Econometric Advances in the Estimation of Housing Price Determinants

António Cunha
ISAG – European Business School, and Research Center in Business Sciences and Tourism (CICET – FCVC); CEOS.PP, Porto Accounting and Business School, Polytechnic Institute of Porto, Portugal

antonio.cunha@isag.pt

Abstract: This paper aims to help improve the estimations produced by researchers who rely on conventional housing market pricing models to determine housing prices. The widespread use of panel data in estimating housing prices is justified by the richness of cross-sectional regional or metropolitan data analysed over several periods. Unfortunately, panel data has slope coefficient heterogeneity and cross-sectional dependence, producing inconsistent and misleading estimates of the coefficients using the Ordinary Least Squares (OLS) estimator. Recent advances in econometrics address these panel data limitations, producing better estimates. We analysed the empirical application of these new estimators on housing market panel data, showing that the Fully Modified Ordinary Least Squares Augmented Mean Group (FMOLS-MG) estimator produces the best estimates of the long-term housing market equilibrium and that the Dynamic Common Correlated Effects Mean Group (DCCE-MG) estimator produces the best estimates of the housing market's short-term dynamics. Adopting a trending methodology like Difference-in-Differences (DID) in housing market research to explain the effects of policy decisions on housing prices also has complications related to using the OLS estimator with fixed effects when the data has serial correlation. We show these problems can be overcome using the Feasible Generalised Least Squares estimator in a Seemingly Unrelated Regression Equations (FGLS-SURE) system. Recent econometric developments produce more accurate housing price determinant estimates than conventional econometric methods. These new methodologies can help researchers better estimate the effects of fundamental economic changes and policy decisions on housing prices, which can, in turn, support policymakers in implementing better housing policies.

Keywords: Econometric advances, Housing price determinants, Augmented mean group, Difference-in-differences

1. Introduction

Over the last decade, the housing price increases in several Western countries have forced policymakers and academics to analyse and address the root causes to make housing more affordable. The fact that housing prices continued to rise even during the Covid-19 pandemic, despite significant drops in gross domestic product, and that local and national measures banning the shift of apartments from housing to tourism (e.g., Airbnb) did not stop price increases indicates that policymakers and academics need to understand better how real estate market prices are determined.

Real estate pricing models are based on general equilibrium under adaptive expectations, where the market does not reach equilibrium in the short term, and supply and demand players develop expectations according to past price changes. Long-term housing price estimates are formed on supply and demand equilibrium. In contrast, short-term house price returns are influenced by lagged changes in fundamentals and an error correction model that accounts for the slow adjustment of short-term price changes to long-term equilibrium.

These asset pricing models estimate housing price determinants using panel data due to the richness of cross-sectional regional or metropolitan data analysed over several periods. The problem is that most of these data sets exhibit slope coefficient heterogeneity and cross-section dependence. Heterogeneity causes the pooled Ordinary Least Squares (OLS) estimator to create a correlation between the regressors and the error term, resulting in serially correlated disturbances and producing inconsistent and misleading estimates of the coefficients. Moreover, under cross-section dependence, the pooled OLS estimator provides little gain in precision compared with time series OLS. Recent advances in econometrics, such as the Mean Group (MG) estimator, the Fully Modified Ordinary Least Squares (FMOLS), and the Common Correlated Effects (CCE), address these panel data limitations, improving the precision of estimates.

Moreover, asset pricing models often fail to estimate the effects of policy decisions that impact housing use, such as the liberalisation of short-term rentals. New methodologies developed to measure policy decisions, such as the Difference-in-Differences (DID) approach, can also be applied to the housing market to assess the effects of short-term rental liberalisation on housing prices. However, the OLS with Fixed Effects (OLS-FE) estimator, commonly used in the DID methodology, is prone to serial correlation and biased estimates. Recent research suggests that using the Feasible Generalised Least Squares estimator in a Seemingly Unrelated Regression Equation system (FGLS-SURE) could address these issues.

We analysed an empirical study that tested several of these new estimators on a sample of housing market data. The study showed that the Fully Modified Ordinary Least Squares Mean Group (FMOLS-MG) estimator is the best solution for estimating long-term equilibriums in the housing market, as it allows for slope heterogeneity and controls for cross-sectional dependence. For short-term dynamics, the estimator that provides the best estimates is the Dynamic Common Correlated Effects Mean Group (DCCE-MG) estimator. Regarding the Difference-in-Differences (DID) methodology, we analysed a study that shows the Feasible Generalised Least Squares Seemingly Unrelated Regression (FGLS-SUR) estimator has much higher power than the OLS with Fixed Effects (OLS-FE) estimator, leading to more precise estimates and more valid results.

The paper is organised as follows: section 2 reviews the literature, section 3 explains the methodology, section 4 presents and discusses the results, and section 5 concludes.

2. Literature Review

2.1 Housing Prices

The first academic approach to real estate prices (or housing prices) was developed by David Ricardo (1817) with the Ricardian Rent Theory, which posits that population and the location of economic activities are the main determinants of real estate prices. In the late 20th century, the advent of computerised databases made it possible to combine time-series data with cross-sectional data to produce longitudinal information (Hsiao, 1985) and thus develop econometric estimators for panel data. Unlike time series data, panel data contain information regarding inter-individual differences, allowing for estimating parameters that depend on various socio-demographic or geographic factors, which are important determinants of real estate prices. While cross-sectional data include variations in these variables, they cannot be used to model dynamics such as housing price growth. The estimated coefficient of a cross-sectional database would reflect the differences between housing locations but not the evolution of housing prices in a specific location.

The term "panel" refers to a two-dimensional board with variable X having two subscripts. One dimension is an individual index (i), representing the cross-sections, and the other is time (t), representing the time series. In macroeconomics, the time dimension is usually larger than the number of individuals. In contrast, in microeconomics, such as in real estate pricing models, the number of individuals is typically larger than the number of periods. Due to these characteristics of panel data, along with the Ricardian Rent Theory's concept of location as a fundamental determinant of real estate prices, most real estate price models have been developed under the assumption that panel data can reflect the different price conditions of various locations (Case and Shiller, 1990; Poterba, 1991). By the end of the 1990s, it became widely accepted that real estate prices should be studied using panel data from metropolitan statistical areas (DiPasquale & Wheaton, 1996; Quigley, 1999). In Europe, these areas are known as the Nomenclature of Territorial Units for Statistics (NUTS 3).

Although the use of panel data has become widely accepted, the debate about which fundamentals explain real estate prices remains open. Ricardo (1817) argued that population growth and location would determine real estate price growth. However, several studies based on this assumption have been proven wrong (Mankiw & Weil, 1989; Bakshi & Chen, 1994). DiPasquale and Wheaton addressed this shortfall by developing a stock-flow equilibrium model (DiPasquale & Wheaton, 1992) to explain housing price equilibrium. This model was later developed into a structural model (DiPasquale & Wheaton, 1994) by incorporating the dynamic process of prices increasing or decreasing until a new equilibrium is reached, based on Case and Shiller's findings about market inefficiency (1988). This model considers the main determinants of housing prices to be aggregate income (measured by gross domestic product, GDP), construction costs, and interest rates. Since the housing market is inefficient and does not reach equilibrium in the short term, two additional variables are included in the model to explain housing prices: last-period price change (momentum, also known as lagged price) and mean reversion (to the equilibrium price). This model has been used extensively since then (Malpezzi, 1999; Green et al., 2005).

2.2 Equilibrium and Dynamic Econometric Models

Using panel data brought new challenges to the econometric field, as Pesaran and Smith (1995) showed. When there is a dynamic model, as is often the case in the housing market, the conventional pooled regression imposes common slopes on all regressors, although allowing for fixed or random intercepts. This slope homogeneity across all housing locations seems highly unlikely. When the regressors are serially correlated (as with lagged prices), ignoring coefficient heterogeneity induces serial correlation in the disturbance, leading to biased estimates. Pesaran and Smith (1995) addressed this problem by estimating separate time series regressions for each cross-section and averaging the coefficients over cross-sections (also known as groups). This is the Mean

Group estimator, which produced consistent, unbiased estimates in the presence of large groups and periods, thus solving one of the problems in estimating housing price determinants.

A second problem associated with dynamic panels is endogeneity. The ordinary least squares (OLS) estimator assumes that regressors are exogenous, but in a dynamic process, there is endogenous feedback from the regressors, creating bias. Pedroni (2001) developed a Fully Modified OLS Mean Group (FMOLS-MG) estimator that introduces a semi-parametric correction to the OLS estimator, eliminating the bias induced by the endogeneity of regressors, in addition to the MG estimator feature that allows for slope heterogeneity.

Another problem associated with panel data is cross-section dependence. In housing markets, spatial correlation is prevalent and intuitive, as Ricardo (1817) explains: the economic distance to activities determines real estate prices. In panel data models where the number of cross sections is small and the time series dimension is large, the standard approach is to treat the equations from the different cross sections as a system of seemingly unrelated regression equations (SURE) and then estimate the system using the Generalised Least Squares (GLS) technique (Pesaran, 2006). This approach allows for general (time-invariant) correlation patterns across the errors in the different cross-section equations.

To address the cross-sectional dependence problem, Eberhardt and Teal (2010) extended the FMOLS-MG by adding a dynamic process to the regression that controls cross-section dependence. They extracted the process from the year dummy coefficients of a pooled regression in first differences, representing the level-equivalent mean evolution of unobserved common factors across all panel sections. Afterward, these estimates are augmented in each of the cross-sectional regressions, including linear trend terms to capture omitted idiosyncratic processes, with estimates averaged across sections, as done by Pesaran and Smith (1995). This estimator is called the Fully Modified OLS Augmented Mean Group (FMOLS-AMG) and holds in the case of slope heterogeneity and cross-section dependence.

Regarding cross-sectional dependence, Pesaran (2006) had already developed an estimator where the observed regressors are augmented by cross-section weighted averages of the dependent variable and the individual-specific regressors, calling this the Common Correlated Effects Mean Group (CCE-MG) estimator. This estimator is unbiased in regressions with slope heterogeneity and cross-sectional dependence. However, it is unsuitable for dynamic panels where the lagged dependent variable is one of the regressors.

Chudik and Pesaran (2015) extended the Pesaran (2006) CCE-MG to dynamic panel data models with weakly exogenous regressors to address the CCE-MG shortfall in terms of dynamic panels. This new estimator is considered for panel data autoregressive distributed lag (ARDL) models where the dependent variable of a cross-section is explained by its lagged values, current and lagged values of weakly exogenous regressors, unobserved serially correlated common factors, and serially uncorrelated idiosyncratic error. This new estimator performs well as long as the time series dimension of the panel is sufficiently large.

2.3 Difference-in-Differences Model

The equilibrium and dynamic econometric models should be able to explain housing prices in the long term (equilibrium) and the short term (dynamic) based on observed economic fundamentals. However, unobserved policy decisions or external factors might affect the housing market beyond the economic fundamentals. The seminal work of Ashenfelter and Card (1985), where an unobserved effect was measured as a treatment against a control group, created the conditions for developing the Difference-in-Differences (DID) methodology. Although the model is based on the OLS estimator with fixed effects, it is a quasi-experimental methodology, as it measures whether there is a difference in an outcome due to a hypothesised treatment over a difference in time, enabling the isolation of the treatment effect above and beyond any difference that would have been expected regardless of the treatment. Because it relies on direct before-and-after measures at the same location rather than relying on modeling differences between locations, the DID approach provides a direct measure of the housing price changes attributable to policy decisions, such as the liberalisation of the transfer of houses from residential purposes to tourism purposes.

Recent studies on the effects of tourism on housing prices have employed the DID methodology (Sommerville et al., 2020; Barron et al., 2021; Franco & Santos, 2021), yielding promising results. However, econometricians have criticised the methodology, arguing that the fixed effects OLS estimator underestimates standard deviations due to serial correlation (Bertrand et al., 2004), as seen in the previous section. Parks (1967) had already developed a Feasible Generalised Least Squares (FGLS) estimator that allowed for serial correlation in seemingly unrelated regression equation systems (Zellner, 1962). Hansen (2007) proposed using FGLS in the Difference-in-Differences methodology, concluding that it would result in a more efficient estimator. Hausman and Kuersteiner (2008) showed that when there is serial correlation in the data, FGLS produces better estimates than OLS with fixed effects, especially in the case of long-period datasets, a result later confirmed by Brewer et al. (2017), who recommended the use of FGLS with cluster-robust residuals in DID.

These theoretical models have yet to be widely validated, as they are recent econometric findings. We are interested in reviewing the empirical testing of the panel dynamic models and the DID to discuss the results against the reviewed theoretical models.

3. Methodology

This section presents part of the methodology of two recent research studies about housing prices where recent econometric advances mentioned in the previous section were applied to empirical data. The first study, focusing on dynamic panel models, was conducted in Spain's and Portugal's metropolitan areas (Cunha & Lobão, 2022a). These two countries have synchronised economic fundamentals (GDP, interest rates, construction costs), so the long-term housing market equilibrium should be similar, but they might show different short-term dynamics. In subsection 3.1 below, we demonstrate how recent econometric advances mentioned in the literature review can be used to estimate the dynamic model. The second study focuses on Portugal's two largest metropolitan areas. Over the last decade, Portugal reported housing price growth above the European Union average, making it one of the world's most unaffordable housing markets (price to income). Portugal's government banned new tourism short-term rental registrations, alleging that this activity transferred residential apartments to tourism activities. This second study used a Difference-in-Differences methodology to measure the effect of these short-term rentals (Airbnb tourism) on housing prices (Cunha & Lobão, 2022b). Section 3.2 below demonstrates how recent econometric advances mentioned in the literature can be used to improve the estimations.

3.1 Panel Dynamic Model

The variables used for estimating the house price equilibrium by Cunha & Lobão (2022a) were based on DiPasquale and Wheaton's (1992, 1994) equations: House Price (P), Gross Domestic Product (GDP), Construction costs (C), and Interest rates (R). Data series from Spain's 52 metropolitan statistical areas (MSA) were collected from Statistics and Spain's Ministry of Transportation between 2011Q1 and 2021Q2. Portugal's 25 MSA data series were collected from Portuguese Statistics for the same period. House Price is the median price in euros per square meter of housing. GDP is seasonally smoothed, in millions of euros, based on annual values with quarterly values interpolated based on quarterly national GDP values. Construction cost is an index with a base of 100 in 2015, with the same quarterly value for all MSA within the same country. Interest rates are the implicit national mortgage interest rate, with the same value for all MSA within the same country. All variables are in natural logarithms. The data for Iberia is the merge of the two datasets, with 77 MSA (i) and 42 quarters (t), resulting in 3,234 observations per variable and a total sample of 12,936 observations.

Long-term equilibrium

Following DiPasquale and Wheaton (1992), The long-term equilibrium equation is:

$$P^*_{i,t} = \beta_{0i} + \beta_{1i}GDP_{i,t} + \beta_{2i}C_{i,t} - \beta_{3i}R_{i,t} + \varepsilon_{i,t} \quad (1)$$

Where P^* is the log of the house price level, representing the long-term equilibrium price of Metropolitan Statistical Area i at quarter t, GDP represents aggregate income (using Gross Domestic Product as proxy), C is the construction cost, which is the same for all MSA within the same country, R is the mortgage interest rate, which is also the same for all MSA within the same country and ε is white noise and has zero mean. All variables are expressed in the logarithmic form.

Cunha and Lobão (2022a) employed five different estimation methods, each with distinct stationarity, slope coefficient heterogeneity, and cross-section dependence features. First, they used Pesaran and Smith's (1995) Mean Group (MG) estimator, which allows for slope coefficient heterogeneity and non-stationary but cointegrated data, but does not control for cross-sectional dependence. Second, they utilised Pedroni's (2001) Fully Modified Ordinary Least Squares Mean Group (FMOLS-MG) estimator, which controls for endogeneity and includes MG features. Third, they applied Eberhardt and Teal's (2010) Fully Modified Ordinary Least Squares Augmented Mean Group (FMOLS-AMG) estimator, which controls for cross-sectional dependence and includes FMOLS-MG features. Fourth, they employed Pesaran's (2006) Common Correlated Effects Mean Group (CCEMG) estimator, which also controls for cross-sectional dependence in addition to the characteristics of MG. Fifth, they utilised Chudik and Pesaran's (2015) Dynamic Common Correlated Effects Mean Group (DCCEMG), which shares the features of CCEMG but includes lagged cross-sectional means of variables to remove bias caused by weakly exogenous regressors.

Short-term dynamics

After selecting the best estimator for the house price's long-term equilibrium level, they proceeded to estimate the equation for short-term dynamics, where the long-term price is used in the error-correction term to compute the deviations from observed house price levels, following the specification below, as per DiPasquale and Wheaton (1994):

$$\Delta P_{i,t} = \beta_{0i} + \beta_{1i}\Delta P_{i,t-1} + \beta_{2i}\Delta C_{i,t-1} - \beta_{3i}\Delta R_{i,t-1} + \beta_{4i}\Delta GDP_{i,t-1} - \beta_{5i}(P-P^*)_{i,t-1} + \varepsilon_{i,t} \qquad (2)$$

Where the current period and MSA price change ($\Delta P_{i,t}$) is determined by the lagged price change ($\Delta P_{i,t-1}$), by the lagged changes (*t-1*) in market fundamentals (the same as in the long-term equation), and by the previous period deviation of the house price level from its long-term equilibrium level ($(P-P^*)_{i,t-1}$ (the error-correction term), all variables are expressed in the logarithmic form.

They ran the equation with the conventional OLS Fixed Effects and Random Effects estimators and with MG, CCEMG, and DCCEMG.

Difference-in-differences

The short-term rental activity was fully liberalised in Portugal in August 2014. It became possible to transfer houses and apartments from housing to tourism without political or administrative opposition. The most affected areas were the two largest metropolitan statistical areas (MSA), Lisbon and Porto, composed of several municipalities, each with its short-term rental registry and housing prices.

This dated event allowed Cunha and Lobão (2022b) to apply the Ashenfelter and Card (1985) Difference-in-Differences model, dividing the municipalities into two groups and measuring the effects of this political decision on the housing prices of the most affected group (the treated group). They created a treated group of municipalities for each MSA, comprising the municipalities in the upper quartile of short-term intensity as a percentage of the total housing stock, and a control group with the municipalities in the lower quartile of short-term intensity as a percentage of the total housing stock. Each MSA's treated and control groups are not geographically close, thus validating the stable unit treatment value assumption.

The dependent variable (Price) is the monthly median price per square meter of bank housing valuation in the municipalities within both MSA. This data was obtained from the Statistics Portugal website for all municipalities of the Lisbon MSA and the Porto MSA from January 2011 to December 2019. In the Lisbon MSA, there is data on 17 out of 18 municipalities, and in the Porto MSA, there is data on 12 out of 17 municipalities. There are 108 price observations for each of the 29 municipalities, which gives a total of 3,132 observations. They included two control variables (Density and Salaries) known in the literature to determine housing price increases. The data for these control variables were also obtained from the Statistics Portugal website. The DID specification is as follows:

$$Log(P_{ijt}) = \beta_{0j} + \beta_1 .\ T_{ij}.L_t + \beta_2.Log(D_{ijt}) + \beta_3.Log(S_{ijt}) + \varepsilon_{i,t}, \qquad (3)$$

$Log(P_{ijt})$ is the log price of housing in municipality *i* from MSA *j* in month *t*, $T_{ij}.L_t$ is a dummy variable set to one if the group of municipalities *i* is in the upper quartile of the MSA *j* (Treated) and the month *t* is post-liberalisation (Liberalisation), or zero otherwise, $Log(D_{ijt})$ is the number of residents per square kilometer of the municipalities *i* from MSA *j* in month *t* (annual data repeated monthly); $Log(S_{ijt})$ is the mean salary of municipality *i* form MSA *j* in month *t* (annual data repeated monthly); ε is white noise and has zero mean.

First, they estimate Equation 3 with the conventional OLS-FE estimator. Following the recommendation of Hansen (2007) for the DID methodology, they also use a Feasible Generalised Least Squares (FGLS) estimator in a Seemingly Unrelated Regression Equation (SURE) system (Zellner, 1962), thus correcting for heteroscedasticity and serial and contemporaneous correlation in the residuals, clustering standard errors by municipality.

Results

In the following subsections, we will present the estimates of equations 1, 2, and 3 obtained from sections of the two research papers (Cunha & Lobão, 2022a, 2022b). We will compare the results using conventional econometric methodologies to those obtained with recent advances. We will not present descriptive statistics because the purpose of this study is to compare econometric estimates, not to study housing price determinants.

3.2 Long-Term Equilibrium Results

The estimates of Equation 1 can be found in Table 1 below. MG, CCE-MG, and DCCE-MG results were estimated with Stata command xtdcce2, FMOLS-MG with Eviews, and FMOLS-AMG with Stata command xtmg. The data series are stationary, as shown by the unit root tests in Table 1.

Table 1: Estimates of Equation 1

Estimation method:		MG	FMOLS-MG	FMOLS-AMG	CCE-MG	DCCE-MG
GDPit		0.272***	0.250**	0.227***	0.350***	0.348***
		(0.028)	(0.01)	(0.017)	(0.086)	(0.091)
Cit		1.145***	1.160**	0.033	0.072	0.119
		(0.148)	(0.020)	(0.094)	(0.074)	(0.0767)
Rit		0.246***	0.240***	0.101***	0.108***	0.113***
		(0.020)	(0.003)	(0.021)	(0.021)	(0.020)
CIPS unit root test statistics						
1 lag		-3.986***	-1.569	-20.836***	-23.890***	-25.541***
2 lags		-3.991***	-1.520	-18.160***	-20.189***	-22.510***
3 lags		-4.001***	-1.488	-16.481***	-17.050***	-21.282***
4 lags		-3.920***	-1.448	-15.450***	-15.438***	-17.672***
R^2 (MG)		0.640	0.570		0.920	0.910
F-test / Wald test / statistic		159568***	n.a.	233***	37***	7***
Root Mean Square Error		0.060	n.a.	0.027	0.030	0.030
Slope homogeneity tests (a)		0.000***	n.a.	n.a.	0.000***	0.000***
Mean cross-correlations		0.560***	0.320***	0.080***	0.060***	0.060***

Notes: The dependent variable is *Pit*. *Pit*, *GDPit, Cit,* and *Rit* represent, respectively, House Price, Gross Domestic Product, Construction Cost, and Mortgage Interest rates of MSA *i* at quarter *t*. In models CCEMG and DCCEMG, Construction costs and Interest rates are not added as cross-sectional averages because they are constant in all cross-sections within the same country. The number of cross-section lags in DCCEMG is set to 3 (the cubic root of the number of periods). (a) The slope homogeneity test is the Pesaran and Yamagata (2008). ∗∗ and ∗∗∗ denote statistical significance at the 5% and 1% levels, respectively. Standard errors in parentheses.

The results of the CIPS tests indicate that except for FMOLS-MG, all four other estimates are cointegrated in levels, meaning that the estimated relations are long-term equilibrium equations. We can observe that the methods that do not control for cross-section dependence (MG and FMOLS-MG) produce similar estimates for construction cost and interest rate coefficients. In contrast, methods controlling for cross-section dependence produce different estimates of these coefficients but are similar among themselves. This was expected, as construction costs and interest rates are included in the model as control variables.

As expected, the mean cross-section correlation is high in the MG and FMOLS-MG estimates and small in the other three models that control for cross-section dependence, and the hypothesis of slope homogeneity is rejected. The FMOLS-AMG estimator combines a small cross-sectional dependence with minor standard errors and produces the smallest root mean square error. This method best estimates the error-correction term incorporated in Equation 2.

Short-term dynamics

The estimates of Equation 2 can be found in Table 2 below. OLS-FE and OLS-RE results were estimated with Eviews.

Table 2: Estimates of Equation 2

Estimation method:	OLS-FE	OLS-RE	MG	CCE-MG	DCCE-MG
ΔPi,t −1	-0.244***	-0.238***	-0.131***	-0.240***	-0.238***
	(0.018)	(0.018)	(0.029)	(0.022)	(0.024)

Estimation method:	OLS-FE	OLS-RE	MG	CCE-MG	DCCE-MG
ΔGDPi,t−1	0.043***	0.056***	0.021	-0.008	0.119
	(0.136)	(0.014)	(0.014)	(0.013)	(0.093)
ΔCi,t−1	0.142**	0.186***	0.223**	-0.081	0.123
	(0.056)	(0.056)	(0.101)	(0.089)	(0.103)
ΔRi,t−1	-0.017	-0.043***	-0.007	-0.014	0.003
	(0.013)	(0.013)	(0.012)	(0.013)	(0.018)
(P−P*)i,t−1	-0.039***	-0.001	-0.077***	-0.099***	-0.198***
	(0.008)	(0.002)	(0.011)	(0.016)	(0.024)
R2	0.030	0.070	0.110	0.180	0.330
F-test / Wald / statistic	56***	219***	1***	1***	1***
Root Mean Square Error			0.030	0.030	0.030
Slope homogeneity tests (a)	0.000***	0.000***	0.000***	0.849	0.849
Mean cross-correlations	0.210***	0.260***	0.120***	0.080***	0.020***

Notes: *Pit*, *GDPit, Cit,* and *Rit* represent, respectively, House Price, Gross Domestic Product, Construction Cost, and Mortgage Interest rates of MSA *i* at quarter *t*. *P** was estimated by the FMOLS-AMG of Table 1. In models CCEMG and DCCEMG, Construction costs and Interest rates are not added as cross-sectional averages because they are constant in all cross-sections within the same country. The number of cross-section lags in DCCEMG is set to 3 (the cubic root of the number of periods). (a) The slope homogeneity test is the Pesaran and Yamagata (2008). $**$ and $***$ denote statistical significance at the 5% and 1% levels, respectively. Standard errors in parentheses.

In Table 2, we can observe that the mean cross-correlations are smaller compared to Table 1, even when comparing the OLS-FE method (in Table 2) to MG in Table 1, suggesting that the use of FMOLS-AMG as the estimator for long-term equilibrium price is controlling some of the cross-section dependence in the housing returns estimates. While in conventional estimation methods (OLS-FE and OLS-RE), the theoretical house price determinants (aggregate income, construction costs lagged changes) are significant (except for lagged interest rates changes in OLS-FE), there is slope heterogeneity, and the cross-section dependence is high. When we allow for slope heterogeneity (MG), lagged changes in GDP lose statistical significance. When we control cross-section dependence (CCE-MG), lagged changes in construction costs also become non-significant. This result shows that conventional panel estimation methodologies that ignore slope heterogeneity and cross-section dependence produce inconsistent parameter estimates. CCE-MG and DCCE-MG produce estimates where we cannot reject that the slope coefficients are homogeneous. In the DCCE-MG method estimation, where the cross-section lags are added, we can obtain the most negligible cross-sectional dependence (0.02), although still significant, meaning that there are unobserved factors determining house prices beyond the variables included in the model (for instance, spill-over effects).

House price dynamics indicators are negative and significant (momentum and adjustment speed towards long-term equilibrium) in all the estimation methods except for OLS-RE. In CCE-MG and DCCE-MG estimations, the house price equilibrium fundamentals are not significant. Only momentum and adjustment speed explain housing returns in the short term. Moreover, the R-squared of these two estimation methods is much lower in the short-term equation (0.19; 0.33) than in the long-term equation (0.91; 0.90), suggesting a worse model fit to the data, even considering the significant momentum and adjustment speed.

The DCCE-MG estimate combines low cross-section dependence, slope coefficient homogeneity, and significant coefficients with lower standard errors.

Difference-in-differences

There is a parallel trend regarding the treated and control groups in both MSA before the August 2014 policy decision. The relevance of Density and Salaries in the regression was tested with the Likelihood ratio, rejecting the hypothesis that these two variables do not contribute to the model. The estimates of Equation 3 can be found in Table 3 below. The results were estimated with Eviews.

Table 3: Estimates of Equation 3

	Panel A - Lisbon MSA			Panel B - Porto MSA		
Specifications	**Equation 3**	**Equation 3**		**Equation 3**	**Equation 3**	
Estimator	**OLS-FE**	**FGLS-SURE**		**OLS-FE**	**FGLS-SURE**	
Intercept	14.173***	0.063		23.236***	2.490***	
	(0.955)	(0.081)		(2.075)	(0.162)	
Treated x Liberalisation	0.056***	0.274***		0.0355***	0.161***	
	(0.009)	(0.011)		(0.013)	(0.006)	
Density	-0.343***	0.096***		-1.157***	0.113***	
	(0.047)	(0.002)		(0.171)	(0.005)	
Salaries	-0.691***	0.872***		-1.216***	0.482***	
	(0.141)	(0.011)		(0.240)	(0.025)	
R^2 Adjusted	0.971	0.960		0.944	0.826	
F statistic	236.934***	181.053***		485.537***	1000.363***	
Observations	864	864		635	635	

Notes: The table reports the estimated coefficients from Equation 3 with the observations from the upper and lower quartiles of the sample. The dependent variable is the natural logarithm of housing prices. The independent variable is the interaction between Tourism and Liberalisation, a dummy variable with the value of one if the municipality is in the upper quartile of short-term rental intensity at the end of 2019 and if the month of observation is after August 2014 and otherwise zero. Density is the natural logarithm of the municipality density, and Salaries is the natural logarithm of the municipality mean gross monthly salary. OLS-FE is the Ordinary Least Square estimator with Fixed Effects both at the cross-section and period. FGLS-SUR is the Feasible Least Squares Estimator in a Seemingly Unrelated Regressions Equation system. Robust standard errors clustered at the municipality level are reported in parentheses. *** indicates significance at the 1% level. Standard errors in parentheses.

As expected, the estimated coefficients of the interaction terms are smaller with the OLS-FE estimator (5.6% in Panel A, 3.55% in Panel B) than with the FGLS-SUR (27.4% for Panel A, 16.1% for Panel B), with similar or smaller standard errors. With the OLS-FE as an estimator, the coefficients of the covariates have inverted signs, suggesting less precision in the estimates, as Hausman and Kuersteiner (2008) predicted. With the FGLS SUR estimator, the estimated coefficients of the covariates no longer have inverted signs, and their standard errors are much smaller than with the OLS-FE estimator. The F statistic is also much more significant with FGLS-SUR and confirms the findings of Brewer et al. (2017) regarding higher power tests when using FGLS-SUR.

Since the liberalisation of short-term rental registration in August 2014, housing prices have increased by 108% in the Lisbon municipality and 98% in the Porto municipality (as of December 2019). These municipalities have the largest shares of housing transferred to short-term rentals (over 5%) of the sample. We believe that the coefficients estimated in Equation 3 with the FGLS SUR estimator reflect more precisely the impact of a supply shock on housing prices.

4. Conclusions

In this study, we examined the outcomes of modelling house price dynamics utilising recent advances in econometric methods that permit heterogeneity and control for cross-sectional dependence. Additionally, we explored the application of a Difference-in-Differences methodology with innovative econometric estimators to gauge the impact of a policy decision on housing prices. Our study contributes to the literature by confirming that these new estimators yield higher power estimates by addressing traditional panel data issues such as serial correlation, slope heterogeneity, endogeneity, and cross-sectional dependence.

We conclude that conventional panel models, which assume homogeneous dynamics across sections and neglect cross-section dependence, such as OLS-FE and OLS-RE, yield significantly biased coefficient estimates that may result in erroneous conclusions. We demonstrate that the most effective econometric tool for estimating house prices' long-term equilibrium is the FMOLS-AMG, while for short-term dynamics, the DCCE-

MG proves to be optimal. Additionally, due to serial correlation in panel data, we illustrate that using the OLS-FE estimator in DID leads to biased estimates, whereas the FGLS-SUR estimator yields higher power estimates in this methodology.

These findings are crucial for researchers, housing market investors, and political decision-makers. This knowledge will facilitate a better understanding of the effects of policy decisions on housing prices, the determination of these prices in the short-term dynamics, and the factors influencing the housing market's long-term equilibrium.

References

Ashenfelter, O. & Card, D. (1985). USING THE LONGITUDINAL STRUCTURE OF EARNINGS TO ESTIMATE THE EFFECT OF TRAINING PROGRAMS. *The Review of Economics and Statistics*, *67*(4), 648-660.

Bakshi, G. S. & Chen, Z. (1994). Baby boom, population aging, and capital markets. *Journal of business*, 165-202.

Barron, K., Kung, E. & Prosepio, D. (2021), "The effect of home-sharing on house prices and rents:Evidence from airbnb", *Marketing Science*, Vol. 40 No. 1, pp. 23-47.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004), "How much should we trust differences-in-differences estimates?", *The Quarterly Journal of Economics*, Vol. 119 No. 1, pp. 249-275.

Brewer, M., Crossley, T.F. & Joyce, R. (2017), "Inference with difference-in-differences revisited", *Journal of Econometric Methods*, Vol. 7 No. 1, pp. 1-16.

Case, K.E. & Shiller, R.J. (1990), "Forecasting prices and excess returns in the housing market", Real Estate Economics, Vol. 18 No. 3, pp. 253-273.

Chudik, A. & Pesaran, M.H. (2015), "Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors", Journal of Econometrics, Vol. 188 No. 2, pp. 393-420, doi: 10.1016/j.jeconom.2015.03.007.

Cunha, A.M., & Lobão, J. (2022a). House price dynamics in Iberian Metropolitan Statistical Areas: slope heterogeneity, cross-sectional dependence and elasticities. *Journal of European Real Estate Research*, *15*(3), 444-462.

Cunha, A.M., & Lobão, J. (2022b). The effects of tourism on housing prices: applying a difference-in-differences methodology to the Portuguese market. *International Journal of Housing Markets and Analysis*, *15*(4), 762-779.

DiPasquale, D. & Wheaton, W. (1996). *Urban Economics and Real Estate Markets*. Prentice-Hall, ISBN 0-13-225244-9.

Eberhardt, M. & Teal, F. (2010), "Productivity analysis in global manufacturing production", Department of Economics, University of Oxford Discussion Paper 515.

Green, R.K., Malpezzi, S. & Mayo, S.K. (2005). Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources. *American Economic Review*, *95*(2), 334-339.

Hansen, C.B. (2007), "Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects", Journal of Econometrics, Vol. 140 No. 2, pp. 670-694.

Hausman, J. & Kuersteiner, G. (2008), "Difference in differences meets generalized least squares: higher order properties of hypotheses tests", *Journal of Econometrics*, Vol. 144 No. 2, pp. 371-391.

Hsiao, C. (1985). Benefits and limitations of panel data. Econometric Review, 4(1), 121-174. http://dx.doi.org/10.1080/07474938508800078

Malpezzi, S. (1999). A simple error correction model of house prices. *Journal of housing economics*, *8*(1), 27-62.

Mankiw, N. G., & Weil, D. N. (1989). The baby boom, the baby bust, and the housing market. *Regional science and urban economics*, *19*(2), 235-258.

Parks, R.W. (1967), "Efficient estimation of a system of regression equations when disturbances areboth serially and contemporaneously correlated", *Journal of the American Statistical Association*,Vol. 62° No. 318, pp. 500-509.

Pedroni, P. (2001). Fully modified OLS for heterogeneous cointegrated panels. In *Nonstationary panels, panel cointegration, and dynamic panels* (pp. 93-130). Emerald Group Publishing Limited.

Pesaran, M.H. (2006), "Estimation and inference in large heterogeneous panels with a multifactor error structure", Econometrica, Vol. 74 No. 4, pp. 967-1012, doi: 10.1111/j.1468-0262.2006.00692.x.

Pesaran, M.H. & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics*, *68*(1), 79-113.

Pesaran, M. H. & Yamagata, T. (2008), "Testing slope homogeneity in large panels", *Journal of Econometrics*, Vol. 142, Nº 1, pp. 50-93, https://doi.org/10.1016/j.jeconom.2007.05.010.

Poterba, J.M. (1991), "House price dynamics: the role of tax policy and demography", Brookings Papers on Economic Activity, Vol. 2, pp. 143-183.

Quigley, J.M. (1999), "Real estate prices and economic cycles", International Real Estate Review, Vol. 2, No. 1, pp. 1-20.

Ricardo, D. (1817), *The Principles of Political Economy and Taxation*, London: J.M. Dent and Son (reprinted, 1965).

Somerville, T., Wang, L. & Yang, Y. (2020). Using purchase restrictions to cool housing markets: A within-market analysis. *Journal of Urban Economics*, *115*, 103189.

Zellner, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests foraggregation bias", *Journal of the American Statistical Association*, Vol. 57 No. 298, pp. 348-368.