Acnowledging for spatial effects in the Portuguese housing markets

G. Carvalho, Pedro and Ribeiro, Alexandra

31 November 2007

Acknowledging for spatial effects in the Portuguese housing market

Pedro G. Carvalho
pguedes@ubi.pt
Alexandra Ribeiro
amribeiro@ubi.pt
U. Beira Interior – Covilhã – PORTUGAL

Abstract

The aim of this paper is to revisit a former paper on the Portuguese housing market (1995), acknowledging for spatial effects in order to interpret housing market changes over 1995-2001.
The paper will include a first section devoted to explain the differences between the OLS regression analysis and spatial econometrics, explaining the theoretical background used to develop a spatial lag model with the same database; the second section will show the misspecification problems we found when we ran the same model for after 1995-1998 databases; the third section is devoted to describe new housing literature findings relating housing market evolution with the macroeconomic cycles in Portugal; as a consequence the fourth section will include the method we developed with recent census data, to explain the evolution of the country macroeconomic cycles and the agents' new behavioural attitudes concerning housing; finally and using spatial analysis we can understand the main changes occurred over the 1995-2001 period.
The evaluation of the results contradicts some mainstream scholar and political knowledge to explain spatial inequalities between coast and interior municipalities. Complexity issues seem to be present when we consider the way different market agents make decisions on housing markets, looking this good either as a place to live or an alternative investment asset. In the concluding remarks we raise some new interesting questions for further research.

Keywords: housing markets; spatial econometrics and analysis; macroeconomics cycles; behavioural change; Portugal

JEL: C21 - Cross-Sectional Spatial Models; C51 - Model Construction and Estimation; R11 - Regional Economic Activity: Growth, Development, and Changes; R21 - Housing Demand; E32 - Business Fluctuations; Cycles; D01 - Microeconomic Behaviour: Underlying Principles

Part of the research was done during a visiting scholar period at REAL, University of Illinois at Urbana-Champaign 2000-2001.

Section 1
Motivation

In 1995 we ran a model to explain housing constructed area in Portugal using income proxy, population and housing construction costs as exogenous variables, among others. Introducing some dummy variables for county location pointed out that Algarve and the Coast municipalities revealed a different pattern against the rest of the country. However, heteroskedasticity induced by population was present and some of the variable data could raise reliability problems while statistics about housing was rather sparse and incomplete. At the time, only OLS estimations were made and going through spatial analysis study motivated us to replicate the model, using new instruments, new theoretical background and more powerful software.

In this paper we intend to acknowledge for spatial effects in two different ways:

(i) Including a new set of tests in order to improve the reliability of all the parameter estimates, looking for model stability over time (1995, 1996, 1997 and 1998) although using the same data and model specification;

(ii) Building and using different variables for 1995-2001, to understand the changes and be able to raise some new theoretical hypothesis for what happened in the Portuguese housing market.

Most of the experiments were carried out using GeoDa 0.95i software. [1]

The basic model

Is there a difference between housing prices across the 305 counties in Portugal?[2] As data on housing prices are not available we had to approach housing demand through quantities (square meters) of housing constructed area, implicitly assuming that Portuguese people will construct more where prices were lower.[3]

The model can be described in a summary table as follows:

[1] Free available in http://www.

[2] In the empirical design here presented we used data just for the 275 continental counties (out of 305), so we could compare the results with spatial econometric modeling. In fact, spatial analysis doesn't allow including isolated objects, which will be the case of island counties (the other 30).

[3] Data base in INE (1960-1999). Estatísticas da Construção e Habitação, INE.

Table 1 – Cross section model for 1990

Variable	Units	Main symbol	Log form	Per capita	Description
Housing constructed area	Sq. meters	A	LA	LAPC	Number of residential housing square meters constructed per municipality during the year
Income	$	INC	LINC	LINCPC	Proxy (automobile tax) of income per municipality during each year
Population	Number of individuals	P	LP	INVP (inverse)	Resident population. Census data for 1990
Cost	$/sq. meter	Cost	Lcost	Lm2 (per sq.meter)	Data from National Statistics Institute Local authority's annual reports.

The model was first estimated in absolute values and after transformed in per capita and/or square meter costs, ending up in the logarithm form:

$$lapc_{it} = lincpc_{it} + linvp_{it} + lm^2{}_{it} + \varepsilon_{it}$$

where $\boldsymbol{i} = 1,\ldots.275$ and $\boldsymbol{t} = 1990$

Results are shown in table 2:

Table 2 – Results for the cross section model (1990)

$R^2adj = 0{,}597$ and White χ^2 (13) = 0.0000

Variables	β-Coefficient	t - statistic	Probability
Constant	2,292	5,95	0,0000
linvp90	- 0.095	- 3,82	0,0000
lincpc90	0, 455	7,19	0,0000
lm290	- 0,492	- 7,33	0,0000
DummyALG	0, 697	4, 08	0,0000

As we can see built per capita housing area ($lapc_{it}$) was positively dependant on income, population (negative sign due to the inverse) and location in Algarve region (south tourism coast); negatively dependant on construction costs, which is consistent with economic theory.

Regressing the dependant variable on the population and the squared population we find evidence for marginal effects because there is a threshold for the rise of housing construction with the population increase, *e.g.*, after a certain demographic size there's a marginal decreasing effect (negative sign of the coefficient), so we could suspect that high and low dense municipalities show different patterns for housing construction. However we could not know about the possible correlation between housing construction in each municipality and the constructed area of its relative neighbours.

This is a good reason to update the same model and to proceed with other theoretical studies in which we were involved such as (Carvalho, 1998) where we claimed for the need for local housing policies; Portugal is a very small but heterogeneous country.

Section 2

The role of spatial analysis

Mainstream economic studies are usually focused in agents: each *agent* looks around him and makes his decisions considering all the other variables as if they were exogenous. But one must acknowledge that her decision will affect *other agent* decisions (among others see Wasserman and Faust, 1994; Willer, 1999; Capello, 2001), therefore social interaction becomes crucial and has to be endogeneized in econometric modeling (Anselin, 1999).

In the literature we can find several examples where this issue is addressed, *e.g.*, it does exist social interaction where endogenous and exogenous variables are highly correlated (Manski, 1993) or (Durlauf, 1999); neighborhood effects (Morenoff and Sampson, 1997); economies with strategic interaction (Brueckner, 2003, 2006); spatial externalities, agglomeration and spillover effects (Krugman, 1994).[4]

As known a number of economic policies are explicitly spatial. Housing policy is the case in the US where mortgage is non discriminative across states and banks are obliged to organize personal data for law enforcement purposes; health care/prevention, marketing business, elections, crime, environment and so forth, are also examples to be analyzed attending to spatial occurrences.

Using spatial frameworks it is not plausible to assume that an 'individual' behavior is not able to receive influence from any other neighbors. Spatial dependence is the rule and we know that closer the 'individual' is to others the more she depend on them (Anselin, 1999).

But 'closeness' in space has to be precisely defined[5], hence we need some metrics to arrange data in such a way that we can decrease the dependence effect. In other words we have to know the structure of the space so we can let dependence's influences apart.

Still following (Anselin, 1999), in spatial referenced data each attribute is always associated with location, which is known as the contiguity problem; objects can either

[4] For further review there is a very recent NBER paper by Scheinkman, José A. (accessed 2007) Social Interactions _ Princeton University and NBER.

[5] Other closeness definition is used in social network analysis, where the main idea is that node a_i within a network is considered as central if it can easily interact/communicate with all the others, in the shortest path through other well connected actors Wasserman, S. and K. Faust (1994).

be *points*, defined by coordinates (*x, y*) – such as in cities, stores, crimes, accidents, etc.; *lines*, defining arcs from node to node – such as road networks, transmission lines or *polygons*, as series of connected arcs – such as the boundaries of counties, states, census tracks.

But when we deal with counties, municipalities or towns it can happen that the entire sample is located in the same neighborhood and thus, the sample will not represent the distribution in the location. In this case we should be sure that, none sub-location would be let far apart, therefore the sampling rule should attend for location factors too. For instance, looking for spillover diffusion if one picks up the wrong villages, she will not observe interaction effects. In this case she will have to measure a radiant of interaction first and then, cluster the observations.

If we want to correct for spatial correlation in an event we should avoid spatial autocorrelation among observations and we want them to be far apart (*e.g.* do not ask two people from the same family). The same applies if we want to capture spatial or network autocorrelation, but here we should cluster the sample. When we do not choose the points (in regional-lattice data) they are discrete observations and not a sample, then we need to consider the neighbors. Studying a phenomenon we should follow the required path: to look at the structural part of the phenomenon otherwise; if there is no pattern it means the phenomenon is randomized and spatial analysis do not apply then.

Spatial effects – theoretical considerations[6]

In spatial econometrics each observation is unique so it doesn't help very much in the sake of significance to add new observations as in OLS estimations. However, it does not mean that everything is different from everything else (complex structure definition) and we should restrict some categories. In urban economics, for instance, we know that some observations follow a certain pattern; this is the case for 'distance to the CBD'; therefore, as the distance of the CBD increases it will decrease model validation. This is the heteroskedasticity problem and we should turn it as a function of a geographical variable.

[6] This section closely follows Anselin 2000's Spatial Econometrics classes at U. Illinois at UC.

a) Spatial dependence

In time series we have lags to explain Y_t and Y_{t-1} where the subscript *t-1* stands for the previous observation. Instead, in spatial data *t-1* stands for the neighbor observation. But then which neighbor are we referring to?

Complexity comes from the two-dimensional and multi-directional aspects: which neighbor and which influence direction?. If observations are too correlated, we will have less information than the *N* observations in the sample. In cases where we have substantive spatial dependence we should add a new variable, including *Wy* as explanatory variable in the model. In the case we want to correct for spatial spillovers, spatial dependence will be a nuisance in the model and we have to correct the error term to improve precision distinguishing structural dependence from contagious effects. In a cross sectional data, it could be something identical to interpret the phenomenon through a unique picture, as when we are looking to a football game picture and have to guess what are the players' movement (attack, defense?). Each observation is equivalent and gives us the same equivalent information.

b) Specifying spatial dependence

Usually we have two techniques to deal with: contagion and apparent contagion.

In a disease – we want to know if one contagious the others (contagion), while if there is an event we want to know which is the responsible for them all, *e.g.* pollution, radioactivity (apparent contagious).

In a spatial lag problem we want to know how the dependant variable depends on the value of the neighbor (closest and distant) variable, and the + or – signs will tell us about the kind of reactions. If we want to correct the spillover effects, we want the error term to be the lowest possible, and then we should explore the error term in order to separate the two types of effects.

How should then we specify the models?

In lag dependence models we do it including *Wy*, where the new variable expresses an average of the *n* considered neighbors according to defined criteria. Then the model $Y = XB + e$ will get a new form $Y = rWy + XB + e$, where the *r* coefficient will measure the strength of autocorrelation. If *r* is near 1 we have perfect autocorrelation. As neighbors are defined ex-ante, we 'clean' the structure from the spatial autocorrelation.

Because of interaction effects, the *r* coefficient goes together with *W* and so it can reach the maximum value at 1; however, the minimum value can be less than –1.

c) <u>Interpretation of spatial lags</u>

The first thing we have to be sure is that the unity of observation is the same that the theoretical unity. Studying illnesses at the county level we should be aware that the contaminated are not counties and vice-versa. If we have a phenomenon in a particular cluster and we use a grid rather than this cluster, there will have a lag. But we can get rid of the lag, just by changing the scale. In this case we have just a scale problem and not an interaction problem. We call this process 'data cleaning', looking at the scale and not overselling the results.

d) <u>Interpretation of error terms</u> (specification problem)

In error term we have got the problem of the non-observed effects. If we bring some of them to the model, there the correlation could be gone. It means that we were considering a misspecification error in the model. But if we ignore space we are ignoring a meaningful variable. To look for missing variables is always a useful exercise; even if we loose some model efficiency we will gain robustness for the theoretical explanation.

Spatial econometrics shows us the difference between structure and spatial dependence.

e) <u>Structuring spatial dependence</u>

The main question is that we have two dimensions to model with one data set and the goal is to know which locations (among *i, j)* do interact.

We have *N* observations to estimate *N*(N-1)/2* interactions and the more we add observations, the less information we have, because the number of interactions will rise exponentially. Therefore we need to impose 'some' kind of structure and diminish the sample so we can decrease the interaction even knowing that there's no ideal structure.

It is known that gravity modeling is used to define distance. In social networks, the point or location does not deal with the same concept of distance; location within the social network is the important thing (Carvalho, 2002).

As space is inter-correlated we have multi-dimensional correlation in the same framework, hence we must define a pattern (those that are correlated) and assume this is the structure. Doing so, we reduce it to one-dimensional problem.

f) Spatial weights

The first thing we have to do is to define county neighbors. For each location *I*, we define the *W* matrix of neighbors, adopting the convention of 1 for neighbors and 0 for non-neighbors. Each location is considered a 0 to itself (there's no reflexive locations). There are three main criteria to define neighbors.[7] When we are defining neighborhood through boundaries, we will get a symmetric matrix *W*, because if *i* has a common boarder with *j*, then *j* has the common border with *i*.[8]

General spatial weights

Cliff-Ord weight matrix is defined as the one reflecting the potential interaction between two regions, with $W_{ij} = (1/d_{ij})^a * (b_{ij})^b$, where distance is (*d*) and boundary is (*b*). The distance parameter is negative, because interaction tends to decay with distance.

These weights may contain parameters like scaling factors and weights. Anyway, whenever this is the case we should consider only one of them. Either we fix the factor and go with weights or then we fix the weights and move the scaling factors. But as we know, distance does not work properly all over the locations (*e.g.*, counties that are very large and other that are very small). If we choose the shortest distance it could happen that a large county will have no neighbors. In this case we should use the *k-nearest* neighbors (independent of the distance each state has to have *k* neighbors).[9] We also can use economic weights, blocking the structure by income/capita similarities, by employment measures and so forth. But whatever the criteria we use, it is important to acknowledge that this will drive the results and becomes essential to the interpretation.

Row-standardized weights

All the *W* matrices should be comparable so the sum of *W* should equal 1. It means that the more neighbors we decide to impose, the lower will be each individual effect, or if we decide 4 neighbors it will mean that each one of them will be weighted by 0,25 and deciding for 5 neighbors with no considerations about their size, it will decrease the individual effect to 0,20.

[7] GeoDa software defines them automatically: ROOK, those which have common boundaries; BISHOP – those which have corners with the location and QUEEN – both corners and common borders.

[8] If we define migration flows, then it could be non-symmetric, because people living from one region to another do not have to be the same then the reverse direction flow. We are imposing a structure and we should see if it's adjustably defined, reason why specification is very important.

[9] This is not the case for Portugal and the study we carried on.

If we attend to the number of neighbors and their geographical size, we have to re-scale them in such a way that the sum of W still equals 1.

Spatial Autoregressive coefficients

What will it happen to the model if we change the weights?

We must not forget that for r, there are asymmetric effects, going from less than -1, up to $+1$. The more neighbors we will consider the lower value we will get for r.

And in spatial regressions, correlation has not the conventional tradition of a high R^2, and we can have meaningful econometric results with very low values for R^2.

Characteristics of Spatial Weights

In spatial econometrics it is very important to clearly acknowledge what it's being done, hence it is reasonable to check the W matrix in several forms. First of all we must check if there are no islands, which means, any location with no neighbors.[10] If they exist they must be excluded from the data, because they have no (contiguous) connections, so they do not interact with other locations. In the W output we can check it looking at some statistical results (frequency of connections, number of islands, number of strongly connected, etc.).

Higher order contiguity

Similarly to social network analysis, we face a difficult problem in spatial contiguity that we do not have in regular time series. While in time series we have sequences of data, such as 2000 is preceded by 1999, 1998 and 1998 and 1999 can not be just one time-lag from 2000, in spatial data it is possible that two observations appear to be one dimension-lag from the same location. Therefore we must use the recursive definition (or step path concept):

j is contiguous to *i* in order *p*

j is first order contiguous to k

k contiguous to *i* of order *p-1*.

This is known as the concept of redundancy. Because 5 can be first order contiguous of 1; but as the relation is interactive, 1 is first order contiguous of 5. In this case 1 is second order contiguous of itself (through 5).

In spatial econometrics we have to skip redundant contiguities, otherwise we are overlapping neighborhoods and committing a misspecification error.

[10] In this case study this is the reason why we considered just 275 out of 305 municipalities, once 30 are the islands of Azores and Madeira.

The way to deal with it is to calculate the square of the contiguity matrix. That will tell us the number of steps to get the interaction: if they equal 1 it is important and if they are greater than 1, they should be excised.

Section 3

Revisiting 1990's model[11]

Selecting he meaningful spatial regimes

The first thing we have done was to create six spatial regimes to chose two of them, those that fitted the country and presenting the best results (magnitude of explanation and higher *t*-values), respectively ALG (municipalities located in Algarve region) and COAST (municipalities with Atlantic border).[12] The ALG and the COAST regimes contain the municipalities with Atlantic coast, which means that they partially overlap reason why we never used them together.

Spatial ANOVA for 2 spatial regimes

There is a highly significant difference between the average of the dependent variable between the two regions (ALG, 3 times more than non-ALG and COAST, 1.5 times more than non-COAST). Those were the chosen regimes for all the following experiments.

Testing for Heteroskedasticity using regimes

The basic model presents values for adjusted R^2 of 0,579 and 0,540 (COAST and ALG respectively); all the variables are significant at a $p<0.01$ although the variable for population - *linvp90* - has not the expected negative sign due to the marginal effects we explain before. We will get to this later (p.12).

The null hypothesis for error normality in both regimes is rejected (Jaques-Bera test with very low probability values).

The null hypothesis for homoskedasticity is clearly accepted in both regimes, considering the Koenker-Bassett test (probabilities of 0.76 and 0.79). However, the results are different for individual coefficient stability. In fact, stability for housing cost is rejected in COAST regime, which leads us to abandon this regime in the final model.

[11] The software used for the results in this section were ArcView 3.2. and SpaceStat.

[12] The other four regimes referred to LIT (coast, including the set of the counties within a certain distance from the west coast of the country); CDIST (the 18 administrative counties); AMP and AML (metropolitan areas of Porto and Lisbon).

Spatial homogeneity

Using the spatial expansion model (1^{st} order), heterogeneity is the rule with exception for the variable population in both regimes and heteroskedasticity is still remaining. Using the spatial expansion model (2^{nd} order), heterogeneity disappear (exception to the constant) and heteroskedasticity is still present.

Spatial Weights

In order to run spatial analysis, we created several weight matrices (W_{it}).

In Table 3 we show the results (just the probabilities) for seven W matrices:

Table 3 – Results for the Weight matrices

Diagnostic	*rook*	*queen*	*invd1s*	*kn_3_3s*	*kn_3_2s*	*2_2*	*2_3*
Moran's I (error)	0,003312	0,003335	0,000000	0,008665	0,010042	0.000002	0.000689
LM (error)	0,006769	0,006840	0,000916	0,014659	0,015393	0.000014	0.002379
Robust LM (error)	0,267000	0,278842	0,497805	0,374953	0,349358	0.946761	0.427486
K-R (error)	0,012743	0,014310	0,906666	0,077545	0,309251	0.000403	0.000028
LM (lag)	0,000006	0,000007	0,00000	0,000057	0,000088	0.000000	0.000001
Robust LM (lag)	0,000147	0,000171	0,000067	0,000892	0,001274	0.000100	0.000078
LM (SARMA)	0,000019	0,000022	0,000001	0,000204	0,000296	0.000000	0.000004

To chose the adjusted matrix we looked sequentially at Moran's I probabilities, LM (error and lag), Robust (error and lag) and LM (SARMA).

The conclusion is that the ***spatial model*** should be the correct specification in this case, once though the LM (error) presents low probabilities and the robust LM never indicates that possibility. The 1990's model remained significant for the spatial lag with the two referred spatial regimes; however, probabilities are slightly greater overall.

Nevertheless heteroskedasticity still remains after running the (FGLS) model and the Wald test they do not allow us to reject the null hypothesis of homoskedasticity in both cases, although for a strict margin.

Spatial Lag and Spatial Error models

We estimated both regression models for *lag* and *error* dependence and the results indicate that:

(i) Spatial lag is significant with ρ = 0,869, with no heteroskedasticity for the COAST regime; there is no spatial autocorrelation remaining in both cases, once the Likelihood Ratio rejects the null hypothesis of $\rho = 0$;

(ii) Spatial Error model presents significant λ estimates, but heteroskedasticity and spatial autocorrelation are still remaining in both regimes.

After these results, we concluded that the best-fit indicator for all the models is the spatial lag model with the regime COAST.

Comparative results – acknowledging for spatial effects

Comparing this spatial model with the 1990's OLS estimation, we loose some fitness (LIK is a little bit higher in absolute value – 179.460 vs. 160.322) but we achieve to correct for spatial dependence; besides, the advantage is mostly theoretical because acknowledging the heteroskedasticity problem, which totally disappears with the spatial lag specification.

Comparing with previous work (Carvalho, 1998), there are some slightly different estimates in the coefficients, but a lot more precision incorporating the spatial interaction between the municipalities, which changed the magnitude and significance of the coefficients, as stated in table 4:

Table 4 – Comparative results

	Normal OLS			Spatial lag model		
Variables	Coefficient β	**t** - statistic	Prob.	Coefficient β	**z** - value	Prob.
Constant	2,292	5,95	0,0000	2,19662	5,01	0,0000
linvp90	- 0,095	- 3,82	0,0000	0,23268	5,99	0,0000
lincpc90	0,455	7,19	0,0000	0,45206	7,96	0,0000
lm²90	- 0,492	- 7,33	0,0000	-0,27341	- 11,8	0,0000
DummyALG	0, 697	4, 08	0,0000	--	--	--
Lambda (λ)				0,813	6,46	0,0000

Another difference between the two models refers the different sign for the variable population (*linvp90).* OLS estimation the predicted sign is negative, once the higher the value of resident population in each county the lower will be its inverse, affecting thus negatively the housing constructed area, which follow other similar theoretical literature on housing demand; however, specifying a spatial lag model, enables us to capture the decreasing rate of housing construction in highly dense municipalities once the lag smoothes the differences among neighbour counties. Moreover, with other specification, including both the variable and the logarithm of the variable (*invp90* and *linvp90*) we get to improve the model fitness and the sign turns as expected; this confirms the previous theoretical hypothesis that population plays a non-linear role on market structure and there is a decreasing effect on constructed housing, the more dense is the miunicipality.

Estimating the model with both *linvp90* and *l2invp90* (power 2) the results are consistent with our hypothesis and population coefficients become negative.

<u>IV estimation</u>

In all the regressions we ran the IV (instrumental variables) for spatial lag model, the coefficient of the lagged variable ρ was always higher than 1. We tested for endogeneity with all the previous exogenous variables and the results were consistent with non-endogeneity.

Evaluative conclusions

The updated model, acknowledging for spatial effects, is broadly consistent with the former study, but showed more accurate estimates (Table 4), namely because municipality location in ALG happens to be less conclusive related to all the other COAST municipalities. The first model captured (with the dummy variable), a short-term effect - the tourist boom construction in ALG during the 90's - which revealed to be non significant correcting for spatial effects.

Section 4

Replicating the model for sequent periods

A number of discussions and debate with peers told me that variables' quality and reliability should be introducing a bias and it could be interesting to recover the model using different data sets for after 1990 years.

As a consequence we ran the basic model for different years with the same variables and sources (1995, 1996, 1997 and 1998); the model lost most of its significance at acceptable levels. This unabled the panel model we wanted to design but enabled new theoretical considerations on housing markets in Portugal:

- First, the proxy data for the income variable (municipal car tax) is not consistent for most recent years. The explanation is that Portugal abandoned the central regime for mortgage credit in 1992 and therefore, we knew that housing loans were highly correlated with car sales and other consumption goods, because people got loans for everything with the mortgage; after the credit liberalization process in 1992, those two variables became clearly disentangled;
- Second, accessing different credit sources, housing construction became more homogenous across continental space, and different spatial regimes turned housing construction more dependant on population than on the available

income in each county and probably changed agents' behavioural concerning consumption and investment decisions;

- Third, national transport infrastructure improvement changed labour mobility patterns an disconnected 'place to work' from 'place to live' :
- Fourth, housing is a very special commodity for Portuguese culture and the construction sector plays an important role for mayors to negotiate municipality development plans;
- Finally, we need more accurate databases on housing prices and housing construction in order to get reliable variables.

Section 5
Updating housing theory through recent literature

According to Goodman (2005), most of the older American cities lost population during the last 20th century three decades, but while cities such as Boston or New York experimented stable dwelling units or even reveal slight increases, others such as Buffalo, St. Louis, Cleveland, Detroit, and Pittsburgh lost large fractions of their dwelling units.

Riddel (2004) also recognized that recent research concludes that the market for owner-occupied housing is often inefficient and adjusts slowly to changes in market conditions quoting Case and Schiller (1989) and DiPasquale and Wheaton (1994, 1995), who found strong evidence that it takes several years for market changes to be fully incorporated into housing prices. Mankiw and Weil (1989) also find that market prices respond slowly to changing demographic forces.

Selective programs and subsidies have an impact on both the financial position and the housing conditions of the household to whom they are allocated. They also affect the equilibrium outcome in housing markets. A study by Nordvik (2006) analyzes how the housing stock in Norwegian municipalities is affected by selective targeted interventions on the supply and demand-sides of the market. The empirical analysis shows that additions to the stock of public housing, increases the total housing stock. For every 100 new public units built, 60 units are added to the total housing stock. Demand-side subsidies are also shown to increase the size of the housing stock.

Lastrapes (2006) says other dynamic models have been developed relying on the asset view of housing: housing is a durable good, the demand for which reflects both the service flow and asset value of housing units. On the margin, the return from housing must equal the return on alternative assets. This view of housing demand is consistent with more general models of durable goods, as in Obstfeld and Rogoff (1996) and Kau and Keenan (1980), and has been used often in policy analysis of housing markets, as in Bruce and Holtz-Eakin (1999), Poterba (1984), and Miles (1994).

He (2003) highlights the existence of two different submarkets – existing and new homes and that stock market returns may directly and indirectly affect consumer confidence according with macroeconomic cycles. Pollakowski and Ray (1997) admit the diffusion of price changes and that housing prices in one location can be affected either by neighbour housing prices or housing types. Both He (2003) and Pollakowski and Ray (1997) consider that subnational results are likely to be economic rather than behavioural; spillover effects would not necessarily occur between neighbouring areas but would be based on economic interrelationships, admitting an opening research area to see if housing price behaviour is a function of economic behaviour.

In Portugal we argue there is a difference between agents' behaviours in housing markets because the country lost competitiveness in tradable goods and housing became a very special durable asset neither competing with the financial assets (that still is a weak market) nor tied with demographic flows (because of cultural values).

Chen (2006) state there is strong statistical evidence that the long run movements of aggregate consumption, disposable income, housing wealth, and financial wealth are tied together. However, the evidence also suggests that short run variations in the Swedish housing market are largely dissociated with consumer spending. Meanwhile, it is shown that the strength of the linkage between consumption and housing wealth is not sensitive to different model specifications and various measures of key variables.

This literature review stimulated us to look for the connection between macroeconomic cycles and housing starts in Portugal (1998-2005).[13]

Macroeconomic cycles and Housing in Portugal (1998-2005)

Housing, as a durable good, impacts economy and society through the construction sector and this implies to understand the supply side of the market, the agents' segments and behaviours more than housing production itself. "*The starting point to the theory of two-sided markets is that an end-user does not internalize the welfare impact of his use of the platform on other end-users*" (Tirole, 2004); although the main examples

[13] We would like to thank Jaime Silva, a former student I supervised during his undergraduate project for providing the estimation results.

developed among this microeconomic stream are applied to network services production such as telecommunications and banking services, in our viewpoint this also applies to the case of the housing market.

In order to understand if there is a relationship between stock market and housing starts the model we referred to (OLS log-linear model) came like:

$$Y = \beta_1 + \underset{(+)}{\beta_2 X_2} + \underset{(-)}{\beta_3 X_3} + \underset{(+)}{\beta_4 X_4} + \underset{(+)}{\beta_5 X_5} + \beta_6 X_6 + \varepsilon_i$$

where Y stands for new housing starts, X2 for GDP, X3 for mortgage interest rate, X4 for general stock market index, X5 for unemployment rate and X6 for construction costs.

The results for 2 models are presented in the next table:

Table

Housing Supply Estimates

Note: t-values in parentesis; * p<0,05 ** p<0,01 significance levels

Variables	Y = log New Starts Model 1	Y = New permits Model 2
Constant	17,203 (15,837)**	25,620 (5,85)**
Log GDP	Not significant	-1,143 (-3,03)**
Log Interest Rate	Not significant	-0,189 (-2,45)*
Log Stock Market	-0,569 (-5,181)**	-0,198 (-3,40)**
Log Unemloyment	-0,665 (-8,015)**	-0,470 (-6,54)**
Log do Índice de custos	Not significant	Not significant
F- stat	35,89**	21,58**
Time Period	1998:1 2005:3	1998:1 2005:3
R^2	0,719	0,769
Observations	31	31

We can see that all the independent variables show a negative relationship with the dependent variable either it is represented by the new built housing or by the new permits, and do explain 77% of it.

These results contradict some of the results in the literature for other countries once expansive periods usually have a positive direct and indirect impact on new housing starts; the explanation we suggest for this result is that in Portugal, housing represent a type of alternative investment during recessive periods when the overall economic dynamics is weak once housing depreciates at a lower rate (over 1%) than other durable goods (Davis e Heathcote, 2002). In Portugal housing market looks like functioning against the macroeconomic cycle.

Moreover, construction costs were not significant in both models, which can explain that in this global housing market agent's behaviour is not exclusively oriented by economic rationale. The result for the interest rate confirms this because it plays a significant role in model 2 and not in the model 1.

Another relevant point is that stock market index show that supply either than demand, leads the market forces. One should expect that positive stock market returns would raise people confidence to invest in housing through the capital market; however in bull market periods, housing supplier agents shifted their assets to capital markets, seeking higher rents but housing demand is increasing, pulled by low interest rates.

As we did not find any reliable evidence we assume there must be some misspecification problem as a consequence that none applied spatial analysis.

Section 6
The acknowledgment for spatial effects

The aim of this section is to raise some new issues to understand the Portuguese housing market development in the late half of 20th decade.

Our urban, regional and territorial background built during several concrete experiments we have done tells us that housing market in Portugal is inefficient and mostly pulled by the supply side; this is contrasting with most of the mainstream literature in the field that mostly focus on the demand side of the market.

Our hypotheses are:

(i) Housing construction is an alternative industry to capital application during recessive macroeconomic cycles and weak economic dynamics on tradable goods industries.

(ii) Housing market is supply pulled and inefficient, spatially clustered around the central city regions;

(iii) There was a shift in the spatial pattern of the country in the last 20th century decade, namely what should be considered as lagged municipalities.

Methodology

Due to remaining specific problems on the National Institute of Statistics data and the lack of reliable longitudinal data sets for housing prices it would be required some caution with results interpretation and extended conclusions.

As a consequence it was important to verify our intuition through a proxy approach. We need to know if supply exceeded demand over the second half of the 20th century and if it did randomly happen across the 275 municipalities of the continent. This procedure required some starting assumptions. First of all (i) to assume that housing market was cleared in 1995; (ii) demand was fully fulfilled by existent stock, ignoring shelter quality levels and (iii) spatial homogeneity among municipalities. Therefore we built the housing supply in 2001 summing the existent stock in 1995 with every year new constructions (1995-01) and the vacant number in the last year getting a Housing Supply equation like:

[1] $HS_i^{2001} = Stock_i^{1995} + \sum_{1995}^{2001} NHC_i + VH_i^{2001}$ with $i = 1,\ldots 275$

where *HS* stands for housing supply in 2001, *Stock* for the existent houses in time 0 (1995), *NHC* for new housing construction (1996-2001), *VH* for vacant houses in time 1 (2001) and *i* for each municipality.

Knowing the average people by household (IND/AG_{2001}) and the population variation between time 0 and 1 ($VPop_{95\text{-}01}$) we got the housing demand needs in 2001 through the following equation:

[2] $HD_i^{2001} = \dfrac{VPop_i^{95-01}}{IND/AG_i^{2001}}$ with $i = 1,\ldots.275$

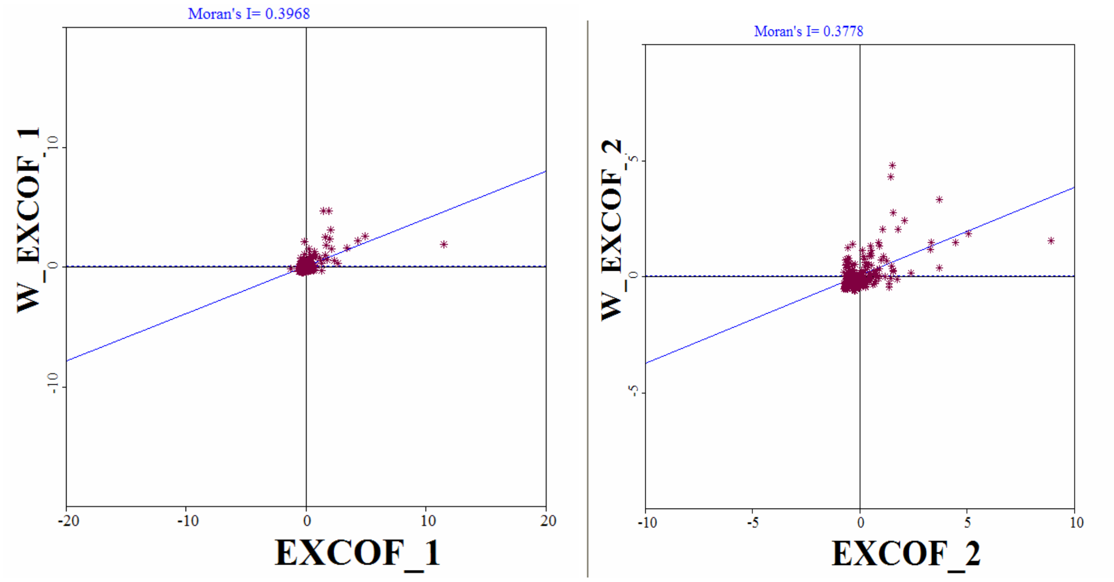
where *HD* stands for housing demand in 2001. Using both equations we got the Supply Surplus (*ExcOf_1*) through [1] – [2]; we also built an alternative proxy for Supply Surplus (*ExcOf_2*) through Census data (*Housing stock*$_{2001}$ – *Classic Households*$_{2001}$).

To relate these variables with income per municipality we used the National Statistics index for average purchasing power per municipality IPCC.

Moran Scatter Plot

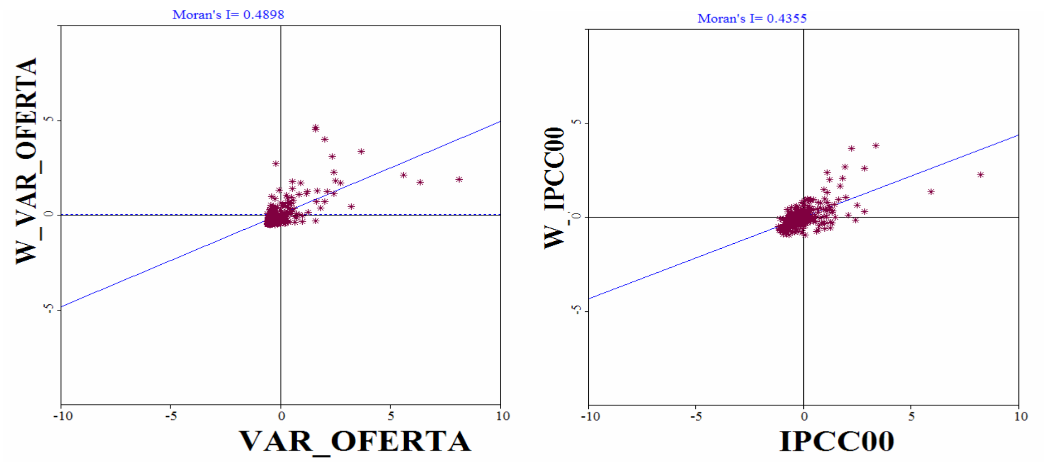
Moran's *I* spatial autocorrelation statistic is visualized as the slope in the scatter plot with the spatially lagged variable on the vertical axis and the original variable on the horizontal axis. The variables are standardized to facilitate interpretation and categorization of the type of spatial autocorrelation (cluster or outlier).

The slope of the regression line is Moran's *I* statistic, indicated at the top of the window. The four quadrants in the scatter plot correspond to different types of spatial correlation. Spatial clusters in the upper right (high-high) and lower left (low-low) quadrants, and spatial outliers in the lower right (high-low) and upper left (low-high) quadrants. Note that the magnitude of Moran's *I* as such does not indicate significance, nor are the statistics directly comparable across weights and variables.



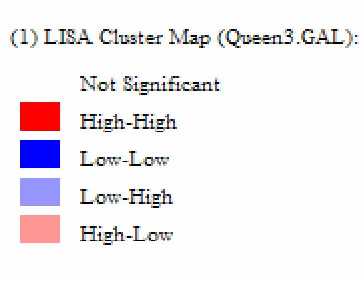
The two figures above show there is no significant difference between the two forms we proxy for housing supply surplus although Excof_1 is less sparse.

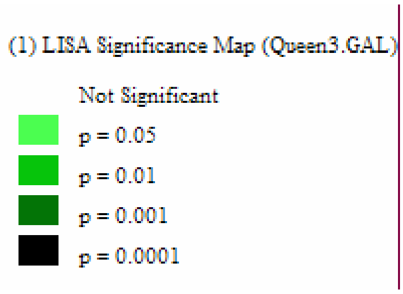
Considering the housing supply change 1995-2001 (VAR_OFERTA) and the income per capita for 2000 (IPCC00), the Moran's *I* show the same kind of spatial clustering.



LISA Maps

Local measures of spatial autocorrelation are implemented as LISA maps for the univariate case as well as for the bivariate and standardized rate case. All three work in the same fashion. The significance map shows the locations with significant Local Moran in different shades of green, depending on the degree of significance. The cluster map (LISA map) shows the significant locations by type of association.





It is important to look at the spatial patterns of average growth of housing supply to examine the spillover effects. If comparison is made of the spatial clustering of both growth rates and initial and final housing stock, then the dynamism between regions can be related to their neighbours' dynamism. Therefore if a neighbour relation has a positive slope (HH and LL) spillover and complementarities do exist and spatial interactions predict people mobility between municipalities.

With this legend we present the maps for the variables we choose: Housing supply change (Var_Oferta), Supply Surplus (ExcOf_1), Vacancy Rate (TX_VAGOS) Demand Variation (Var_Procura), Purchasing Power Index (IPCC00) and Population Change 1995-2001 (VarPop95_01).

«See Annex1»

Looking at the spatial autocorrelation the HH clustering is limited to the metropolitan areas Lisbon and Oporto, while the LL clustering occur in the interior regions around the main mid-cities showing the urbanization process with higher significance in the 3 administrative cities where Universities or Colleges were built during the 20th century 80 and 90 decades and new road infrastructures were built (Vila Real - north, C. Branco - center and Évora - south).

«See annex2»

Considering the higher number of clustering in vacant housing we assist to the depopulation effect on Alentejo region and within the inner metropolitan areas and also in some suburban regions around the metropolitan Oporto area and also in the south Algarve capital - Faro.

«see annex3»

The clustering effect is again concentred around the metropolitan regions of Lisbon and Oporto, showing their persistent pulling effect, together with the clustering of economic activity.

«See Annex4»

Nevertheless, the purchasing power is broadly located in the whole northern region of the country revealing an interesting spatial pattern, besides the metropolitan regions.

Spatial econometrics

We also intended to know if there is an expected interrelation between the new housing construction and population change, its purchasing power and housing costs. We regressed through OLS squares and spatial lag models the equations:

Supply change = Population change

Supply change = Purchasing power

Supply change = cost

There is no significant changes considering the theory that housing will be built in the places where population raises and income is higher. The "new thing" considering former studies presented is that Housing Cost does not relate negatively as it could be expected with new housing constructions.

The explanation we find is that either the reliability of the statistics for cost is not good or that people does not anymore build their housings because of their needs or cost but because interest rates are low and they do not have other alternative capital applications, functioning as an alternative investment during recession periods as we predicted before.

Conclusions

Considering the different purposes of the paper, we can easily conclude that spatial analysis is appropriate for economic markets such as housing where what happens in one municipality would be intertwined with what happen in neighbour municipalities.

The second conclusion is that for housing market analysis, spatial analysis enables us to specify econometric models more accurately; therefore we should always acknowledge that.

Thirdly, when we deal with unreliable data we should always explore spatial analysis through instruments such as Moran's I and Lisa maps frameworks once they constitute a helpful guide to see the 'picture' we want to search.

Finally we conclude there is a long way to go on housing research because a number of mainstream findings in the literature concerning housing markets evolution and the macroeconomic cycles of the countries should pay deeper attention to the agents' new behavioural attitudes concerning housing.

In Portugal the evaluation of the results contradicts some mainstream scholar and political knowledge to explain spatial inequalities between coast and interior municipalities. Complexity issues seem to be present when we consider the way different market agents make decisions on housing markets, looking this good either as a place to live or an alternative investment asset. In the concluding remarks we raise some new interesting questions for further research.

References

Bruce, D., and Holtz-Eakin, D. (1999). "Apocalypse Now? Fundamental Tax Reform and Residential Housing Values," *Journal of Housing Economics,* **8,** 249–271.

Carvalho, Pedro G. (1995) on Housing Market in Portugal: a cross section analysis for 305 municipalities. Msc dissertation, FEUC, September, 1995.

Carvalho, Pedro G. (1998) *The need of local housing policies: a research based argument. In Housing Patchworks*, Chapter 5, Delft University, Ed.

Carvalho, Pedro G. (2002) Uma análise das Redes Institucionais: uma aplicação para a competitividade nas cidades, PhD, UBI, 2002, June.

Case, K.E., Schiller, R.J. (1989) The efficiency of the market for single family homes. *American Economic Review*, 79, 125–137.

Chen, Jie (2006) Re-evaluating the association between housing wealth and aggregate consumption: New evidence from Sweden, *Journal of Housing Economics,* 15, 321–348.

Davis, M. e Heathcote, J. (2002), "Housing and the business cycle", Board of Governors of the Federal Reserve System, USA and Georgtown University, USA, http://www.federalreserve.gov/pubs/feds/2004/200411/200411pap.pdf
[accessed 15.11.2005]

DiPasquale, D., Wheaton, W. (1994) Housing market dynamics and the future of housing prices. Journal of Urban Economics, 35, pp:1–27.

DiPasquale, D., Wheaton, W. (1995) *Urban Economics and Real Estate Markets*. Prentice Hall, Englewood Cliffs, NJ.

GOODMAN, Allen C. (2005) Central cities and housing supply: Growth and decline in US cities, *Journal of Housing Economics*, 14, pp: 315–335.

He, Ling T. (2003) The Effects of Real Stock Returns on Sale of New and Existing Homes, *Journal of Housing Research*, Volume 13 (2), 199-217.

Kau, J. B., and Keenan, D. (1980). "The Theory of Housing and Interest Rates," *Journal of Financial Quantitative Analysis,* **XV,** 833–847.

William D. Lastrapes (2002) The Real Price of Housing and Money Supply Shocks: Time Series Evidence and Theoretical Simulations, *Journal of Housing Economics*, 11, pp: 40–74.

Mankiw, N.G., Weil, D.N. (1989) The baby boom, the baby bust, and the housing market. *Journal of Regional Science and Urban Economics*, 19, pp: 235–258.

Miles, D. (1994). *Housing, Financial Markets and the Wider Economy*, New York: Wiley.

Nordvik, Viggo (2006) Selective housing policy in local housing markets and the supply of housing, *Journal of Housing Economics*, 15, pp: 279–292.

Obstfeld, M., and Rogoff, K. (1996). *Foundations of International Macroeconomics*, Cambridge: MIT Press.

Pollakowski, Henry O. and Ray, Traci S. (1997) Housing Pricing Diffusion Patterns at Different Aggregation Levels: An Examination of Housing Market Efficiency, Journal of Housing Research, Voliume 8 (1), 107-124.

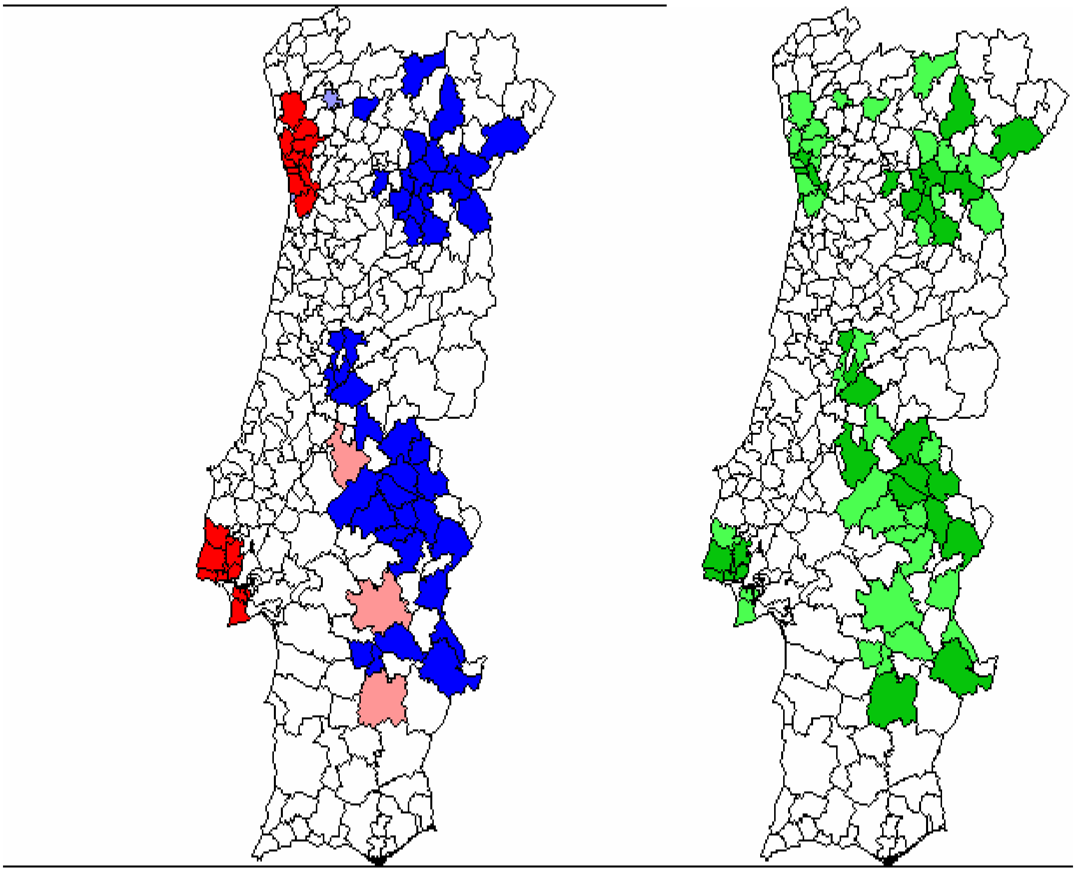
Poterba, J. M. (1984). "Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach," *Quarterly Journal of Economics,* 99, pp: 729–752.

Riddel, Mary (2004) Housing-market disequilibrium: an examination of housing-market price and stock dynamics 1967–1998, *Journal of Housing Economics*, 13, pp: 120–135.
Scheinkman, José A. (accessed 2007) Social Interactions _ Princeton University and NBER.

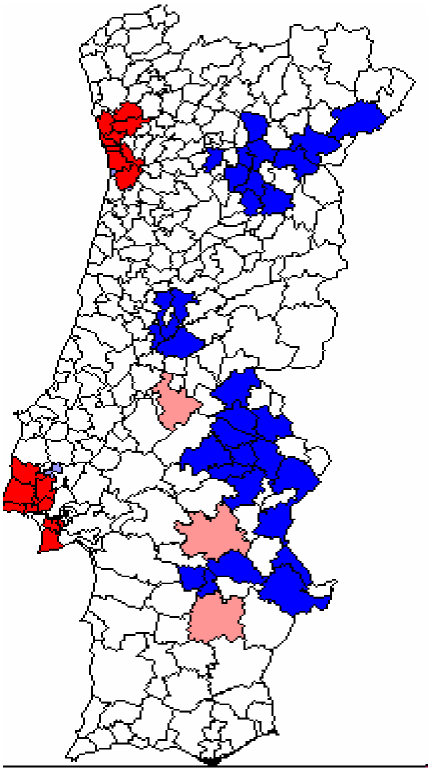
Tirole, J. e Rochet, J. (2004), "Defining Two-Sided Markets", Toulouse; http://www.brousseau.info/semnum/pdf/2004-03-01_tirole.pdf [accessed 22/02/2006]

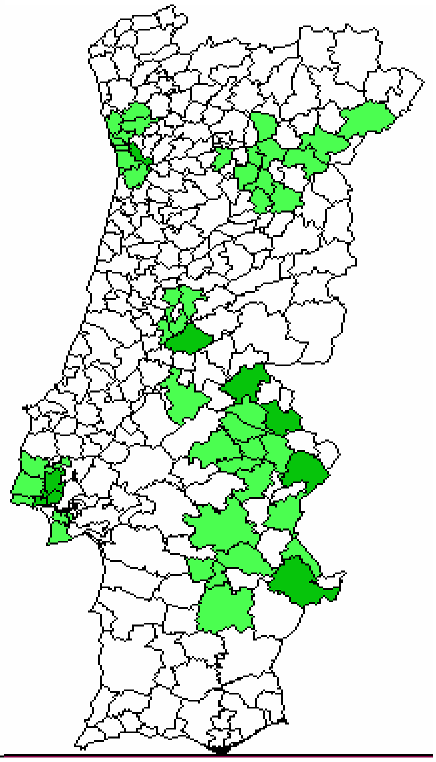
ANNEX
«Annex1»

Var_Oferta (housing supply change)

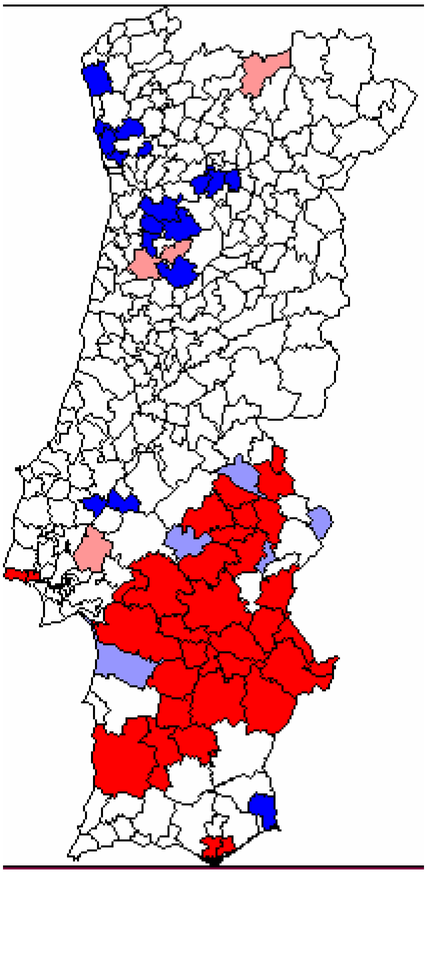


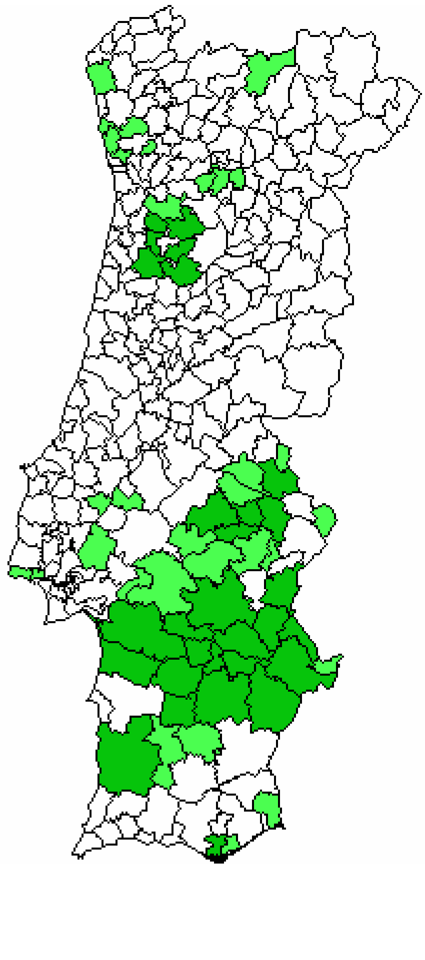
«Annex2»
ExcOF_1



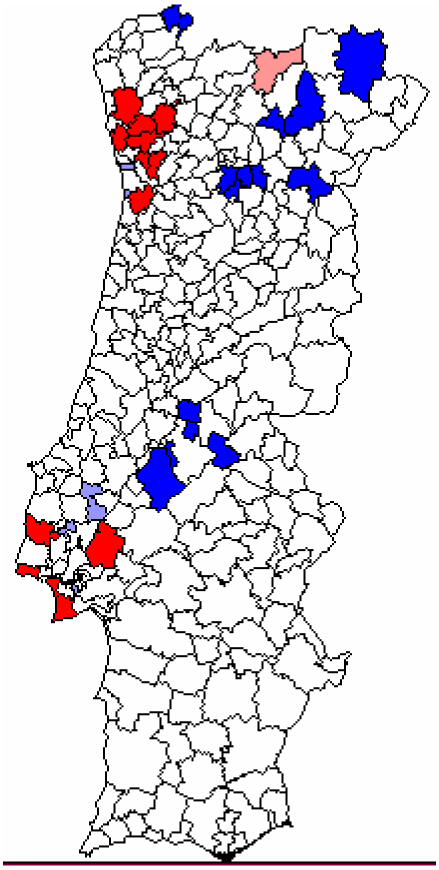


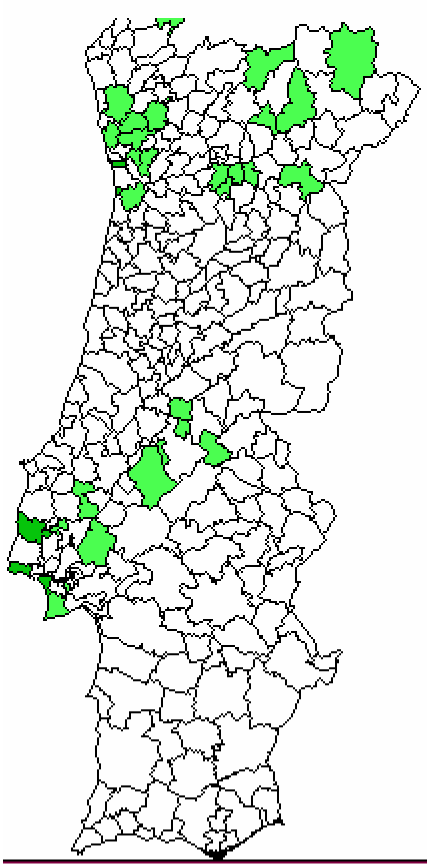
TXVAGOS02 (Vacancy rates)



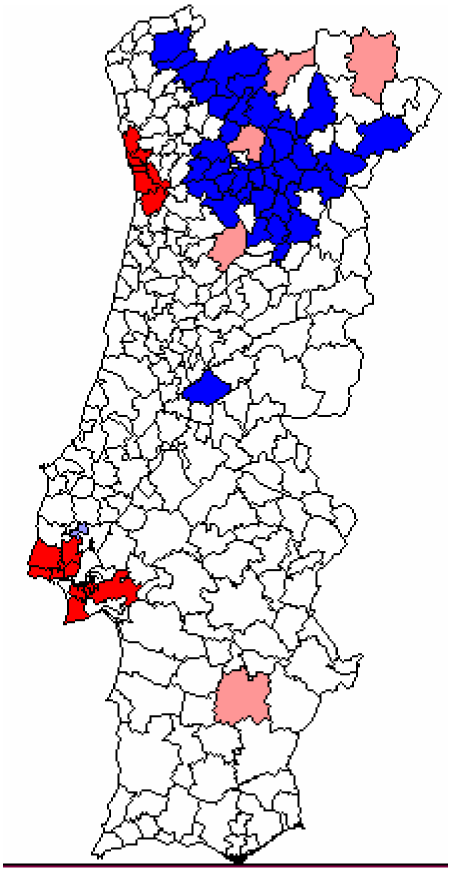


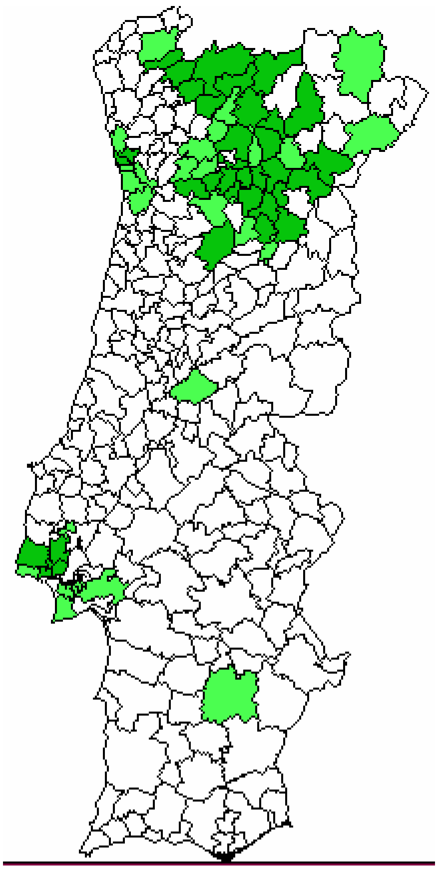
«Annex3»
Var_Procura (demand change)





«Annex4»
IPCC00 (Purchasing power index)





Regression results

Novas construções (95-01) = f (Var_procura)

REGRESSION

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

```
Data set              : habpais3
Dependent Variable    :     NCT95_01  Number of Observations:
275
Mean dependent var    :      1666.79  Number of Variables   :
2
S.D. dependent var    :      2647.99  Degrees of Freedom    :
273

R-squared             :     0.259905  F-statistic           :
95.8717
Adjusted R-squared    :     0.257194  Prob(F-statistic)
:1.34704e-019
Sum squared residual:1.42709e+009  Log likelihood         :
-2516.25
Sigma-square          :5.22745e+006  Akaike info criterion :
5036.5
S.E. of regression    :      2286.36  Schwarz criterion     :
5043.74
Sigma-square ML       :5.18943e+006
S.E of regression ML:        2278.03
-----------------------------------------------------------
------------
    Variable    Coefficient      Std.Error    t-Statistic
Probability
-----------------------------------------------------------
------------
    CONSTANT      1401.601       140.5078       9.975251
0.0000000
    VAR_PROC     0.7220088       0.073739       9.791411
0.0000000
-----------------------------------------------------------
------------
```

REGRESSION DIAGNOSTICS

```
MULTICOLLINEARITY CONDITION NUMBER   1.215555
TEST ON NORMALITY OF ERRORS
TEST                      DF          VALUE            PROB
Jarque-Bera                2           9995.289
0.0000000
```

DIAGNOSTICS FOR HETEROSKEDASTICITY

```
RANDOM COEFFICIENTS
TEST                      DF          VALUE            PROB
Breusch-Pagan test         1           2541.474
0.0000000
```

```
Koenker-Bassett test   1             168.24
0.0000000
SPECIFICATION ROBUST TEST
TEST                  DF           VALUE            PROB
White                  2           199.8673
0.0000000
========================= END OF REPORT
==============================
```

Var_Oferta = f(Poder de compra)

```
REGRESSION
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set              : habpais3
Dependent Variable    :  VAR_OFERTA  Number of Observations:
275
Mean dependent var    :     3546.19  Number of Variables   :
2
S.D. dependent var    :     5596.79  Degrees of Freedom    :
273

R-squared             :    0.576578  F-statistic           :
371.747
Adjusted R-squared    :    0.575027  Prob(F-statistic)     :
0
Sum squared residual:3.64741e+009  Log likelihood          :
-2645.28
Sigma-square         :1.33605e+007  Akaike info criterion :
5294.56
S.E. of regression    :      3655.2  Schwarz criterion     :
5301.79
Sigma-square ML      :1.32633e+007
S.E of regression ML:     3641.88
-----------------------------------------------------------
------------
    Variable    Coefficient      Std.Error    t-Statistic
Probability
-----------------------------------------------------------
------------
    CONSTANT     -6239.802       553.3475      -11.27646
0.0000000
      IPCC00       147.214       7.635289       19.28074
0.0000000
-----------------------------------------------------------
------------


REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   4.813156
```

TEST ON NORMALITY OF ERRORS
TEST DF VALUE PROB
Jarque-Bera 2 6264.294
0.0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST DF VALUE PROB
Breusch-Pagan test 1 169.1957
0.0000000
Koenker-Bassett test 1 13.85629
0.0001973
SPECIFICATION ROBUST TEST
TEST DF VALUE PROB
White 2 24.1497
0.0000057
========================= END OF REPORT
==============================

Novas C_01 = f (custo01)

REGRESSION
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set : **habpais3**
Dependent Variable : **CNFH01** Number of Observations:
275
Mean dependent var : 369.12 Number of Variables :
2
S.D. dependent var : 583.805 Degrees of Freedom :
273

R-squared : 0.255128 F-statistic :
93.5057
Adjusted R-squared : 0.252399 Prob(F-statistic)
:3.27166e-019
Sum squared residual:6.98152e+007 Log likelihood :
-2101.34
Sigma-square : 255733 Akaike info criterion :
4206.68
S.E. of regression : 505.701 Schwarz criterion :
4213.91
Sigma-square ML : 253874
S.E of regression ML: 503.859

Variable Coefficient Std.Error t-Statistic
Probability

CONSTANT -2481.967 296.4162 -8.37325
0.0000000

COST01 0.02968117 0.00306946 9.669835
0.0000000

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 19.3888
TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	4042.279	0.0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	1	31.35287	0.0000000
Koenker-Bassett test	1	3.251856	0.0713426

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	2	45.04712	0.0000000

========================= END OF REPORT ================================

Novas_C_01 = f (custo; poder de compra)

REGRESSION
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : **habpais3**
Dependent Variable : **CNFH01** Number of Observations: 275
Mean dependent var : 369.12 Number of Variables : 3
S.D. dependent var : 583.805 Degrees of Freedom : 272

R-squared : 0.352555 F-statistic : 74.0563
Adjusted R-squared : 0.347794 Prob(F-statistic) :2.10682e-026
Sum squared residual:6.06836e+007 Log likelihood : -2082.06
Sigma-square : 223102 Akaike info criterion : 4170.13
S.E. of regression : 472.336 Schwarz criterion : 4180.98
Sigma-square ML : 220668
S.E of regression ML: 469.753

Variable	**Coefficient**	**Std.Error**	**t-Statistic**	**Probability**
CONSTANT	-1749.918	299.5732	-5.841373	0.0000000
COST01	0.01670989	0.003511428	4.758715	0.0000032
IPCC00	7.731301	1.208454	6.397681	0.0000000

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 26.98342

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	4459.975	0.0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	2	177.6655	0.0000000
Koenker-Bassett test	2	17.35139	0.0001707

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	5	29.17557	0.0000214

========================= END OF REPORT ==============