable and allow for $r_i > 1$ and do not require any particular partitioning of the variables in $\mathbf{z}_{it}$.

Residual-based approaches usually consider the following regression:

$$p_{it} = d_i'\mathbf{d}_{it} + \mathbf{x}_{it}' b + u_{it}, \quad i = 1,2,\ldots,N \tag{3.3}$$

where $d_i'\mathbf{d}_{it}$ represent the deterministic trend. The innovations in ?$\mathbf{x}_{it}$, denoted by $e_{it} = \Delta\mathbf{x}_{it} - E(\Delta\mathbf{x}_{it})$, are allowed to be cointegrated with u_{it}. It is assumed that the vector of coefficients, ß, is the same for all regions, that is, a homogeneous cointegration relationship is assumed.

Applying a sequential limit theory it can be shown that the OLS estimator of ß is $T\sqrt{N}$ consistent and, therefore, the time series dimension is more informative than the cross-section one on the long-run coefficients. However, the OLS estimator is inefficient in the model with endogenous regressors. To obtain an asymptotically efficient estimator, Pedroni (1995) proposed the FMOLS approach that adjusts for the effects of endogeneity and short-run dynamics of the errors (Philips and Hansen 1990).

An alternative approach is the DOLS estimator (Kao and Chiang 2000), which discomposes the error term

$$u_{it} = \sum_{k=-\infty}^{\infty} g'_{ik} \Delta \mathbf{x}_{i,t+k} + v_{it}, \quad (3.4)$$

where v_{it} is orthogonal to all leads and lags of ?$\mathbf{x}_{it}$. From (3.4) and (3.3) we obtain (assuming no time trend and allowing $\mathbf{x}_{i1} = 1 \ \forall i$)

$$y_{it} = b'\mathbf{x}_{it} + \sum_{k=-\infty}^{\infty} g'_{ik} \Delta \mathbf{x}_{i,t+k} + v_{it}. \quad (3.5)$$

In practice the infinite sums are truncated at some small number of k. Kao and Chiang (2000) show that in the homogeneous case the FMOLS and the DOLS have the same limiting distribution. Tough the DOLS estimator outperforms the FMOLS, both have a substantial bias if T is small (Breitung 2005).

Pesaran, Shin and Smith (1999) adapted a parametric model to estimate the cointegration vector based on the error correction format

$$\Delta p_{it} = f_i p_{it-1} + j'_i \mathbf{x}_{it} + v_{it}, \quad (3.6)$$

where it is assumed that the long-run parameters are identical across the cross section units, i.e., $b_i = -j_i / f_i = b$ for i =1, 2, …, N. This method is called the Pooled Mean Group (PMG) estimator. The potential problem of the three methods (FMOLS, DOLS and PMG) is that they assume no correlation between errors in different cross section units (regions).Cross-section dependence can arise due to a variety of factors, such as omitted observed common factors, spatial spill over effects, unobserved common factors, or general residual interdependence that could remain even when all the observed and unobserved common effects are taken into consideration.

A Multifactor Residual Model has been proposed by Pesaran (2006a) where errors u_{it} in (3.3) have the multifactor structure

$$u_{it} = ?'_i \mathbf{f}_t + h_{it} \quad (3.7)$$

In which $\mathbf{f}_t$ is the $m \times 1$ vector of unobserved common effects and $?_{it}$ are the region-specific errors assumed to be independently distributed of ($\mathbf{d}_t$,$\mathbf{x}_t$). A further extension of this model includes the potential correlation of ($\mathbf{d}_t$,$\mathbf{x}_t$) and $\mathbf{f}_t$ by considering the general model for the individual specific regressors

$$\mathbf{x}_{it} = \mathbf{a}_i + \Gamma'_{it}\mathbf{f}_t + \mathbf{v}_{it} \tag{3.8}$$

where $\mathbf{a}_i$ is the $k \times 1$ vector of individual effects, G_i is a $k \times m$ factor loading matrix with fixed components and $\mathbf{v}_{it}$ are the specific components of $\mathbf{x}_{it}$ distributed independently of the common effects and across *I*, but assumed to follow general covariance stationary processes.

Under this multifactor model, Pesaran (2006a) demonstrate that it is possible to obtain Common Correlated Effects (CCE) estimators of the cointegration vector for both the homogeneous and heterogeneous cases that are consistent regardless of whether the common factors ft are stationary or nonstationary. He also proposes a Corss-section Dependence (CD) test of error cross dependence, which does not require an a priori specification of a connection matrix and that is applicable to short *T* and large *N* panels. He also proposes a panel unit root test called CIPS (Pesaran 2006b).

3.2 System Estimators

The single equation estimators exposed above assume that all the regressors are *I*(1) and not cointegrated, which can be avoided by using a system approach. Besides, they usually perform poorly in short *T* panels. To cope with these problems, Groen and Kleibergen (2003) and Breitung (2005) consider the Vector Error Correction Model (VECM) for the *m* dimensional vector $\mathbf{z}_{it}$ given by

$$\Delta\mathbf{z}_{it} = a_i b'_i \mathbf{z}_{it} + \mathbf{w}_{it}, \tag{3.9}$$

where $\mathbf{w}_{it} = \left(u_{it}, e'_{it}\right)'$. Breitung (2005) proposed a computationally convenient two-step estimator based on the fact that the Fisher information is block-diagonal with respect to the short- and long-run parameters. Accordingly, an asymptotically efficient estimator can be constructed by estimating the short- and long-run parameters independently. Monte Carlo experiments performed to compare this estimator with the FMOLS and DOLS suggest that the latter may be severely biased in small samples whereas the bias of the two-step estimator is relatively small.

4. Panel Data on Spanish House Prices

4.1 Preliminary Data Analysis

We have constructed panel data for the Spanish real house prices and its fundamentals for the 50 Spanish *provincias*. The real house price has been obtained as the nominal house price deflated by the Consumer Price Index (CPI) of each *provincia*. The nominal house price is the average price per square meter of houses in the market[6], as reported by the Spanish Ministry of Housing, the CPI has been obtained from the Spanish National Statistics Institute/ *Instituto Nacional de Estadística* (INE). Data on the stock of houses has been constructed by combining the number of new houses in the market provided by the Ministry of Housing and the stock of houses in 2001 from the INE[7]. Regional GDP (also deflated by the CPI) and population figures include the effect of immigration (quite important in Spain since the late 1990s) and also come from the INE. Nominal interest rates are the average rates for new mortgage credits as reported by the Spanish Central Bank. Real interest rates are constructed with the nominal rates and the CPI. We have also constructed an index on real rental prices by deflating by the CPI an index derived from the house renting chapter of the CPI, defining 1995 as base year (with value equal to 1). To account for portfolio diversification we include the yearly returns[8] from the IBEX-35, the Spanish capitalization-weighted stock market index. A list of the variables under consideration is presented in Table 1.

In Figure 1 we show the evolution of the YoY increments in the real price of housing. The total increment in the housing price in the period 1995:2005 was of 95%. The supply of houses is shown in Figure 2, reflecting how some regions (as Madrid, for example) have experience a surge in the number of built houses, whereas others (such as Barcelona) have kept a more or less constant rate of increase in the stock of housing.

4.2 Panel Unit Root Test Results

Once the panel data have been constructed, it is convenient to test whether the price and fundaments are $I(1)$ and if there is significant cross section dependence. Following Holly, Pesaran and Yamagata (2006) we first compute cross-dependence (CD) test

[6] We distinguish between houses in the market and public houses, the latter provided by the State at prices well below their market value

[7] The correct magnitude should have been the number of square meters (as this is the item being priced). We are implicitly assuming that new houses have the same number of square meters than previous ones. This could be wrong due to demographical issues, such as the reduced number of family members in the new generations, but it may be considered as a first-order approximation. An additional problem is that we are not accounting for depreciation of the houses.

[8] Year-over-year (YoY) returns in December.

estimates of the p[th]-order Augmented Dickey-Fuller test statistics for p_{it}, y_{it}, h_{it}, r_{it}, ib_{it} and ren_{it}[9]

The CD statistics reported in Table 2 clearly show that the cross correlations are statistically significant, and thus invalidate the use of first generation panel unit root test that do not allow for error cross section dependence. It is necessary to apply a second generation test such as the CIPS test proposed by Pesaran (2006b)[10].The CIPS test results, summarized in Table 3, show that for p_{it}, y_{it}, h_{it} and ren_{it} the unit root hypothesis cannot be rejected if the trended nature of these variables are taken into account. The problem arises as the test is not able to reject the unit root hypothesis either for the real interest rates or for the first difference of the variables[11]. For the case of real house prices, the unit root hypothesis cannot be rejected for the variable itself, but it can be rejected at a ten per cent significance level for its first difference, so it can be assumed to be I(1). As the test results present a high degree of uncertainty due to the short temporal length and due to the economic considerations commented in Section 2, we assume y_{it}, and h_{it} to be I(1) and r_{it}, I(0).

5. Residual-Based Panel Cointegration

5.1 Dynamic Ordinary Least Squares

To test for a possible cointegration relation as the one derived in Equation (2.34)[12] between the price p_{it} and its fundamentals y_{it}[13], h_{it}, and r_{it}, we estimate the model

$$p_{it} = \alpha_i + \beta_1 y_{it} + \beta_2 h_{it} + \beta_3 r_{it} + u_{it}, \quad i = 1,2,\ldots, N; \quad t = 1,2,\ldots, T, \tag{3.10}$$

where

$$u_{it} = \sum_{k=-q}^{q} \gamma_k' \Delta \mathbf{x}_{i,\, t+k} + v_{it}. \tag{3.11}$$

---

[9] Due to the reduce number of time periods, we should reduce the analysis to a maximum $p = 2$.

[10] Due to the reduce number of time periods, we should reduce the analysis to $p = 1$.

[11] A potential explanation is due to the reduced length of the time period, which is inferior to a whole economic cycle of the Spanish economy (the last recession was in 1993). Consequently, some 'traditionally' stationary variables such as real interest rates (as in Holly, Pesaran and Yamagata 2006) appear to be non stationary, as in Figure 3.

[12] We have not shown how we can pass from the de-trended equation to the trended one. Altough it might be shown , from now on let's assume that this is a proposed cointegration relationship that we want to test independently of its micr economic foundations.

[13] Income per capita (GDP per capita) is employed as a proxy for consumption due to the lack of data.

This is an adapted version of Equation (3.4) (Kao and Chiang 2000) following Breitung and Pesaran (2007) where $?_{ik} = ?_k$. The reason to simplify (3.4) is that there are not enough degrees of freedom to allow for potential heterogeneity of the $?_{ik}$ due to the sample size. Also due to sample size limitations, we should restrict $q = 1$[14]. As commented above, the reason to choose a DOLS estimator instead of a FMOLS is that DOLS has been shown to outperform the latter in Monte Carlo simulations, and both yield asymptotically the same distribution.

Results presented in Table 4 show a coefficient on income of 0.8, which is relatively close to previous literature where it approximately ranges from 0.3 to 3 according to the review of Girouard et al. (2006), under different definitions, data and econometric methods. Holly, Pesaran and Yamagata (2006) propose a value of 1. The coefficient on stock of housing *per capita* is -0.31 (Girouard et al. (2006): the elasticity relatively to housing stock ranges from -0.5 to -8). The coefficient on real interest rates is -8.43 (Girouard et al. (2006): from -0.1 to -9.4). All the variables are significant[15] above the 1 per cent level.

The CD statistic demonstrate that there is a high degree of cross-dependence: as it was mentioned above, this method does not explicitly account for cross-section correlation, which we will try to correct by applying the CCE estimator in the next Subsection. Nevertheless, CIPS test reject the null hypothesis of unit root below the 1 per cent; so Equation (3.10) seems to be a valid cointegration relation to describe the real price of housing.

Results without including the stock of housing per capita yield smaller income elasticity and higher semi-elasticity to real interest rates, but in general demonstrate that our results are somewhat robust.

5.2 Common Correlated Effects

To take into account cross-section dependence, we have decided to run a CCE estimator following Equations (3.7) and (3.8), where we estimate Equation (3.10) with

$$u_{it} = \sum_{k=1}^{q} g_{ik} f_{kt} + v_{it}, \tag{3.12}$$

[14] This is not as short as it may seem; for example Stock and Watson (1993) chose q equal to 2 for the period 1900-1989, which is significantly longer than 1995-2005.

[15] It is important to mention that the standard errors presented have been reescaled by the method presented in Stock and Watson (1993) so the *t*-statistic tends asymptotically to a $N(0,1)$

where f_{lk} can be either stationary or non stationary as long as v_{lt} is stationary and k is a fixed number (Pesaran 2006a).

Results are shown in Table 5. They totally lack significance, as most elasticities are close to zero. In consequence the estimator fails to find a successful cointegration relationship. The reason of this is the sample size: to be able to estimate 3 cointegration coefficients, the estimator generates a total of 8 coefficients per cross-section unit that should be estimated with 11 samples.

5.3 Short-Term Dynamics

Having established by DOLS a panel cointegration relation between the price and its fundamentals, we may turn our attention to the dynamics of the adjustment and estimate the panel error correction model:

$$\Delta p_{it} = a_i + j_i\left(u_{it-1}\right) + q_i \mathbf{W}\Delta\mathbf{p}_{t-1} + d_{1i}\Delta y_{it} + d_{2i}\Delta h_{it} + d_{3i}\Delta r_{it} + d_{4i}\Delta ren_{it} + d_{5i}\Delta ib_t + u_{it} \qquad (3.13)$$

The coefficient f_i provides a measure of the speed of adjustment of house prices to a shock. The half life of a shock to p_{it} is approximately $-ln(2)/ln(1+f_i)$. u_{it-1} is the lagged residual from Equation (3.10). **W** is a $1\times N$ vector of weights w_i so that

$$w_i = \frac{GDP_{i2005}/Pop_{i2005}}{\sum_{i=1}^{N} GDP_{i2005}/Pop_{i2005}} \qquad (3.14)$$

and $\mathbf{p}_t = \left[p_{1t}, p_{2t}, K, p_{Nt}\right]'$ is the vector of real prices. The term $?_i\mathbf{W}?\,\mathbf{p}_{t-1}$ tries to capture the possibility of spill-over effects of the price, as if the price of the houses rises in the richest per capita regions, it may be logical to assume that some of their residents will purchase houses in cheaper regions, both as a capital investment and as a second residence. We also include the effect of portfolio diversification ib_t and the impact of the rental market. ren_{it}. Equation (3.13) is estimated by OLS regressions separately for each *provincia* and for the homogeneous pool of all regions, including fixed effect results (considering the same coefficients for all the regions, but allowing for independent intercepts). Results for the pooled case are presented in Table 6 and examples for the heterogeneous case in Table 7.

For the case of the pooled model, results show the little importance of rent prices and income, and the impact of the long term adjustment and the housing per capita and interest rates. It is important to notice the portfolio diversification effect, as the stock market returns are negatively correlated with house prices (after accounting for the rest of variables). Test statistics still show a high degree of cross-dependence and the null hypothesis of unit root can be rejected with a significance below the 1 per cent.

6. The Panel-VECM Approach

As discussed above, although DOLS and FMOLS approaches are an elegant way to estimate non-stationary panel data models, they may be problematic especially in fairly small samples. In particular, the FMOLS estimator may be severely biased in empirically relevant sample sizes (Pedroni 2000). Another problem with these methods is that they are based on a single equation approach. Consequently, feed-back effects cannot be modelled in this set-up. For all these reasons, a parametric approach may be a promising alternative, in particular, for panels with a small number of time periods. Pesaran et al. (1999) have suggested estimation procedures for cointegrated panel data based on a vector error correction (VECM) format. However, these methods are based on a ML estimator and may have problematic small–sample properties as well as convergence issues.

Following Breintug (2005), we apply a simple asymptotically efficient two-step estimation procedure. The individual specific parameters are estimated in the first step, whereas, in a second step, the common long-run parameters are estimated from a pooled regression. The resulting estimator is asymptotically efficient and normally distributed.[16] Moreover, since the second step of the parametric approach is based on an ordinary least-squares regression, it is straightforward to account for possible contemporaneous correlation among the errors.

6.1 The set-up

We consider a cointegrated VAR(1) model with individual short-run dynamics and deterministic terms. As usual in the panel cointegration framework, we assume that the mean (or trend) and the short-run dynamics may differ across provinces, whereas the long-run relationship is the same for all provinces.[17] The model takes the following general form:

16 The results of some Monte Carlo experiments suggest that the two-step estimator performs better than the FMOLS and DOLS estimator in typical sample sizes.

17 The reason for assuming a homogenous long-run relationship is that the underlying economic principles that are employed to establish the long-run equilibrium should apply similarly in all provinces,

$$\Delta y_{it} = \Psi_i d_{it} + a_i b' y_{it-1} + \Gamma_{ij} \Delta y_{it-1} + u_{it} \qquad (3.15)$$

Where $y_{i,t}$ is a $k \times 1$ vector of endogenous variables that include the real price of housing, per-capita GDP, number of houses per capita and the real interest rate; d_{it} is a vector of deterministic variables (a constant in this case) and Ψ_i is a $k \times j$ matrix of unknown coefficients. u_{it} is an k-dimensional white noise error vector with $E(u_{it}) = 0$ and positive definite covariance matrix $\Sigma_i = E(u_{it}, u_{it})$. The term $b' y_{i,t}$ captures long-run relationships amongst the variables in the model. In this specification, the $k \times r$ ($0 < r < k$) cointegration matrix b is the same for all cross section units, whereas the $k \times r$ loading matrix $a I$ and the error covariance matrices Σ_i are allowed to vary across i.

Following Breintug (2005) we implement a two-step estimation procedure. Since the information matrix of the Gaussian likelihood is asymptotically block diagonal with respect to the "short-run parameters" (a_i, Σ_i) and the matrix of cointegration vectors b, the latter can be estimated efficiently based on some consistent initial estimator of $a I$ and Σ_i ($i = 1, \ldots\ldots, N$). Hence, in the first step, we compute a consistent estimator (as T ? 8) of a_i from estimating separate models for all N cross section units. In our empirical analysis, we restrict ourselves to just one cointegration relationship and, thus, use the two-step estimator suggested by Engle and Granger (1987).[18] At the first estimation stage, the restriction that the cointegration vectors are the same for all cross section units is ignored, but this does not affect the asymptotic properties of the final estimator of b.

At the second stage, the system is transformed such that the cointegration matrix b can be estimated by the ordinary least-squares of the *pooled* regression

$$\hat{z}_{it} = b' y_{i,t-1} + \hat{v}_{it} \qquad \text{for } i = 1, \ldots\ldots N \text{ and } t = 1, \ldots\ldots T.$$

where $\hat{z}_{it} = (\hat{a}_i' \hat{\Sigma}_i^{-1} \hat{a}_i)^{-1} \hat{a}_i' \hat{\Sigma}_i^{-1} \Delta \tilde{y}_{it}$ where $\Delta \tilde{y}_{it} (\tilde{y}_{i,t-1})$ denote the residual vectors from the least squares regression of $\Delta y_{it} \left(y_{i,t-1} \right)$ on $\Delta y_{it-1}, \ldots\ldots, \Delta y_{it-p+1}$ and d_t. Finally, $\hat{v}_{it}$ is defined as $\hat{v}_{it} = (\hat{a}_i' \hat{\Sigma}_i^{-1} \hat{a}_i)^{-1} \hat{a}_i' \hat{\Sigma}_i^{-1} u_{it}$. Breitung (2005), based on a sequential limit theory, shows that the two-step estimator has a normal limiting distribution. From this result, it follows that the long-run parameters are asymptotically normally distributed and, therefore, inference on the cointegration parameters involves standard limiting

whereas the adjustment process towards the long-run equilibrium may differ due to behavioural and institutional characteristics.

18 The ML estimator of Johansen (1991, 1995) could also be used.

distributions. To estimate the covariance matrix of b^{2S}, the finite sample moments of $y_{i,t-1}$ and $\hat{v}_{it}$ can be used and, therefore, the ordinary t statistics of the elements of b^{2S} have a standard normal limiting distribution.

6.2 Results from the Spanish Housing Market

In the present analysis we decided to keep the real interest rate as an exogenous explanatory variable in the Panel-VECM. Given that this latter variable is formed by a common nominal interest rate minus the province specific inflation rate, we consider it as weakly exogenous. The first step in the analysis is to check the non-stationary properties of the data used. To that end, we apply panel unit-root tests. As discussed in Section 4, the series are clearly I(1). Next, we proceed with the estimation of the individual VECM models. Given that the number of endogenous variables is 3, there might be at most 2 cointegration relationships. Even thought it is possible to test the exact dimension of the cointegration space (see Theorem 2 in Breitung, 2005) we assume it is of dimension one. Next, and taking into consideration this restriction, we normalise the cointegration vector such that the coefficient associated with the real price of houses is fix to one.

We apply the two-step estimator suggested by Engle and Granger (1987) to each province. Table 8 shows the estimated loading factors, that is, the vector of parameters $\hat{\alpha}_i$. For concreteness, we just show the elements corresponding to the house price equation. The parameters are thus an indicator of the speed of adjustment of house price to the corresponding “equilibrium” values. To facilitate the comparison, we show the ratio of each loading factor with respect to the mean of the provinces. The results show that there is not a clear pattern. Some provinces, such as Baleares, Barcelona, Malaga and Navarra, appear to be very dynamic, whereas Albacete, León, Gerona, Valladolid are relatively static.

Next, we proceed with the second step of the estimation procedure and compute the long-run coefficients. Table 9 shows the estimates of the common coefficients with the associated t-statistics. The numbers can be regarded as long-run elasticity. The estimated coefficient of real GDP is positive and significant. In particular, it is 1.24 which is a value in line with the estimates of Holly et al (2006) for the U.S. cities. Regarding the elasticity of the number of houses per capita, the sign is negative but the level of significance is not very high. Finally, the price of houses appears to respond negatively to the real interest rate and very strongly. The estimated coefficient is -2.50 and the level of significance is high.

The analysis so far has assumed that there is one co-integration relationship between the variables in the model. A full assessment of this assumption would require the application of properly constructed tests, a task that is currently out of the scope of the present paper. Nevertheless, we might gain some confidence in the results by analysing the stationary properties of the residuals of the long-run equation. To that end, we first study the correlation properties of the residuals. The cross-section dependence test yields a value of -2.05. Given that the CD-test is asymptotically normally distributed, we might reject the hypothesis of no cross-section dependence. Hence we should apply the CIPS panel unit-root test. The estimated CIPS is -2.5, whereas the truncated CIPS is -2.2. These numbers indicate that we can confidently reject the null of I(1) and thus a stationary relationship seems to exist among house prices, real GDP, the number of houses and the real interest rate.

An interesting exercise is to compute the contributions, in the short-run dynamic equation, of the different variables in the model. In this regard, Figure 4 and Figure 5 show the contributions to the growth in the real price of housing for Madrid and Barcelona respectively. In these cases, the disequilibrium or error-correction term has contributed positively to the increase in the price of housing between the years 1997 and 2003. In the last part of the sample, the contribution has been lees pronounced, even negative.

Another interesting result is the contributions to the growth of real GDP per capita. We can see from Figure 6 and Figure 7 that the disequilibrium term has explained a large portion of the growth experience in the first half of the sample in these two provinces. It should be noticed how during the first half of the period, real GDP growth was substantially fuelled by the fact that the observed price has been above its equilibrium value in a sort of wealth effect, whereas in the second half, a considerable amount of the GDP growth is due to persistence effects of the price. It is also interesting to see that the contribution of the number of houses per capita has dragged to output growth, indicating a possible deficiency in the supply of housing.

7. Concluding Remarks

This paper shows a formal analysis of the interaction between house prices and some fundamentals such as consumption, stock of houses or interest rates. The theoretical relations have been developed departing from the framework of DGSE models and they yield some interesting results, such as the possibility of ever rising real housing prices as long as the consumption and capital technology may grow faster than the real state one. This framework also allows developing a de-trended equilibrium equation that links fundamentals with housing prices. This analysis was of a partial equilibrium nature and should be extended to include general equilibrium effects.

Using different families of State-of-the-Art panel econometric methods we have found a robust relationship among house prices, income, interest rates and stock of houses per capita for the Spanish real estate market in the period 1995:2005. A prominent result of this paper is to show the importance of the (dis)-equilibrium term in explaining the short-term dynamics of real house prices across Spanish provinces. Another important result is the potential evidence that a considerable share of regional GDP growth might have been motivated by wealth and expectation effects due to house prices being above their fundamental values.

Further research should improve those results by extending the sample time-period and performing international comparissons.

References

Arce, O., and J. D. López-Salido (2006). "House Prices, Rents, and Interest Rates Under Collateral Constraints", *Bank of Spain Working Papers*, 0610.

Ayuso, J., and F. Restoy (2006). "House Prices and Rents: An Equilibrium Asset Pricing Approach", *Journal of Empirical Finance*, forthcoming.

BIS (2002). *Quartely Review*, September.

Beyer, A. and R. Farmer (2006). "On the Indeterminacy of Determinacy and Indeterminacy", *American Economic Review*, forthcoming.

Breitung, J. (2005). "A Parametric Approach to the Estimation of Cointegration Vectors in Panel Data", *Econometric Reviews*, 24(2), pp. 151-173.

Breitung, J., and H. Pesaran (2007). "Unit Roots and Cointegration in Panels", forthcoming in *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Kluwer Academic Publishers.

Brunnermeier, M. K., and C. Julliard (2005). "Housing Prices and Inflation. What Fuels Housing Bubbles?", mimeo.

Campbell, J. Y., and R. J. Shiller (1988). "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", *Review of Financial Studies*, 1 (3), pp. 195-228.

Capozza, D. R., P. H. Hendershott, C. Mack and C. J. Mayer (2002). "Determinants of Real House Price Dynamics". *NBER Working Paper* 9262.

Capozza, D. R. and P. J. Seguin (1996). "Expectations, Efficiency, and Euphoria in the Housing Market", *Regional Science and Urban Economics*, 35, pp. 369-386.

Case, K., J. Quigley and R. Shiller (2001). "Comparing wealth effects: the stock market versus the housing market", *NBER Working Paper* 8606.

Girouard N., M. Kennedy, P van den Noord and Christophe André (2006). "Recent House Price Developments: The Role of Fundamentals", *OECD Economics Department Working Papers*, 475.

Groen, J.J.J. and F. Kleibergen (2003). “Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models”, *Journal of Business and Economic Statistics*, 21, 295-318.

Dobson, W. and G. Hufbauer (2001). World Capital Markets: A Challenge to the G-10. Institute for International Economics.

Engle, R.F., and C.W.J. Granger (1987), “Co-integration and Error Correction: Representation, Estimation, and Testing”, *Econometrica*, 55, 251-276.

Heathcote, J., and M. Davis (2005). “Housing and the Business Cycle”, *International Economic Review*, 46 (3), pp. 751-784.

Holly, S., H. Pesaran and T. Yamagata (2006). "A Spatio-Temporal Model of House Prices in the US," *IZA Discussion Papers 2338*, Institute for the Study of Labor (IZA).

IMF (2004), *World Economic Outlook*, pp. 71-89, September.

Jacobsen, D. H. and B. E. Naug (2005). “What Drives House Prices?”, Norges Bank Economic Bulletin, 1, pp. 29-41.

Johansen, S. (1991). “Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models”. *Econometrica* 59, pp. 1551–1580.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.

Kao, C., and M. Chiang (2000). “On the Estimation and Inference of a Cointegrated Regression in Panel Data”, *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics*, 15, JAI Press, pp. 179-222.

McCarthy, J., and R. W. Peach (2004). “Are Home Prices the Next Bubble?”, *Federal Reserve Bank of New York Economic Policy Review*, pp. 65-78, December.

Nagahata, T., Y. Saita, T. Sekine and T. Tachibana (2004), “Equilibrium land prices of Japanese prefectures: A panel cointegration analysis”, *Bank of Japan Working Paper Series*, 04-E-9.

Ngai, L. R., and , C. A. Pissarides (2004). “Structural Change in a Multisector Model of Growth”, *London School of Economics*, mimeo.

Maza, A. and J. Villaverde (2007). "A State-Space Approach to the Analysis of Economic Shocks in Spain", *Journal of Policy Modeling* 29, pp. 55–63.

Pedroni, P. (1995). "Panel Cointegration: Asymptotic and Finite Sample Propierties of Pooled Time Series Test with an Application to the PPP Hypothesis", *Indianma University Working Papers in Economics*, 95-013.

Pedroni, P. (2000). "Fully Modified OLS for Heterogeneous Cointegrated Panels", *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics*, 15, JAI Press, pp. 93-130.

Pesaran, M. H. (2006a). "Estimation and Inference in large Heterogeneous Panels with a Multifactor Error Structure", *Econometrica*, 74, pp. 967-1012.

Pesaran, M. H. (2006b). "A Simple Panel Unit Root Test in the Presence of Cross Section Dependence", mimeo, University of Cambridge.

Pesaran, M.H., Y. Shin and R.P. Smith (1999). "Pooled Mean Group Estimation of Dynamic Heterogeneous Panels", *Journal of the American Statistical Association*, 94, 621-624.

Phillips, P.C.B. (1991). "Optimal Inference in Co-integrated Systems", *Econometrica*, 59, 282-306.

Phillips, P.C.B. and B.E. Hansen (1990). "Statistical Inference in Instrumental Variable Regression with I(1) Processes", *Review of Economic Studies*, 57, 99-125.

Ruíz, P. and San Martin, I. (2004). "El Alquiler de la Vivienda en España, en una Encrucijada", *BBVA Real Estate Research*, 16-22, July.

Stock, J., and M. Watson (1993). "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783-820.

Tsatsaronis, K., and H. Zhu (2004) "What Drives Housing Price Dynamics: Cross-Country Evidence" *BIS Quarterly Review*, March, pp. 65-78.

Appendix

Figure 1: Evolution of Real Price Increments in Spain and some *Provincias* (YoY)

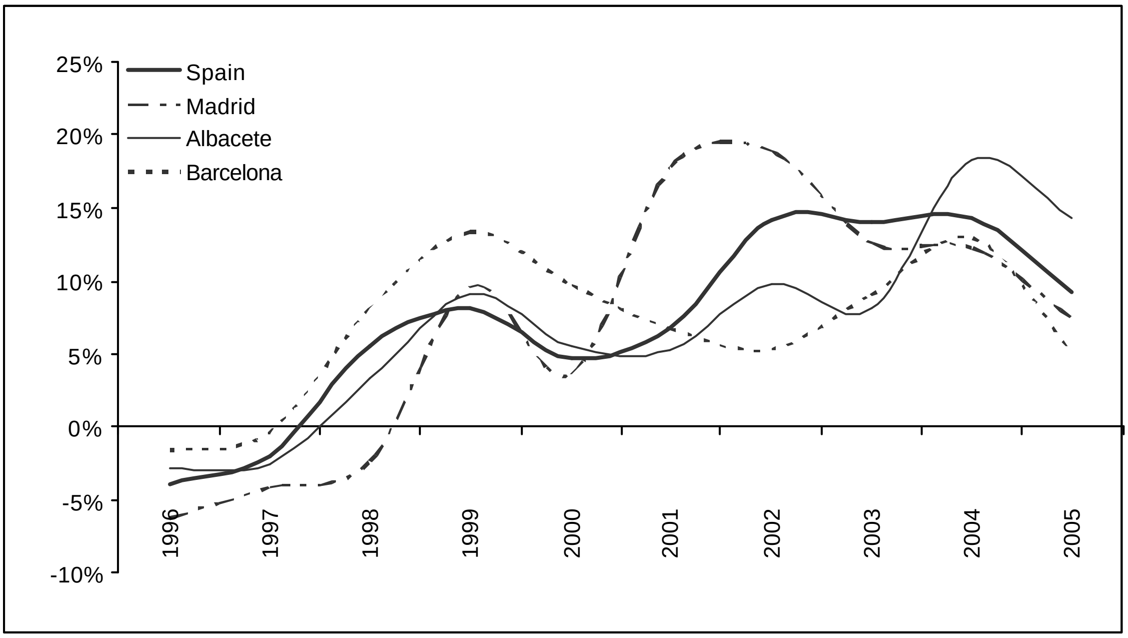

Figure 2. Evolution of the Increment of New Houses (YoY)

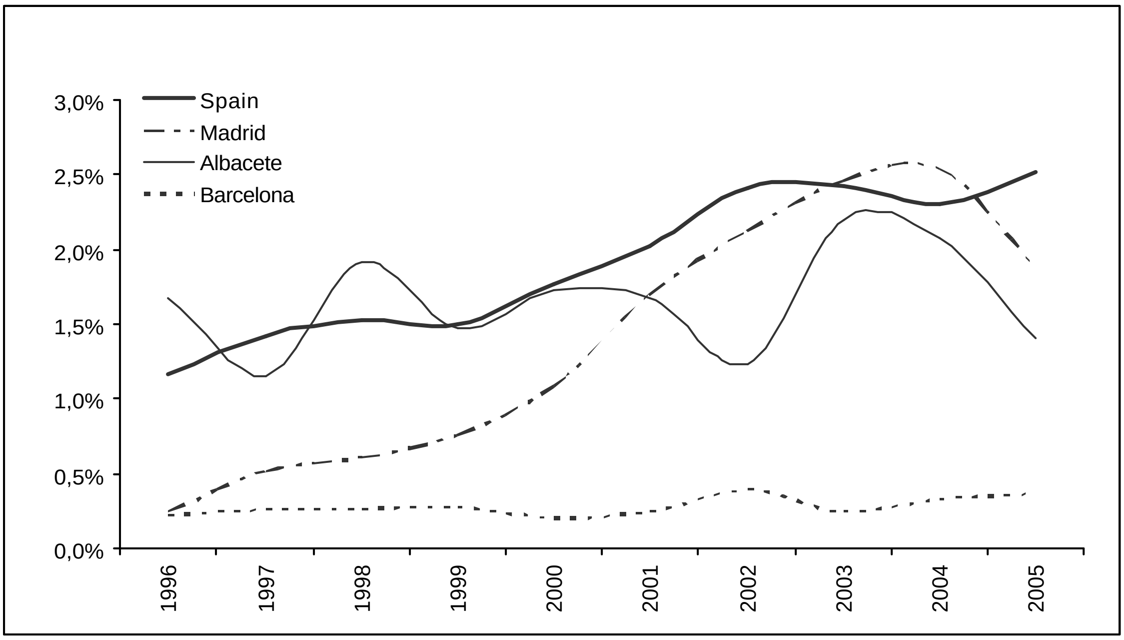

Figure 3 Real Interest Rates for the 50 *Provincias* (Overlapped)

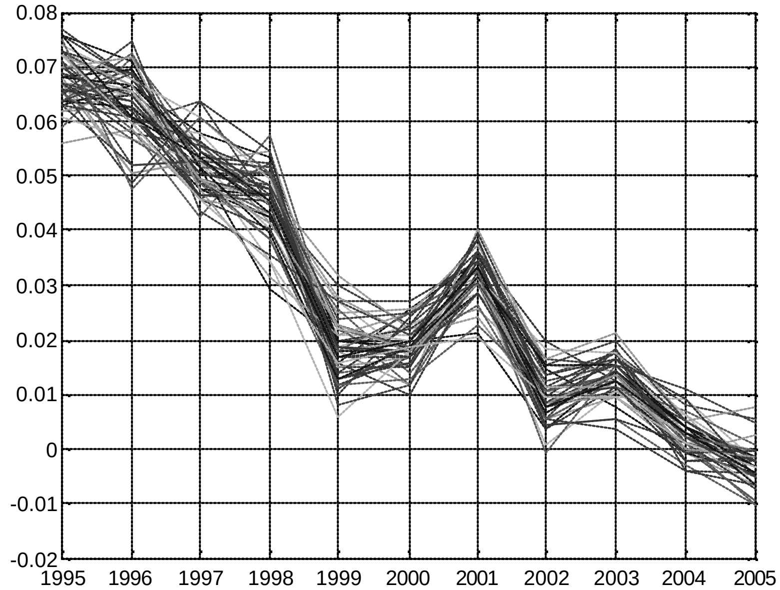

Figure 4 Panel-VECM Contributions to Growth Price of Housing (Madrid)

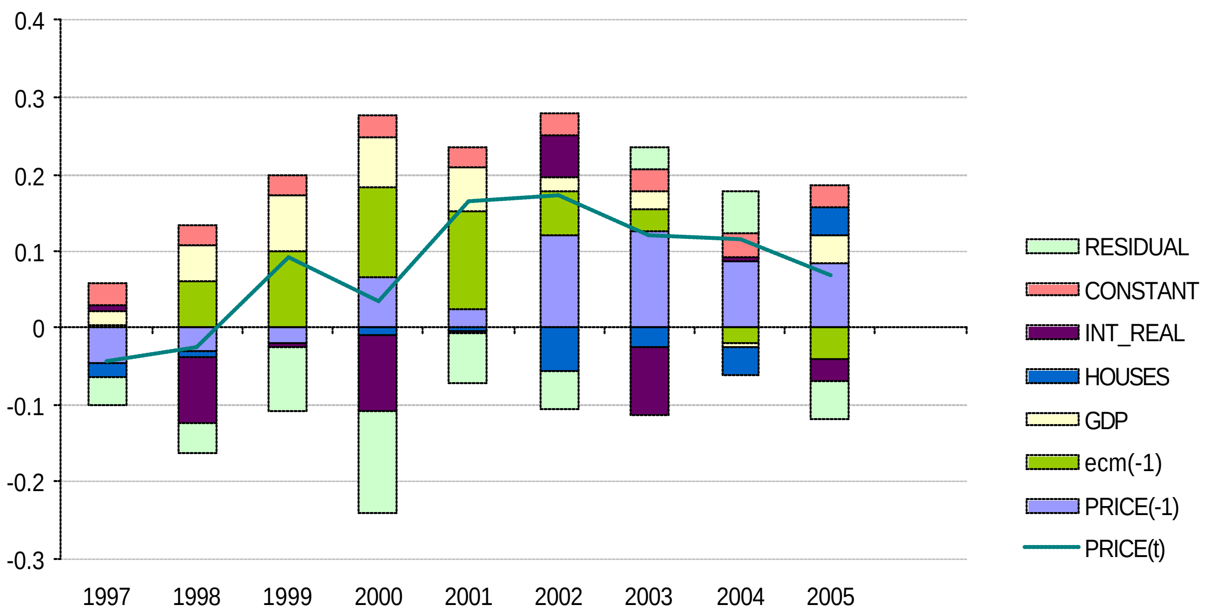

Figure 5 Panel-VECM Contributions to Growth Price of Housing (Barcelona)

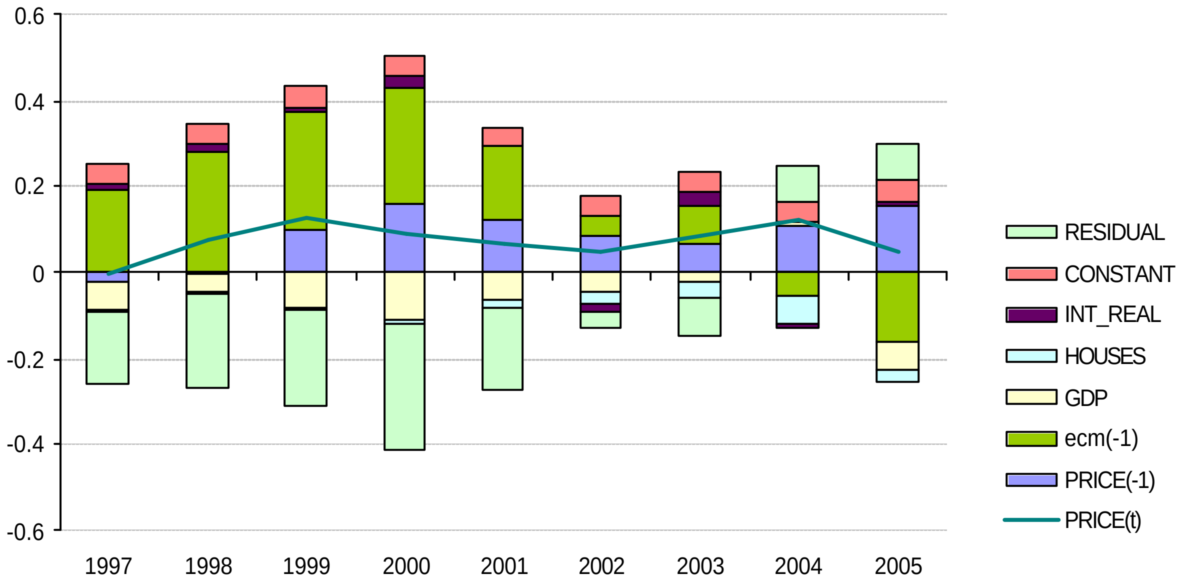

Figure 6 Panel-VECM Contributions to Per-Capita GDP Growth (Madrid)

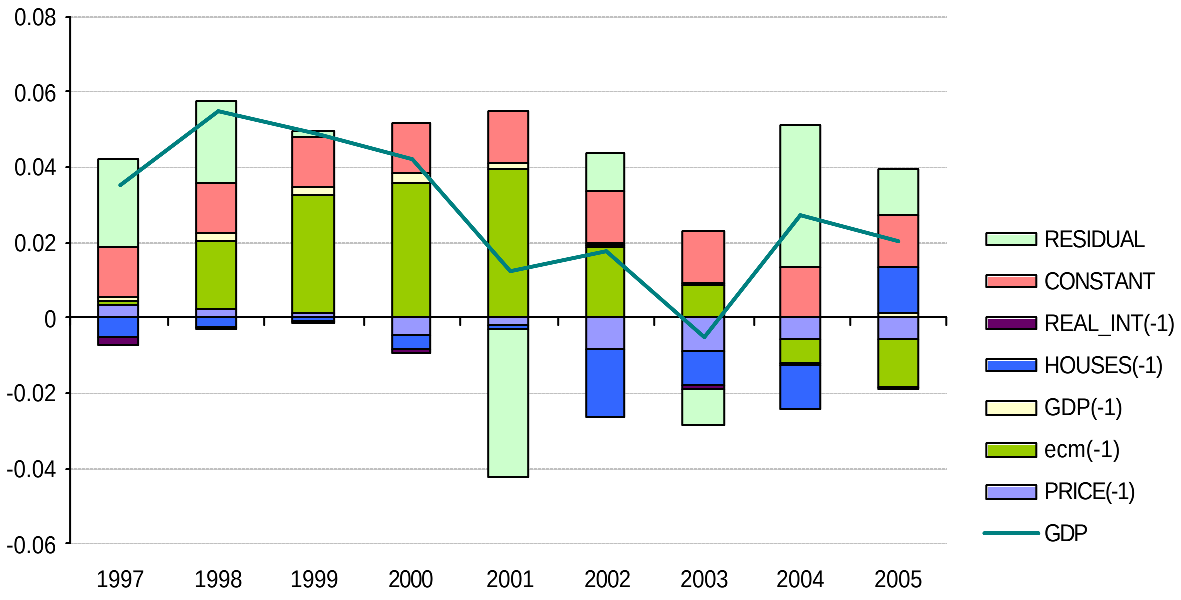

Figure 7 Panel-VECM Contributions to Per-Capita GDP Growth (Barcelona)

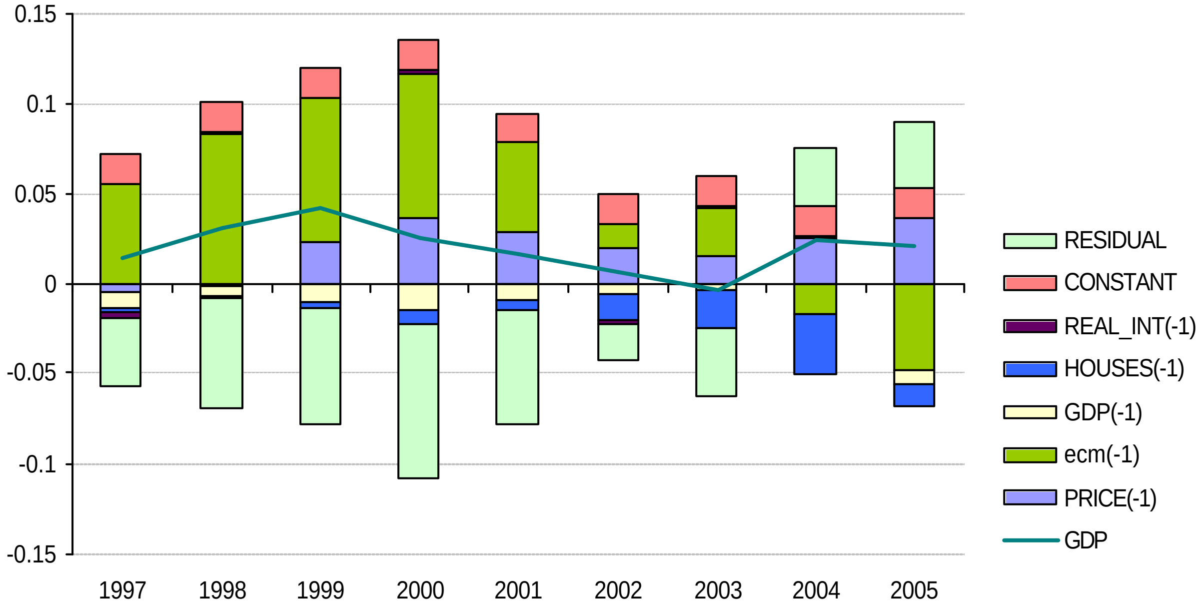

Table 1: List of Variables and their Descriptions

$P_{it,g}$	Regional Consumer Price Index (2005 = 1)
$P_{it,h}$	Regional housing price per square meter
GDP_{it}	Regional GDP
POP_{it}	Regional population
RB_t	Spanish average mortgage interest rates
$RENT_{it}$	Regional nominal rent index (1995 = 100)
HOU_{it}	Stock of houses
p_{it}	Natural logarithm of the real house price $p_{it} = \log(P_{it,h} / P_{it,g})$
y_{it}	Natural logarithm of the real per capita GDP $y_{it} = \log(GDP_{it,h} / (POP_{it}\ P_{it,g}))$
h_{it}	Natural logarithm of houses per capita $y_{it} = \log(HOU_{it,h} / POP_{it})$
r_{it}	Real interest rates $r_{it} = RB_t - \log(P_{it,g} / P_{i\text{-}1t,g})$
ib_{it}	YoY returns of the stock-market (IBEX)
ren_{it}	Natural logarithm of the real rental cost $ren_{it} = \log(RENT_{it,h} / P_{it,g})$

Notes: Annual data between 1995 and 2005 (T = 11) for 50 *provincias* (N = 50).

Table 2: Residual Cross Dependence of ADP(p) Regressions

CD Test Statistic		
	ADF(1)	ADF(2)
p_{it}	7.09	6.75
y_{it}	20.95	12.00
h_{it}	15.79	8.60
r_{it}	88.29	80.63
ren_{it}	41.53	34.67

Notes: p^{th}-order Augmented Dickey-Fuller test statistics, ADF(p) are computed for each cross-section unit separately. For all the variables except r_{it} an intercept and a linear trend are included. r_{it} includes only an intercept. $CD = \sqrt{2T / N(N-1)} \sum_{i=1}^{N-1} \sum_{i=1}^{N-1} r_{ij}$ tends to $N(0, 1)$ under the null hypothesis of no error cross section dependence. $?_{ij}$ is the pair-wise correlation of the residuals from the ADF.

Table 3: Pesaran's CIPS Panel Unit Root Test Results

	CIPS	CIPS*
Intercept and Trend		
p_{it}	-2.36	-2.11
y_{it}	-2.06	-1.73
h_{it}	-1.89	-1.89
ren_{it}	-1.31	-1.31
Only Intercept		
r_{it}	-0.67	-0.71
? p_{it}	-2.17*	-2.01
? y_{it}	-1.98	-1.83
? h_{it}	-1.63	-1.63
? ren_{it}	-1.49	-1.49
ib_{it}	-0.07	-0.10
No Intercept No Trend		
? r_{it}	-0.67	-0.73
? $^{2}y_{it}$	-2.89**	-2.63**
? $^{2}h_{it}$	-1.87**	-1.87**
? $^{2}ren_{it}$	-1.57*	-1.57**
? ib_{it}	-0.05	-0.09

Notes: The reported values are CIPS(1) statistics, which are cross section averages of Cross-sectionally Augmented Dickey-Fuller test statistics CADF(1). CIPS* refers to the truncated version of the test statistic, more suitable for small T panels. Relevant lower 5%/10% critical values for the CIPS statistic are -2.86/-2.71 for the intercept and trend case, -2.19/-2.07 for the only intercept case and -1.58/-1.46. Relevant lower 5%/10% critical values for the CIPS* statistic are -2.75/-2.73 for the intercept and trend case, -2.16/-2.05 for the only intercept case and -1.57/-1.46."**" "*" mean that the test is significant at the 5 and 10 per cent level respectively.

Table 4: Estimation Results: Cointegration Relation Estimated by DOLS

DOLS Results		
	Estimated Model	*Without h_{it}*
y_{it}	0.85 (0.27)	0.71 (0.26)
h_{it}	-0.31 (0.13)	--
r_{it}	-8.43 (1.30)	-8.51 (1.29)
R^2	0.96	0.95
CD	14.06	12.83
CIPS	-1.81***	-1.78***
*CIPS**	-1.88***	-1.84***

Notes: Estimated model is Equation (3.10), "Without h_{it} " refers to the same model without including houses per capita. The estimation method is Panel DOLS with $q = 1$. Standard errors are given in parenthesis and have been reescaled by the method presented in Stock and Watson (1993) so the *t*-statistic tends asymptotically to a $N(0,1)$. R^2 refers to the adjusted *R*-squared statistic. CD is the cross-dependence test statistic with ADF(1), which tends to $N(0, 1)$ under the null hypothesis of no error cross section dependence. CIPS and CIPS* refer respectively to the standard and truncated cross section averages of Cross-sectionally Augmented Dickey-Fuller test statistics CADF(1) with no intercept and no trend. Relevant lower 1% critical values for the CIPS and CIPS* statistic are -1.78 and -1.77 respectively. "***"means that the test is significant at the 1 and 10 per cent level .

Table 5: Estimation Results: Cointegration Relation Estimated by CCE

CCE Results		
	Pooled	*Mean Group*
y_{it}	0.17 (0.18)	0.35 (0.22)
h_{it}	0.16 (0.40)	-0.02 (0.56)
r_{it}	0.22 (0.46)	0.03 (0.56)

Notes: The estimation methods are Panel Common Correlated Effects, both Pooled and Mean Group. Standard errors are given in parenthesis .

Table 6: Estimation Results: Homogeneous Error Correction Model (OLS)

Error Correction Model Results	
	Pooled Model
u_{it-1}	-0.16 (0.02)
W? $\mathbf{p}_{t-1}$	--
? y_{it}	0.06 (0.12)
? h_{it}	-0.39 (0.13)
? r_{it}	-1.07
? ren_{it}	-0.17
ib_t	-0.03
R^2	0.21
Half Life	4.01
CD	33.45
CIPS	-2.15***
*CIPS**	-1.78***

Notes: Estimated model is Equation (3.13). The "Pooled Model" refers to the homogeneous model with independent intercepts (fixed effects). The estimation method is OLS. R^2 refers to the adjusted *R*-squared statistic. CD is the cross-dependence test statistic with ADF(1), which tends to $N(0, 1)$ under the null hypothesis of no error cross section dependence. CIPS and CIPS* refer respectively to the standard and truncated cross section averages of Cross-sectionally Augmented Dickey-Fuller test statistics CADF(1) with no intercept and no trend. Relevant lower 1% critical values for the CIPS and CIPS* statistic are -1.78 and -1.77 respectively. "***"means that the test is significant at the 1 and 10 per cent level .

Table 7: Estimation Results: Heterogeneous Error Correction Model (OLS)

Error Correction Model Results				
	Barcelona 1	*Madrid 1*	*Barcelona 2*	*Madrid 2*
u_{it-1}	-0.68 (0.35)	-0.25 (0.23)	-0.47 (0.14)	-0.07 (0.16)
W? $\mathbf{p}_{t-1}$	1.41 (1.30)	-0.23 (0.72)	0.33 (0.19)	1.03 (0.36)
? y_{it}	-4.93 (5.71)	-4.37 (2.10)	--	--
? h_{it}	8.97 (10.1)	-2.10 (3.72)	--	--
? r_{it}	-0.52	-1.14	-2.39	-1.05
? ren_{it}	6.22	3.84		
ib_t	0.13	0.14	0.05	-0.04
R^2	0.91		0.62	
CD	5.61		33.87	
CIPS	-2.77***		-2.01***	

Notes: Estimated model is Equation (3.13); coefficients are for 2 regions (Madrid and Barcelona) whereas test statistics are for the whole sample. Two different versions of (3.13) are estimated: #1 includes the full set of variables whereas #2 is a reduced subset. The estimation method is OLS. R^2 refers to the adjusted *R*-squared statistic. CD is the cross-dependence test statistic with ADF(1), which tends to *N*(0, 1) under the null hypothesis of no error cross section dependence. CIPS and CIPS* refer respectively to the standard and truncated cross section averages of Cross-sectionally Augmented Dickey-Fuller test statistics CADF(1) with no intercept and no trend. Relevant lower 1% critical values for the CIPS and CIPS* statistic are -1.78 and -1.77 respectively. "***"means that the test is significant at the 1 and 10 per cent level

Table 8: Speed of Adjustment Price of Housing to Disequilibirum
Panel-VECM

Alava	6.00	*León*	0.04
Albacete	0.01	*Lleida*	1.28
Alicante	0.75	*Lugo*	0.35
Almeria	0.23	*Madrid*	0.51
Asturias	0.22	*Malaga*	3.17
Avila	1.16	*Murcia*	0.88
Badajoz	1.53	*Navarra*	2.35
Baleares	2.22	*Orense*	1.81
Barcelona	2.31	*Palencia*	0.74
Burgos	2.56	*Las Palmas*	0.27
Caceres	0.78	*Pontevedra*	0.39
Cádiz	0.60	*La Rioja*	0.07
Cantabria	0.87	*Salamanca*	2.43
Castellon	0.24	*Tenerife*	0.55
Ciudad Real	0.40	*Segovia*	0.22
Cordoba	0.11	*Sevilla*	0.90
A Coruña	0.18	*Soria*	0.15
Cuenca	0.63	*Tarragona*	0.09
Gerona	0.02	*Teruel*	1.30
Granada	2.14	*Toledo*	0.59
Guadalajara	0.56	*Valencia*	0.81
Guipuzcoa	3.29	*Valladolid*	0.08
Huelva	0.08	*Vizcaya*	1.36
Huesca	0.11	*Zamora*	1.29
Jaén	0.17	*Zaragoza*	1.19

Notes: The numbers indicate the ratio of the loading factor $\hat{\alpha}_i$ with respect to the mean of the provinces.

Table 9: Estimation Results: Cointegration Relation Panel-VECM

y_{it}	1.24 (0.18)
h_{it}	-0.12 (0.06)
r_{it}	-2.50 (0.94)
CD	-2.05
CIPS	-2.5***
*CIPS**	-2.2***

Notes: Estimated coefficients in the second step of the Breitung (2005) Panel-VECM procedure. Standard errors are given in CIPS and CIPS* refer respectively to the standard and truncated cross section averages of Cross-sectionally Augmented Dickey-Fuller test statistics CADF(1) with no intercept and no trend. Relevant lower 1% critical values for the CIPS and CIPS* statistic are -1.78 and -1.77 respectively. "***"means that the test is significant at the 1 and 10 per cent level.