*Erika Ghiraldo
*Raffaele Ciula
**Maurizio Festa

# Housing prices: an analysis of the dynamics of italian market development

**Keywords: house prices, co-movements, convergence, panel data, principal components**

**Abstract** Considering the time series of house prices at regional level in this article we present a study on the dynamics of house prices in Italy at both short and long term.
In the short run, we analyse real house price appreciation in order to investigate the existence of a common national impact, a city specific fixed effects and the persistence parameters with the aim to investigate co-movements among the regional time series.
For the long run, we propose a preliminary analysis on the existence of a "convergence" process among Italian regional house prices series.
From the empirical point of view we implement, for both periods, a principal components technique and a panel data model using a time series (from the first half year 2000 to second half year 2011) of a cross section (19 regions) released by Real Estate Market Observatory belonging to the Italian Revenue Agency.
At this level of analysis, the results show the existence of a common trend among the series of house prices although, for many regions, there are a significant local fixed effects. Finally, the long term convergence analysis of house prices appreciation is not fully showed by the statistical procedures.

## INTRODUCTION

Recent experience has shown that the evolution of housing prices has a strong impact on many macroeconomic aspects of the economy and is consequently of great interest to families, policy makers and all who are involved in the real estate market. The recent economic crisis, which originated in the real estate sector, has shown that a significant proportion of financial risk is related to complex financial instruments whose purpose is to protect against risk. The housing market is able to trigger systemic crises that generate repercussions in the real sector; it is therefore important to trace trends in order to have a better understanding of the economy perspective.

Knowing the housing pricing dynamic is of fundamental importance in order to adequately assess and avoid to over- or under-estimate the *direction* of the series. We must likewise remember that the trend in prices is influenced by sub-regional movements that direct it. Housing prices in fact shows these trends far more than any other differences in local goods, especially those related to the wide variety of regional features. The study of the evolution of the time series viewed as a whole or the deepening of their "co-movements" allows an understanding of the series' response to *shocks* occurring in the system.

*Official of Statistics and Studies Real Estate Market - Revenue Agency
**Head of Statistics and Studies Real Estate Market - Revenue Agency

Based on this premise, this article proposes a time series analysis of home prices in regions of Italy, for the period 2001-2011.
The main objective is to verify the existence of a common factor that by "driving" the regional time series of prices in a common direction, may tend to generate a type of convergence. More specifically, the analysis is distinguished in the temporal aspect by separating the short and long term. The short term analysis focuses on a study of the "co-movement" of the series in an attempt to determine the components of the change in housing prices and the relationships between the trends of entire series. We test the hypothesis of long run convergence of the series.
Essentially two techniques apply for both analyses: a panel model is estimated and the results are validated using the principal components technique. The regression applied to the data *panel* follows the analysis proposed by Gyourko & Voith in 1992, which uses a function that breaks down the change in housing prices into three components: the overall effect of time, a local parameter, and finally a persistence coefficient.
The structure of correlation of the time series is analysed by a principal component technique both in the short and the long run.
The rest of the paper is organised as follows. The second section describes the literature on the convergence of housing prices, and the third section presents the panel data model to estimate price dynamics. The results and issues of both models are discussed in the fourth and fifth paragraph.
The sixth paragraph discusses the hypothesis of convergence in the long term and finally, the last section is devoted to the description of the main results, problems and future lines of research.

## LITERATURE REWIEV

The concept of convergence is well known in literature that deals with economic growth models, beginning with neoclassical theories (Solow, 1956 and Romer, 1986) which aim to investigate various aspects of per capita income growth through an examination of their factors.
In the real estate market, housing price dynamics in the various subnational areas and their possible convergence have been a widely debated topic for long time. Basically two types of convergence can be found in literature. The first concept refers to time convergence or ripple effect (discussed above in Meen, 1999 and Holmes and Grimes, 2008) in which convergence is defined as the tendency of prices to react in a similar way to a *shock*. In this sense, according to the theory, a latent factor leads the relative housing prices to converge towards a steady-state. In more detail, the ripple effect spreads the shock of the regional prices throughout all the regions, thereby causing a long run co-movement. Usually this effect is attributed to various factors including the immigration of families, spatial arbitrage, and also the presence of similar spatial patterns in housing prices. The use of this definition of convergence does not assume that the prices between the various areas are the same, but that a variation extends in a very similar way to other areas. For example the article by Zhang (2010) analyses the possibility of a ripple effect in regional housing market prices in the United Kingdom using the fractional co-integration. This method exploits the fact that the variables (the price ratio for each region) are not stationary and have a common trend that drive them. The results show that, with some exceptions, regions have a long run trend that makes them converge to the national price.
The other notion of convergence could be called a cross-sectional convergence, and suggests that the areas in which low prices are initially observed tend to have higher growth rates than regions that initially show higher prices, so as to create a convergence in the long term. In the literature there are many articles that investigate the processes of convergence between the metropolitan areas of the same country or between different countries. This concept belongs to the notion of relative convergence or "*t-convergence*", introduced by Philips and Sul (2007), which suggests a decrease in

the dispersion of housing prices over time. This definition of convergence in housing prices between geographical areas and over time is taken up in the article by Kim and Rous (2012), which tests the hypothesis of convergence in a panel of 48 US states. The results show a lack of the “t-convergence” among all the states, and instead point out a strong convergence among subsets of states.
From an empirical point of view, in order to test for convergence the most used techniques is the analysis of time series co-integration. The fractional co-integration is used in the article by Barros et al. (2012) to test the relationship between housing prices among the states in the U.S.
Gupta and Miller (2012) analyse the hypothesis of convergence among metropolitan areas in Southern California by testing the co-integration between housing price rates. Granger causality test is used to demonstrate the statistical causality among the various metropolitan areas.
Seemingly unrelated regression estimations (S.U.R.E.) is used to test the hypothesis of housing market convergence in Britain against the prices of various regions (Holmes, 2007).
Panel data models are also widely used. One example is the article by Hiebert and Roma (2010), which sets out to analyse the convergence of housing prices among some European cities.
The paper by Gyourko and Voith (1992) shows a model which breaks down the growth of house prices into three components: a global aggregate, a local variation and a persistence of individual time series. The authors analyse the rates of changes in house prices in 56 metropolitan areas in the United States in the period 1971-1989 in order to test the influence of each component. More specifically, Gyourko and Voith highlight the existence of a significant common national component in long run prices in all metropolitan areas.
Holmes and Grimes (2008) jointly use co-integration methods and multivariate analysis such as the principal component approach throughout their study. The authors use the principal components to test the hypothesis of long run price convergence between the English regions on the ratio between regional and the national prices; they test the stability of the first component as well. The results suggest that prices tend to converge towards a long run equilibrium. If a shock on housing prices spread from any region would result in a wave (ripple effect) that would have the same effect on all regional prices.

## EMPIRICAL MODEL

The model proposed by Gyourko and Voith (1992) suggests, in theory, a link between the rate of change in housing prices, a national component common to all series, a specific factor for each geographical area and, finally, some parameters that capture the persistence of the individual series. Based on this model and on a later development (MacLean, 1994) we use the following model:

$$100 * \ln(P_{it}/P_{it-1}) = B_t T_t + A_i + C_i * 100 * \ln(P_{it-1}) + E_{it} \qquad [1]$$

in which:
- $P_{it}$ is the housing price variable. In particular, it is the deflated of price of region i at time t, computed as the average regional prices for regional capitals, which will be described lather in this article;
- $T_t$ is the time index for the semester for each time t.

Coefficients $B_t$, $A_i$ $C_i$ are estimated in the model. The coefficient $B_t$ is related to variable $T_t$ to indicate common trends for all series. This can therefore be interpreted in economic terms as the national variation. $A_i$ coefficients are fixed effects for each region. Finally $C_i$, are the coefficients of the lag price and can be viewed as persistence and they allow to show the price divergence or convergence over time.
To complete the model and in order to carry out further tests it is assumed that the error $E_{it}$ is

independent and identically distributed (IID) with zero mean and constant variance equal to $\sigma^2$ or $E_{it} \sim IID(0, \sigma^2)$.
The goal of each of these three components is therefore to explain a specific effect of price appreciation. The national component represented by the $B_t$ coefficient to detect a common to all series that somehow "guides" them all together in the same direction. If $B_t$ are significant we will observe a co-movements across the series. The vector $B_t$ would therefore capture the differences between the price changes over time.
However it is still obvious that the series may deviate from the national trend considering the specificities that characterise each region. In the model formulated, these effects are taken into account by introducing the $A_i$ component to capture only the differences between the regions, while ignoring the temporal aspect.
The $C_i$ coefficients concern the fact that the time series make show an autocorrelation.
In order to verify the statistical significance of the three components we provides tests of several hypothesis using the usual F test.
The statistical significance of all $B_t$ is tested to set the hypothesis $H_0$: $B_t = 0$ for each period t.
The rejection of the hypothesis would imply that there is a time varying common component for all series. On the contrary, the impossibility of rejecting $H_0$ would be an indication of the absence of a global component which can be interpreted with the fact that housing prices variations is attributed only to local factors.
The statistical significance of the local component needs to be tested, verifying that all the coefficients $A_i$ are jointly not zero. If $A_i = 0$ for all i, then all the regions do not show deviations from national trends, or local specificities are not as influential. If the null hypothesis is rejected, however, the local characteristics are significant and may cause deviations from the national trend.
The significance of differences between the regional home prices appreciation is verified testing the hypothesis $H_0$: $A_i = A_j$ for all $i \neq j$. If we cannot reject the null hypothesis then the local effects are not significant, or in other words, there is an equal appreciation across the regions.
The joint test on the hypothesis $H_0$: $C_i = 0$ allows us to understand if the persistence in the appreciation differ or not across the regions. If the coefficients $C_i$ are jointly not significant than the series follow a common trend and the random shocks have effect on
In order to verify the joint significance of $C_i$ coefficients we set the hypothesis system $H_0$: $C_i = 0$ for all regions i. If the $C_i$ are not jointly significant, then the series do not show persistence and the shocks influence the series in a single period without repercussions in subsequent periods. On the other hand if the coefficients are significantly different from zero, then the series tend to "remember" the past and the shocks are not absorbed in a single period.
Finally, we test the hypothesis of equality of the coefficients $C_i$, $H_0$: $C_i = C_j$ for all $i \pm j$ with the aim to verify the equality of appreciation among the regions.
In summary, then, the following tests are applied:

- $H_0$: $B_t = 0$ per ogni t, $H_1$: otherwise;
- $H_0$: $A_i = 0$ per ogni i, $H_1$: otherwise;
- $H_0$: $A_i = A_j$ per ogni $i \neq j$, $H_1$: otherwise;
- $H_0$: $C_i = 0$ per ogni i, $H_1$: otherwise;
- $H_0$: $C_i = C_j$ per ogni $i \pm j$, $H_1$: otherwise;

Prices in every region are expected to differ each other because of many advantages to live in a place in comparison with others.
If we assume that the initial regional growth rate levels off and regions with economic disadvantage increase the growth rate than the differences between those regions should decrease.

Since the greater benefits are generally reflected in higher price levels, we should observe less and less price appreciations for the better off regions and these could case the catch up of the disadvantages regions. If this occurs then the $C_i$ vector coefficients should have a negative sign. In this case, the price series for each region would tend to return to its mean price level.
On the contrary, if the $C_i$ coefficients are greater than zero, we will see prices increasing over the time.

**The panel model estimate**

As seen above, the model's specification suggested by the theoretical relationship formulated by Gyourgo and Voith, requires two-dimensional observations, or observations that vary both in time and per unit. To achieve this goal, beginning with the data made available by the Inland Revenue's Real Estate Market Observatory (OMI), semi-annual average housing prices have been recorded in time series, broken down into 19 Italian regions.
The OMI gives semi-annual purchase and sales prices, offers or appraisals for the OMI zones of each municipality[1]. Based on these values, the average six months is calculated for capital only, as the average quote for all types of residential real estate. The averages thus obtained are used to generate a weighted average of prices for each region, using as weights the relative housing stock[2]. This operation is conducted for each semester from the first semester of 2001 through the end of 2011 to form the panel of N=19 regional time series consisting of the average prices observed for 22 semesters (I semester 2001 - II semester 2011)[3]. Each series is deflated using the regional consumer price index, including tobacco products, published by ISTAT. The dynamic of the regional series is shown in Figure 1, the latter points out that prices had an increasing trend up to 2008 when they began to fall. The only exception is in Friuli-Venezia Giulia, which shows a downward trend for most of the period.
The dependent variable, as predicted by the model, is constructed by taking the natural logarithm of the ratio between semester price t and the price for the preceding semester t-1, thereby dropping the initial semester observation (I 2001).
In order to estimate the model in the equation (1) we build a series of dummy variables:

- to estimate the $B_t$ coefficients, we set T-1 dummy variables using the number of semesters T,[4] indicated in the tables of results with "d_I_2002" to "d_II_2011";
- to estimate the $A_i$ coefficients we set N dummy variables, indicated with "d_Abr" to "d_Ven";
- $C_i$ are the coefficients of the lag price of each region and are indicated with "Pt-1_Abr" to "Pt-1_Ven".

We estimate the model using an OLS method on the pooling of cross-sectional data and time series.
The estimates reported in the Appendix (Table 1) show that the fixed effects are all statistically significant with the exception of those relating to Abruzzo, Sardegna and Friuli-Venezia Giulia. The same coefficients are very high in the Valle D'Aosta, Lombardia, Umbria and Molise. The $C_i$ coefficients show similar results in terms of significance. With regards to sign, the $C_i$ conforms to expectations with all negative. The time varying component is statistically significant for almost all semesters with always a positive sign, except for the first half of 2002. These results suggest a positive impact on the Italian real house appreciation.

1 For a discussion on the formation of homogeneous areas and the Observatory database you can refer to the "Operating Manual of the Bank Data Centre Real Estate Market, 2008".

2 The number of homes in each municipality is taken from the archives of the Italian Cadastre.

3 Belong to the panel all regions with the exception of Trentino Alto Adige for which no data are available related to the *stock*.

4 As for the dummy $T_t$, the first half was lost to build the dependent variable.

In the bar graph (Figure 2) we show the coefficients $A_i$ and $C_i$. We can see that when $A_i$ coefficients rise, $C_i$ also rise but with opposite sign. This suggests that bigger fixed effects imply higher persistence.
On the basis of this analysis we could assume that if a region has higher fixed effect than it has a greater persistence and we can observe a mean reverting process and deviation from national trend.

**Test Results**

As we have seen, in order to obtain some initial result we perform a series of tests on the estimated coefficients.
We perform a F test on $B_t$ coefficients comparing the full model with a restricted model that imposes the nullity of $B_t$. We obtain a p-value close to zero and we can conclude that there is a significant time varying component among the series. There is a common trend that draw the series in the same direction triggering the "wave" effect. From this result we can deduce that the differences across the areas could remain unaltered over time but we cannot establish, at least at this stage of the analysis, if there are series affected by a major or minor effect or there are series that are not affected in any way.
Performing the joint test on Ai coefficients we observe that the local differences among the regions are significant and then in order to explain the house appreciation rates we cannot ignore them. Less formally, a joint reading of the results of the two tests discussed here shows that if you look at a certain national price increase it is difficult to see very different rates across regions. However, the effect of the national trend summarised by $B_t$ is weakened by the importance of the specific effects that can cause differences in $A_i$ due to factors that summarise the characteristics of each region. Regional peculiarities may be more or less pronounced and manifested in different forms; and it can reasonably be argued that regions with higher fixed effects and therefore with higher $C_i$ show a greater tendency to drift from the national rate.
For a more detailed study of the similarities and differences between the fixed effects of each region pairwise tests are run between the $A_i$ coefficients.[5] The maps of Figure 3 included in the appendix show the results of all comparisons. Most of regions do not show significant differences with a large number of other regions; however, a more thorough study points out that each region shows significant differences with at least five regions. More attention should be given to the fact that there is a group of regions formed of Lombardia, Valle d'Aosta, Umbria and Molise, which has specific effects at the same time not significantly different from each other, but with marked differences from most of the other regions. This is true for those regions where the estimated A and C coefficients were found to be higher. It would appear that these regions have specific, more influential effects, which confirms what has already emerged from the fixed effects analysis and persistence graphs.
The "time" dimension represented by each regional time series cannot be ignored in an exercise of this type. The joint test on persistence underlines that, globally considered, this series effect is significantly non-zero, that is each one has memory of its on past value. In addition, the fixed effects parameters are also more evident for some regions and less for others. Even in this case, the pairwise tests point out similar differences that are not negligible. Even this test shows the group of regions formed of Lombardia, Valle d'Aosta, Umbria and Molise reinforcing the idea that regions with strong fixed effects have price series more dependent on their past. Recalling that the coefficient bears a negative sign, this may mean that external shocks have a less intense effect on the price series and it assures that the adjustment of the variations around the average of the

5 For this system of hypotheses the t test statistic is used.

same series is faster. In other words, these series exhibit different trends from the general trend and are more related to their pattern.

## PRINCIPAL COMPONENTS ESTIMATE

If we consider our data as a time series rather than as a panel, literature, especially literature regarding a financial time series analyses, suggests for a more detailed analysis of the covariance structure the use of the principal components analysis (PCA), which is typical tool of multidimensional analyses (see for example Curto *et al*., 2006, or Feeney & Hester, 1964).

The goal of the principal component analysis[6] is to reduce a large number of variables that represent the phenomenon analysed to a smaller number of latent variables obtained as a linear combination of the originating variables. The transformation takes place in such a way that the first component explains the greatest part of the variance. The principal components are mutually uncorrelated in the construction and sorted in descending order with respect to the variance. The total variance (sum of variances) can also be shown as preserved in the transition from the observed variables to the principal components. The principal components do not correspond to any directly observable characteristics but need to be interpreted.

Before proceeding with the principal components analysis, it is necessary to verify that the method is feasible and appropriate for the available data examining the matrix of simple correlations, using the Kaiser-Meyer-Olkin (KMO) test and the Bartlett sphericity test.[7] All three tests confirm the existence of a correlation structure that would allow application of the PCA. In particular:

- the correlation matrix show high and significant correlation coefficients for all series with Friuli-Venezia Giulia and Valle d'Aosta the only exceptions. Friuli Venezia Giulia is the only region to show negative correlation coefficients with all regions that are especially notable when compared with those of Liguria, Puglia and Sardegna. As mentioned, Friuli exhibits a decreasing trend in prices in most of the period, with a very high difference in comparison with the growth trend of the three regions mentioned. It also shows no significant correlation with the other four regions. The Valle d'Aosta region on the other hand exibits the lowest and least significant correlations of all those remaining, with no particular trend. Generally, it can be said that the regions show extensive interconnections with others, suggesting that it may be sufficient in one or two components to capture the majority of the correlations highlighted.
- a KMO index, 0.74, shows good correlation and then supports the convenience of a principal components' analysis;
- the Bartlett sphericity test shows a p-value close to zero, leading to the rejection of the null hypothesis.

After completing the preliminary analysis we proceed to extract the principal components. To choose the number of components for consideration we make a visual inspection of the eigenvalues (Figure 4). It is clear that the highest eigenvalues $\lambda$ are associated with the first two components ($\lambda_1 = 15.1$ and $\lambda_2 = 2.43$), explaining 79.4% and 12.7% respectively of variability. It was therefore decided that the first two components would be used for the next phase of interpretation.

An analysis of the weights associated to each region in the first component, PC1, points out a general uniformity. In fact, most of the regions has a value around 0.25. Almost all regions contribute similarly to the explanation of most of the variability in the data.

6 For a discussion on the technique of principal components one can refer for example to work Jolliffe (2002).

7 The index varies between 0 and 1 and a value close to zero indicates that the principal component analysis would be inadequate, while a value close to 1 indicates the presence of a correlation structure which can be incorporated in an appropriate manner by the main components.

Consequently the first principal component PC1 suggests the presence of a global short term national effect that influences all regions. This result confirms what has already emerged from the analysis of the $B_t$ coefficients regression.
However, two exceptions stand out. The first is Friuli Venezia Giulia which, as already mentioned, is the only region with decreasing prices for most of the period, as well as the only region which is weighted negatively. The same PC1 also suggests the presence of a group of regions from Valle d'Aosta, Lombardia and Molise, with weights lower than in other regions. It is also interesting that these same regions have fixed effects and a high persistence estimate. It therefore seems reasonable to assume that these regions suffer less from the national situation as they are characterised by significant fixed effects. Such effects make the series more dependent on itself and therefore less affected by the national trend.
Although the presence of local fixed effect the PC1, that explains almost 80% of variance, confirm the importance of a national global component.
It is also interesting to analyse the weights of principal component, PC2. From this analysis we can see a group of regions with high weights as Lombardia, Molise and Friuli. We rimind that these regions have high values of fixed effects and high persistence too. Then, the second component can represent the specific fixed effects for the regions. However, conclusions cannot be drawn about the existence of an effect common to the two regions. In other words, the results indicate that some local effect may set apart individual regions but this impact is not necessarily common. In fact it is not possible to detect the presence of a common factor that only a formal analysis of co-integration could show. The limited number of observations available for each series does not in fact make this technique possible.
One useful way to interpret the results of the principal component analysis is to display them in a biplot[8] (Figure 5). The biplot is a graphic representation that allows a simultaneous display of observations and variables, allowing us to explore the correlation structure between the variables, represented in this case by the regions. This graph shows the contribution of each variable to the explanation of a principal component. In our case, the graph suggests the results that emerged from the analysis of the weights of the two components. The graph underline the correlations among some regional groups such as Calabria, Liguria, Puglia, Basilicata or even between Piemonte and Emilia Romagna, Campania, Toscana and Marche. These regions in fact show similar weights for both components. It also clearly shows the difference between Friuli Venezia Giulia and the other regions. The results of the first two principal components confirm the findings about the panel data estimation on $A_i$ and $C_{i.}$
The graph on the right of the same figure shows, instead, the loadings of the first two components per time period. By dividing the graph into quadrants, we can distinguish different trends over the period. After an initial period (until approximately 2002) prices begin an increasing trend characterised by high rates, which decreases to finally turn negative in 2008.

**An attempt to interpret the first principal component**

The joint analysis of the estimated $B_t$ coefficients in the regression model and the first principal component suggests the existence of a global trend that, although often diluted by the effects of the individual regions, is significant, and definitely not negligible.
A study of the factors that affect price series performance requires the implementation of a multivariate time series model that allows a structural and an economic policy analysis. Such a

8 Because there are several ways to construct a biplot graph, it is necessary to specify that we used the command biplot available in the software R. This command follows the theoretical lines of the original article that introduced the method (Gabriel, 1971).

study is beyond the scope of this article which is designed solely to provide an initial reflection on the interpretation of the first component, *i.e.* the existence of an unobserved factor that leads to a co-movement of the regional price series. Current studies in literature are useful in this respect. Disposable income is described as one of the most important fundamentals on the *demand side* of the housing market. More specifically it highlights a strong connection between an increase in personal income and an increase in housing prices caused by a rise in the demand for housing (Capozza *et al.* (2004), Kim & Rous (2012), Hwang & Quigley (2006)), and in a more recent work from the Bank of Italy (Nobili & Zollino 2012) on the Italian market. Given this result it seems reasonable to associate the first component with the disposable income. However the literature also highlights important links with other macroeconomic variables such as interest rates, demographic trends and employment rates. A recent article (Kholodilin, 2012) summarises the main results about the determinants of housing prices. The author gives an overview of the categories of indicators used in the models, including income, interest rates, demographic variables, the credit market, labour, land supply, housing supply and even other variables not related to a particular category. Income is defined as real GDP *per-capita* income or employment income, while interest rates in most of the models is regarded as the real interest rate; demographic variables considered obviously include, above all, the population, but also variables such as the marriage rate, the number of singles or divorce rate. Price rates are understood as referring to inflation or expected inflation. Among the indicators not attributable to a specific class we find the widely used construction prices, the corresponding real estate index or even the housing stock value.

More generally, it is recognized that reversals or changes in the economic cycle affect different areas in a similar way, even in the presence of structural differences between the regions (Oikarinen, 2007).

## LONG RUN CONVERGENCE

Analogous with the economic theory of rising incomes and their convergence proposed by Barro and Sala-i-Martin (Barro & Sala-i-Martin, 1992) it may be reasonable to expect, in absolute terms, that regions with higher prices in a certain period show lower rates of increase (rise at a slower rate) in order to reduce the differences. In our case, the correlation coefficient between the global prices variation for the period as a whole (2001-2011) and their level during the initial period (Semester 1 – 2001) shows a negative sign, close to zero. However, if we remove Lazio and Liguria that, in addition to having high prices show very high growth rates, the correlation coefficient is -0.27. Less formally, if we refer to the relative concept of convergence we should observe a decrease in the time of dispersion of the values observed. By measuring variability through the standard deviation for each year, there is a decline, although not pronounced, beginning in 2006, however the downward trend becomes more notable if we exclude Lazio and Liguria.

These first analyses show then that the hypothesis of a long term convergence is rather weak, but in the next step we apply two more formal analysis.

Based on the theoretical model suggested by Gyourko & Voith, 1992 we estimate this relation:

$$100 * \ln(P_{it}/P_{it-1}) = B_t T_t + A_i + C * 100\ln(P_{it-1}) + E_{it} \qquad [2]$$

As in formula (1) $P_{it}$ is the variable price for each region i at time t and $T_t$ is a time variable for each semester. The coefficients $A_i$ and $B_t$ are the same of formula (1), while C now represents a single mean divergence or convergence coefficient for all of the regions. If C becomes negative we could say that regional prices in the long run will tend to converge. The variable $E_{it}$ is the error that we assume $E_{it} \sim IID(0, \sigma^2)$.

Applying the model to the same data described above we obtain the estimated coefficients reported

in Table 3. Coefficient C is significant and equal to -0.104, indicating that a lower rate of increase of 0.10% is associated with regions that on average showed real prices over 1%. Thus, the regions with higher rates should under equal conditions show a lower rate of increase over all regions with high price levels. Still, since the value of the coefficient is rather low, such a situation can occur in a very long period of time and possibly make a transition period necessary.

Unlike the short run, in this model the parameters associated to $T_t$ variable are less significant. In the long term the varying component is less important. This result seems to strengthen the impact of fixed effects on the behaviour of the series. In terms of convergence, there may be a double interpretation. In terms of convergence, there may be a double interpretation. On the one hand we find that the absence of a global component and the diversity of local effects might lead to somewhat similar prices with very different variation rates while, on the other hand, the same differences in fixed effects may cause a slowdown or the total exclusion of the phenomenon of convergence over the long term as well.

With the aim to examine in depth the long run analysis, on the basis of the article by Holmes and Grimes (2005), we present another study on the price series. In order to verify the long run convergence Holmes and Grimes suggest to test the stationary, with a formal unit root test, of the first principal component on regional house price differentials. This type of convergence refers to the concept of the ripple effect: if a shock impacts on a region, the long run effect of house price prices may be arise, like a wave, across the entire country.

We calculate the ratio

In summary the analysis is performed in two phases:

- in the first phase we calculate for each region in each period the ratio between the average price and the average national (which then becomes the benchmark price). On these series we apply a principal components analysis and we extract the first component indicated as LPC (long period principal component);
- in the second phase on first component LPC we apply a series of stationary tests. If we accept the hypothesis of stability we have evidence of a convergence process across regions. In fact, if the first component is stable then so will all the other components, indicating a strong convergence between the series.

On our series we extract the first principal component[9] which represents about 55% of variability and then we verify a presence of unit root applying three tests, known in the literature, described in the appendix (Table 4). Two of the three tests exclude the stationary of the series, and only the last one accept the hypothesis. Although the data estimated for the panel model have revealed that price trends tend to converge, albeit at a very low rate, the results of the tests on the principal component do not fully confirm a convergence of housing prices in the long term. Both results suggest uncertainty about the trend of prices or at least show a slow transition process. We must say that the series are too short to allow any firm conclusions.

It can be assumed that this result is due to the incidence of various factors such as the presence of local effects that significantly affect housing prices or even the persistence of structural gaps. As suggested by the growth theory, the regions with an initial economic gap, for instance in the labour market, should decrease it increasing the economic attractiveness causing an increasing in house prices. The catch-up strategies implemented by business lead to greater labour demand therefore we should observe flows of employees and families from areas where housing prices are higher to regions with lower prices but with future chances of development. Furthermore, the income growth

9 It should be noted that the number of deviations from the national average regional prices have conducted preliminary tests to assess the adequacy of the method of principal components discussed above.

in the areas should lead to a higher rate increase in house prices compared to the regions with initial economic advantages.
If it does not happen, the adjustment mechanism will not work for the long run equilibrium.
The results confirm that the long term convergence analysis and the techniques used to test it are still debated and controversial as evidenced in Holmes (2007).

## CONCLUSIONS

In this paper we presented a discussion on two topics. On the one hand in the short run we verified the existence of co-movements between regional series of housing prices in Italy. In the other hand, in the long term, we investigated the hypothesis of a process of convergence between the price series. Both analysis are conducted using semi-annual data from 2001 to 2011 from Observatory of the Real Estate Observatory Market of Revenue Agency.
In the short run, in order to test a co-movement in the series we used a *data panel* regression technique and a principal component approach. The results show the existence of a significant latent factor that drives the series. The series therefore shows a tendency to respond in a similar way to shocks from one or more factors. As no specific analysis have been under taken on the price determinants, reference was made to those already present in the literature and it was assumed that the global trend may be associated to variables as income, demographics trends, interest rates and, in more general terms, upswings and recessions in the economy. For this point we intend, in the future, to conduct an *ad hoc* study that implements multivariate time series analysis techniques with the inclusion of potentially relevant macroeconomic variables.
The short term analysis also showed that in addition to the global factor each region has specific fixed effects that tend to deflect the series to a greater or lesser degree based on the performance of the global trend. For this second aspect we think to conduct a further study through an analysis that examines the peculiarities of a certain area.
We also investigated how the specific effects are tied to persistence, in fact if the former is higher, then the persistence of the series is also higher as well as the link with its past. Regions with marked specific effects and therefore with a strong persistence tend to dampen the overall effect and to return to their own path rather than following the common trend.
About the long run analysis we used a panel regression and a principal component technique too. The results of both techniques are in part discordant. The estimated C parameter has a negative sign and a value of 0.10 which indicates a low convergence.
The stationary tests on LPC allow to accept the hypothesis of convergence in only one case over three. The long term convergence appears to be weak and in any case to require a rather long adjustment process.
The main problem of the study is shortness of the time series that do not permit to carry out formal cointegration tests in order to verify a global convergence or a convergence across a regional subsets so called club convergence.
Future research should employ a formal analysis co-integration or even implement multivariate models with exogenous variables to better explain the series and ad hoc analysis to understand specific local effects.
On the other hand, although we used short time series and we did not implement more formal econometric analysis, we presented a preliminary study with the aim to provide information on the price dynamics by studying jointly sub-national areas levels.

**References**


Agenzia del Territorio, (2008), Manuale della banca dati dell'Osservatorio del mercato immobiliare, (pag. 59). Available at: www.agenziaterritorio.it.

Barro, Robert J & Sala-i-Martin, Xavier, (1992), "Convergence," Journal of Political Economy, University of Chicago Press, vol. 100(2), pages 223-51, April.

Barros, C. P., Gil-Alana, L. A. & Payne, J. E., (2012), "Co-movements among U.S. state housing prices: Evidence from fractional co-integration", Economic Modelling, Volume 29, Issue 3, May 2012, Pages 936-942, ISSN 0264-9993.

Bartlett MS., (1950), "Tests of significance in factor analysis", British Journal of Psychology, 3(Part II):77-85.

Capozza, D. R., Hendershott, P. H., & Mack, C., (2004), "An anatomy of price dynamics in illiquid markets: analysis and evidence from local housing markets". Real Estate Economics, 32(1), 1-32

Curto, J. D., Pinto J. C. & Fernandes, J. E., (2006), "World Equity Markets: A New Approach for Segmentation (in English)," Czech Journal of Economics and Finance (Finance a uver), Charles University Prague, Faculty of Social Sciences, vol. 56(7-8), pages 344-360, July

Feeney G. J. & Hester, D. D., (1964), "Stock Market Indices: A Principal Components Analysis", Cowles Foundation Discussion Papers 175, Cowles Foundation for Research in Economics, Yale University

Gabriel, K. R., (1971), "The biplot graphical display of matrices with application to principal component analysis", Biometrika 58, 453–467

Gupta, R. & Miller S., (2012), "The Time-Series Properties of House Prices: A Case Study of the Southern California Market", The Journal of Real Estate Finance and Economics, Springer, vol. 44(3), pages 339-361, April.

Gyourko, J., & Voith, R., (1992), "Local market and national components in house price appreciation", Journal of Urban Economics, 32, 52-69.

Hiebert, P. & Roma, M., (2010), "Relative house price dynamics across euro area and US cities: convergence or divergence?", Working Paper Series 1206, European Central Bank

Holmes, M. J., (2007), "How Convergent are Regional House Prices in the United Kingdom? Some New Evidence from Panel Data Unit Root Testing", Journal of Economic and Social Research, 9: 1-17

Holmes, M., & Grimes, A., (2008), "Is there long-run convergence among regional house prices in the UK?", Urban Studies, 45(8), 1531–1544.

Hwang, M. & Quigley, J. M., (2006), "Economic Fundamentals In Local Housing Markets: Evidence From U.S. Metropolitan Regions", Journal of Regional Science, 46: 425–453.

Jolliffe I.T., (2002), "Principal Component Analysis", Series: Springer Series in Statistics, 2nd ed., Springer, NY, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4.

Kaiser, H.F., (1974), "An Index of Factor Simplicity." Psychometrika Vol. 39, No. 1

Kim, Y. S. & Rous, J. J., (2012), "House price convergence: Evidence from US state and metropolitan area panels", Journal of Housing Economics, Elsevier, vol. 21(2), pages 169-186.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin, (1992), 'Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?', Journal of Econometrics 54: 159–178.

Maclean D., (1994), "The Role of House Prices in Regional Inflation Disparities", Technical Report 1994-67 Bank of Canada

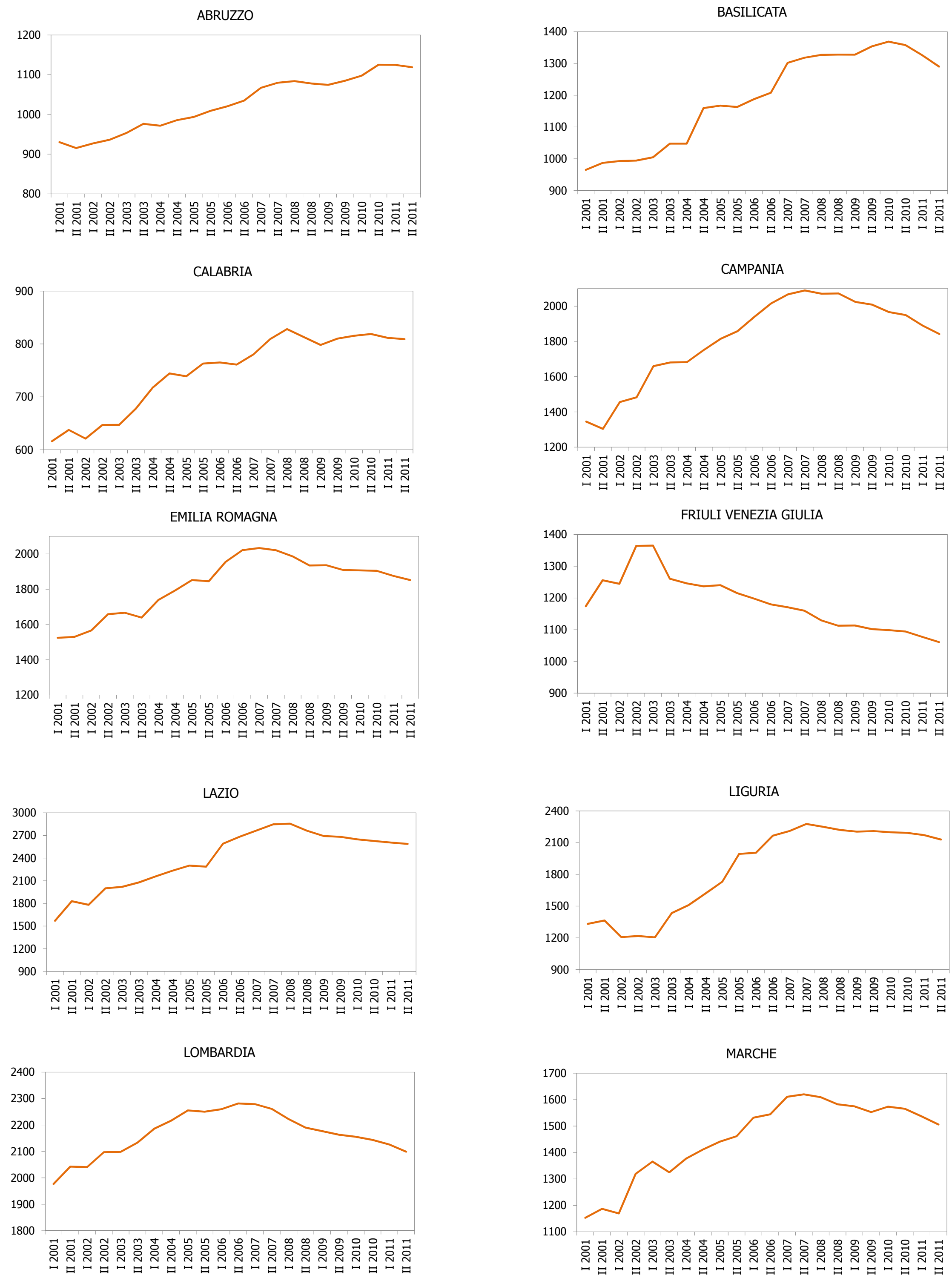
Nobili, A. & Zollino, F., (2012), "A Structural Model for the Housing and Credit Markets in Italy" (October 26, 2012). Bank of Italy Temi di Discussione (Working Paper) No. 887.

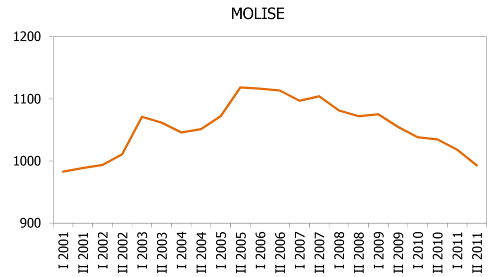
Romer, P. M., (1986), "Increasing Returns and Long-run Growth," Journal of Political Economy, University of Chicago Press, vol. 94(5), pages 1002-37, October.

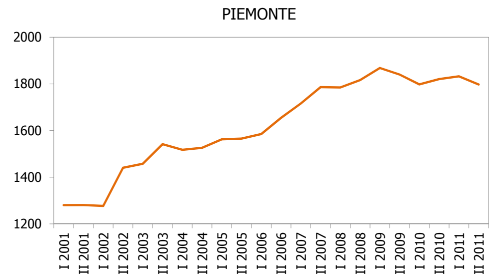
Solow, R. M., (1956), "A Contribution to the Theory of Economic Growth". Quarterly Journal of Economics (The MIT Press) 70 (1): 65–94.

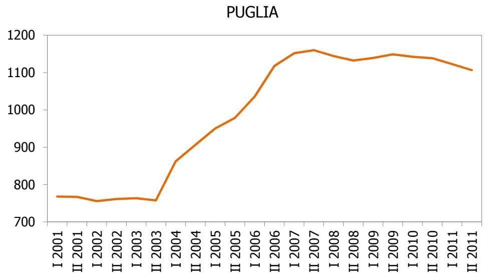
# APPENDIX

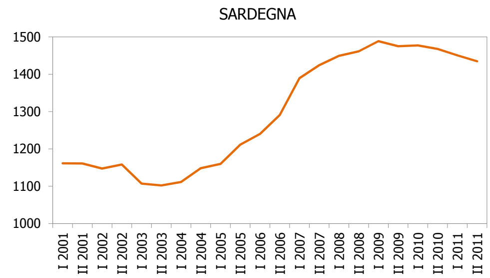
**Figure 1** Deflated regional prices time series

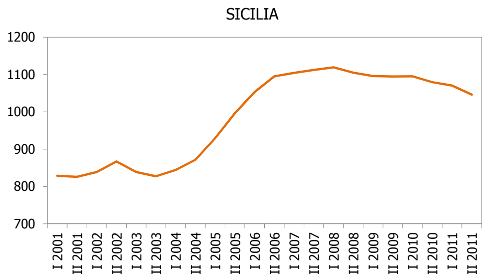


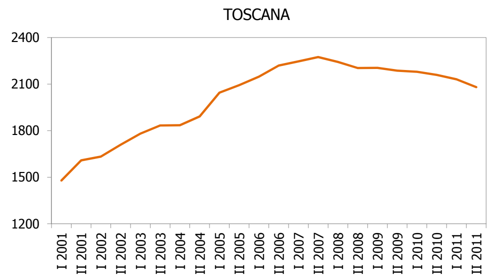

MOLISE
1200
1100
1000
900
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011

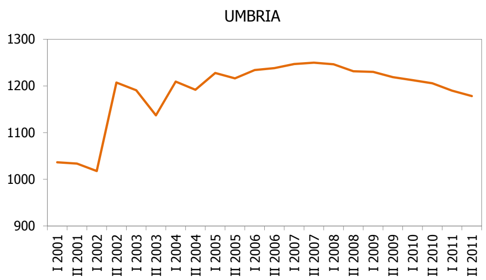

PIEMONTE
2000
1800
1600
1400
1200
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011

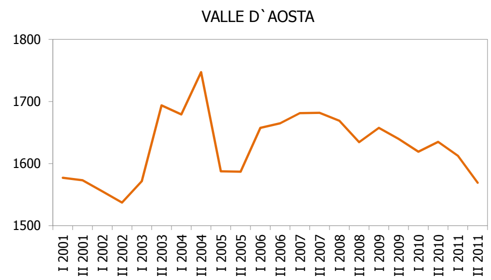

PUGLIA
1200
1100
1000
900
800
700
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011


SARDEGNA
1500
1400
1300
1200
1100
1000
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011


SICILIA
1200
1100
1000
900
800
700
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011


TOSCANA
2400
2100
1800
1500
1200
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011


UMBRIA
1300
1200
1100
1000
900
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011


VALLE D`AOSTA
1800
1700
1600
1500
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011

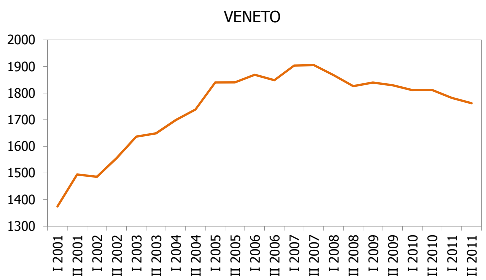

VENETO
2000
1900
1800
1700
1600
1500
1400
1300
I 2001 II 2001 I 2002 II 2002 I 2003 II 2003 I 2004 II 2004 I 2005 II 2005 I 2006 II 2006 I 2007 II 2007 I 2008 II 2008 I 2009 II 2009 I 2010 II 2010 I 2011 II 2011

**Table 1** Coefficient values of panel data estimation in the short run

| Coeff. | Variables | Estimate | Std Error | t Stat | P-value |
|---|---|---|---|---|---|
| | d_Abr | 55.68 | 68.862 | 0.809 | 0.419 |
| | d_Bas | 84.92 | 37.233 | 2.281 | 0.023 * |
| | d_Cal | 106.94 | 43.532 | 2.457 | 0.015 * |
| | d_Cam | 156.92 | 34.309 | 4.574 | 0.000 *** |
| | d_Emi | 179.47 | 53.799 | 3.336 | 0.001 *** |
| | d_Fri | 25.77 | 67.533 | 0.382 | 0.703 |
| | d_Laz | 173.18 | 29.999 | 5.773 | 0.000 *** |
| | d_Lig | 66.77 | 20.596 | 3.242 | 0.001 ** |
| | d_Lom | 416.09 | 126.638 | 3.286 | 0.001 ** |
| $A_i$ | d_Mar | 168.55 | 44.841 | 3.759 | 0.000 *** |
| | d_Mol | 353.61 | 113.385 | 3.119 | 0.002 ** |
| | d_Pie | 104.78 | 39.354 | 2.662 | 0.008 ** |
| | d_Pug | 63.41 | 25.906 | 2.448 | 0.015 * |
| | d_Sar | 14.46 | 39.861 | 0.363 | 0.717 |
| | d_Sic | 75.85 | 36.285 | 2.090 | 0.037 * |
| | d_Tos | 174.99 | 39.176 | 4.467 | 0.000 *** |
| | d_Umb | 356.69 | 73.621 | 4.845 | 0.000 *** |
| | d_Val | 551.90 | 143.173 | 3.855 | 0.000 *** |
| | d_Ven | 247.44 | 53.977 | 4.584 | 0.000 *** |
| | Pt-1_Abr | -0.08 | 0.100 | -0.828 | 0.408 |
| | Pt-1_Bas | -0.12 | 0.053 | -2.294 | 0.022 * |
| | Pt-1_Cal | -0.16 | 0.066 | -2.468 | 0.014 * |
| | Pt-1_Cam | -0.21 | 0.046 | -4.566 | 0.000 *** |
| | Pt-1_Emi | -0.24 | 0.072 | -3.347 | 0.001 *** |
| | Pt-1_Fri | -0.04 | 0.095 | -0.426 | 0.671 |
| | Pt-1_Laz | -0.22 | 0.039 | -5.723 | 0.000 *** |
| | Pt-1_Lig | -0.09 | 0.028 | -3.216 | 0.001 ** |
| | Pt-1_Lom | -0.54 | 0.165 | -3.296 | 0.001 ** |
| $C_i$ | Pt-1_Mar | -0.23 | 0.062 | -3.762 | 0.000 *** |
| | Pt-1_Mol | -0.51 | 0.163 | -3.134 | 0.002 ** |
| | Pt-1_Pie | -0.14 | 0.054 | -2.666 | 0.008 ** |
| | Pt-1_Pug | -0.09 | 0.038 | -2.452 | 0.015 * |
| | Pt-1_Sar | -0.02 | 0.056 | -0.397 | 0.692 |
| | Pt-1_Sic | -0.11 | 0.053 | -2.113 | 0.035 * |
| | Pt-1_Tos | -0.23 | 0.052 | -4.456 | 0.000 *** |
| | Pt-1_Umb | -0.51 | 0.104 | -4.853 | 0.000 *** |
| | Pt-1_Val | -0.75 | 0.194 | -3.868 | 0.000 *** |
| | Pt-1_Ven | -0.33 | 0.073 | -4.584 | 0.000 *** |
| | d_I_2002 | -2.34 | 0.875 | -2.673 | 0.008 ** |
| | d_II_2002 | 2.91 | 0.875 | 3.320 | 0.001 *** |
| | d_I_2003 | 0.69 | 0.900 | 0.765 | 0.445 |
| | d_II_2003 | 1.37 | 0.916 | 1.492 | 0.137 |
| | d_I_2004 | 2.61 | 0.933 | 2.798 | 0.005 ** |
| | d_II_2004 | 3.64 | 0.960 | 3.792 | 0.000 *** |
| | d_I_2005 | 3.62 | 1.000 | 3.621 | 0.000 *** |
| | d_II_2005 | 3.58 | 1.026 | 3.490 | 0.001 *** |
| | d_I_2006 | 4.70 | 1.057 | 4.452 | 0.000 *** |
| | d_II_2006 | 4.82 | 1.108 | 4.352 | 0.000 *** |
| $B_t$ | d_I_2007 | 5.16 | 1.145 | 4.507 | 0.000 *** |
| | d_II_2007 | 4.34 | 1.187 | 3.654 | 0.000 *** |
| | d_I_2008 | 2.80 | 1.208 | 2.319 | 0.021 * |
| | d_II_2008 | 2.00 | 1.196 | 1.669 | 0.096 . |
| | d_I_2009 | 2.69 | 1.170 | 2.299 | 0.022 * |
| | d_II_2009 | 2.37 | 1.170 | 2.027 | 0.043 * |
| | d_I_2010 | 2.20 | 1.159 | 1.896 | 0.059 . |
| | d_II_2010 | 2.26 | 1.150 | 1.964 | 0.050 . |
| | d_I_2011 | 1.14 | 1.150 | 0.995 | 0.320 |
| | d_II_2011 | 0.42 | 1.126 | 0.376 | 0.707 |

**Figure 2** $A_i$ and $C_i$ coefficients

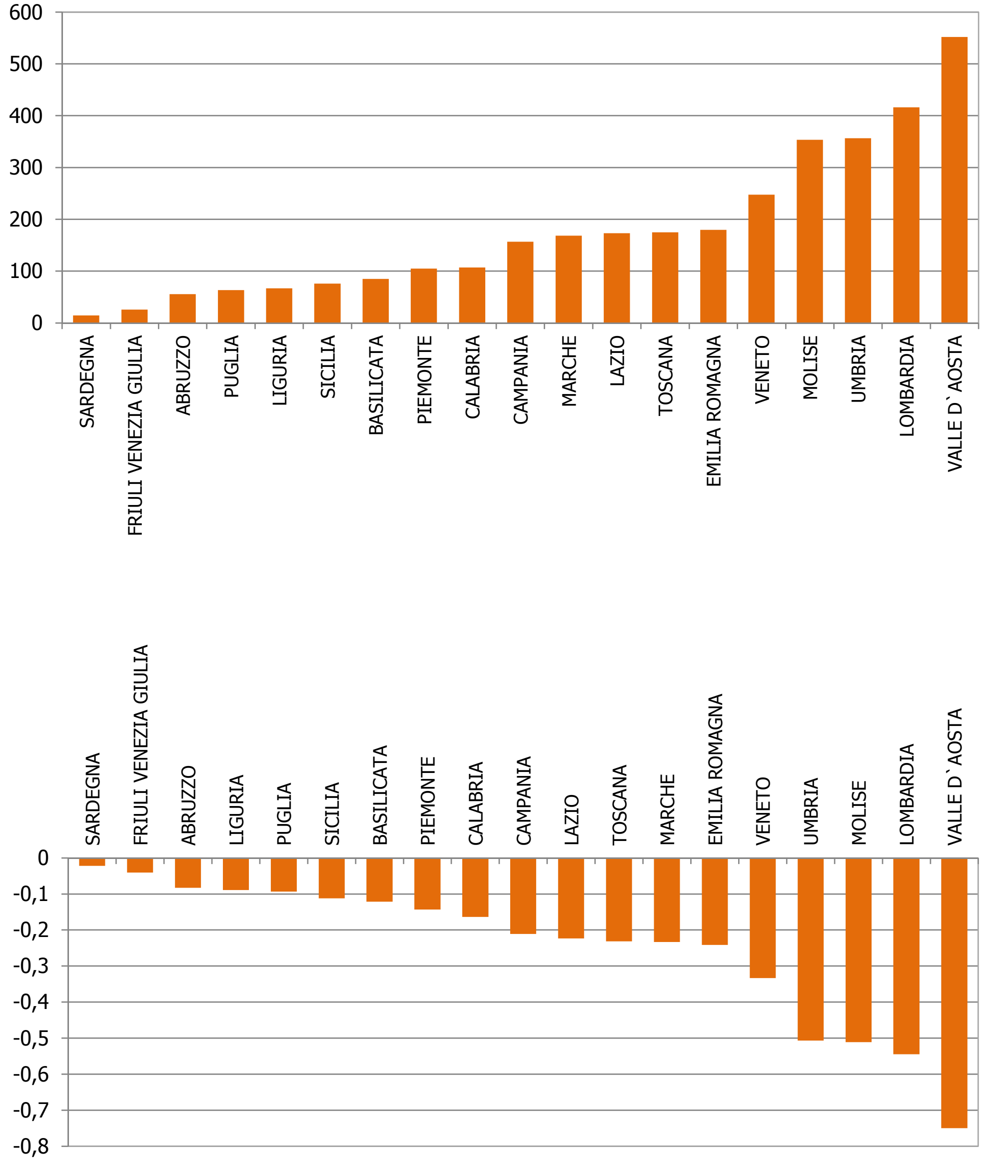
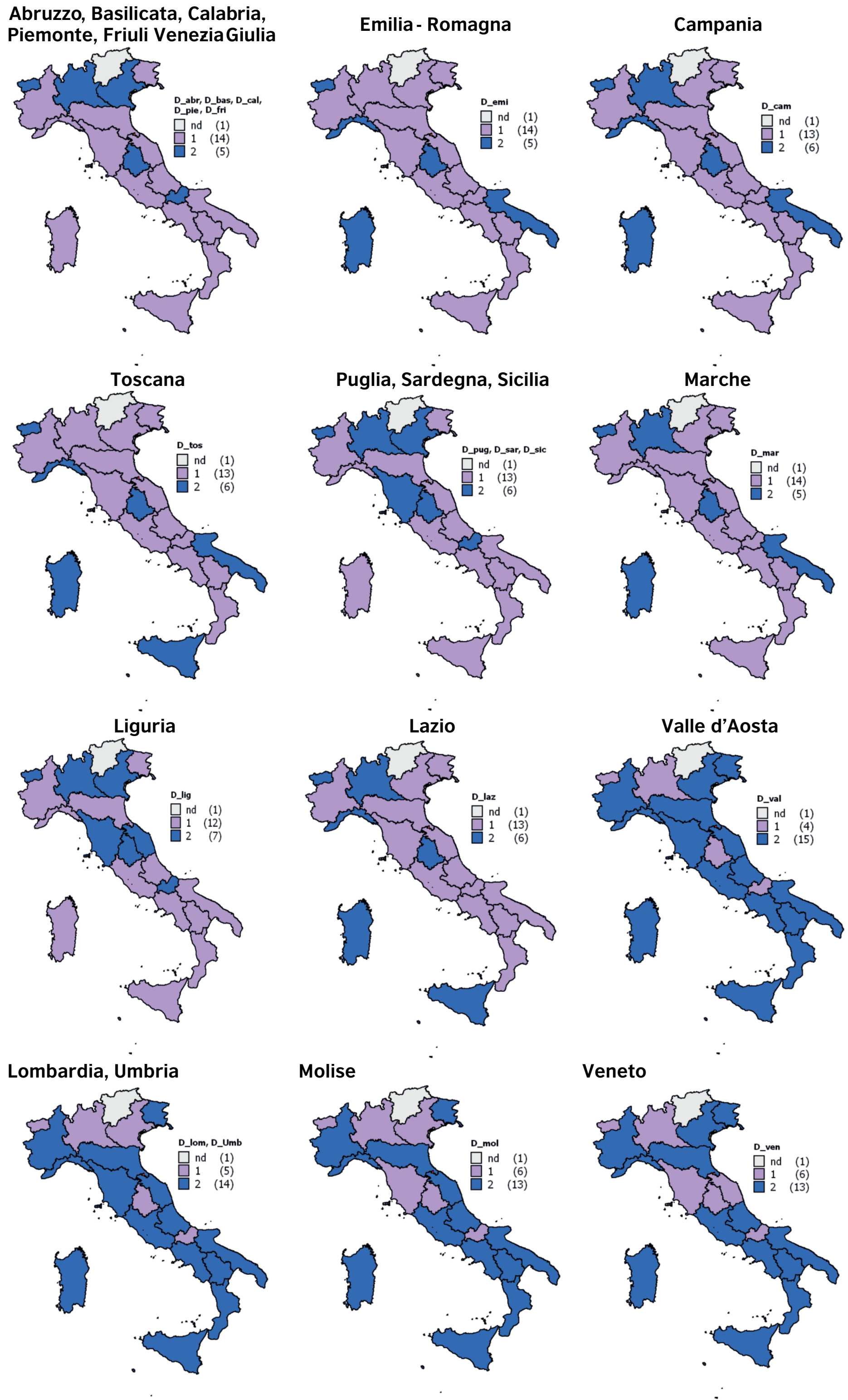


**Figure 3** $A_i$ significant maps



Each map shows the results of the pairwise tests on coefficient $A_i$ for every region. For example, the coefficient about Abruzzo was compared with the all other coefficients as follows tests:
$H_0$: $A_{Abruzzo} = A_{Basilicata}$ against $H_1$: $A_{Abruzzo} \neq A_{Basilicata}$...
and so on until test $H_0$: $A_{Abruzzo} = A_{Veneto}$ against $H_1$: $A_{Abruzzo} \neq A_{Veneto}$
if we accept the null hypothesis than the differences are not statistically significant.
In the map, the modality 1 indicates not statistical difference between the coefficients, on the contrary the modality 2 indicates statistical significant difference between the coefficients.
We can read the legend of the map as follows:

- in cluster 1 we have the regions which the coefficients are statistically different;
- in cluster 2 we have the regions which the coefficients are not statistically different;
- in nd (not available) we have the Trentino Alto Adige region that is not included in the analysis.

We have grouped the regions with the same results. For instance, the first map shows that the pairwise tests on Abruzzo's coefficient produce the significant differences with all other regions except for Veneto, Umbria, Lombardia, Valle d'Aosta and Molise. In a similar way we obtain the same results for the Basilicata, Calabria, Piemonte and Friuli-Venezia Giulia.

**Table 2** PC1 and PC2 weights

| VARIABLE | PC1 | PC2 |
|---|---|---|
| ABRUZZO | 0.236 | 0.223 |
| BASILICATA | 0.246 | 0.161 |
| CALABRIA | 0.25 | 0.084 |
| CAMPANIA | 0.252 | -0.089 |
| EMILIA ROMAGNA | 0.249 | -0.113 |
| FRIULI VENEZIA GIULIA | -0.186 | -0.386 |
| LAZIO | 0.254 | -0.008 |
| LIGURIA | 0.25 | 0.102 |
| LOMBARDIA | 0.185 | -0.422 |
| MARCHE | 0.254 | -0.05 |
| MOLISE | 0.153 | -0.471 |
| PIEMONTE | 0.243 | 0.129 |
| PUGLIA | 0.25 | 0.118 |
| SARDEGNA | 0.217 | 0.321 |
| SICILIA | 0.246 | 0.109 |
| TOSCANA | 0.254 | -0.059 |
| UMBRIA | 0.21 | -0.269 |
| VALLE D'AOSTA | 0.126 | -0.283 |
| VENETO | 0.242 | -0.184 |
| Eigenvalue | 15.102 | 2.429 |
| % Cumulative variance explained | 79.5% | 92.3% |

**Figure 4** Eigenvalues of the principal components

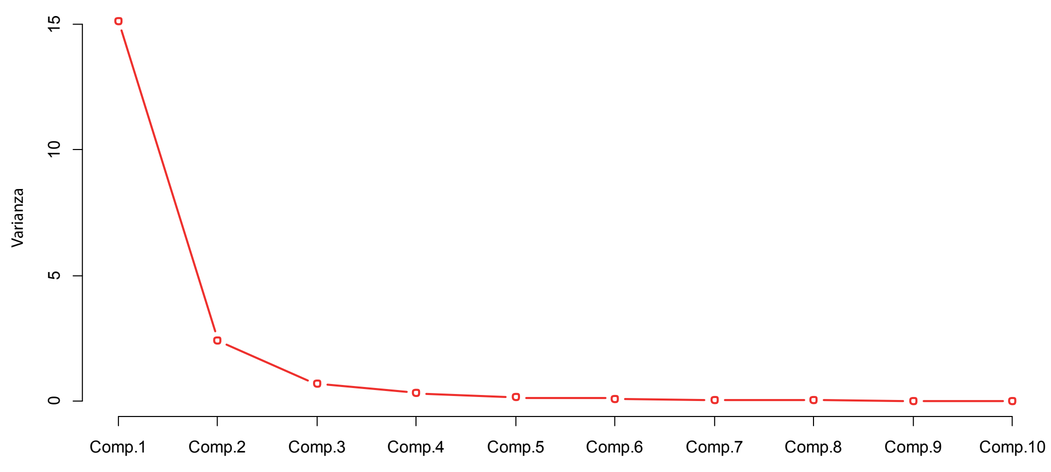



**Figure 5** Biplot and scatter plot of the first two components

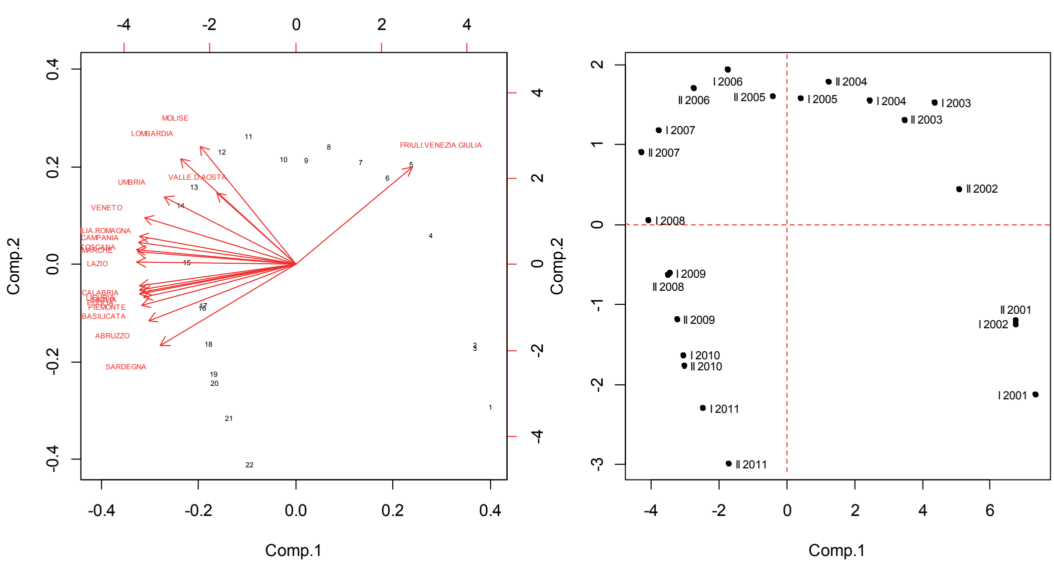


**Table 3** Coefficient values of panel data estimation in the long run

| Variables | Estimate | Std Error | t Stat | P-value |
|---|---|---|---|---|
| Pit-1 | -0.10 | 0.020 | -5.231 | 0.000 *** |
| d_Abr | 73.03 | 13.545 | 5.392 | 0.000 *** |
| d_Bas | 75.05 | 13.833 | 5.425 | 0.000 *** |
| d_Cal | 70.09 | 12.904 | 5.432 | 0.000 *** |
| d_Cam | 79.59 | 14.678 | 5.422 | 0.000 *** |
| d_Emi | 79.13 | 14.699 | 5.383 | 0.000 *** |
| d_Fri | 73.28 | 13.852 | 5.290 | 0.000 *** |
| d_Laz | 83.27 | 15.213 | 5.474 | 0.000 *** |
| d_Lig | 80.31 | 14.675 | 5.473 | 0.000 *** |
| d_Lom | 80.34 | 15.052 | 5.337 | 0.000 *** |
| d_Mar | 77.07 | 14.241 | 5.412 | 0.000 *** |
| d_Mol | 72.57 | 13.615 | 5.330 | 0.000 *** |
| d_Pie | 78.52 | 14.451 | 5.433 | 0.000 *** |
| d_Pug | 73.38 | 13.447 | 5.457 | 0.000 *** |
| d_Sar | 75.55 | 14.000 | 5.396 | 0.000 *** |
| d_Sic | 72.82 | 13.461 | 5.410 | 0.000 *** |
| d_Tos | 80.78 | 14.881 | 5.428 | 0.000 *** |
| d_Umb | 74.35 | 13.848 | 5.369 | 0.000 *** |
| d_Val | 77.05 | 14.483 | 5.320 | 0.000 *** |
| d_Ven | 78.91 | 14.609 | 5.402 | 0.000 *** |
| d_I_2002 | -2.66 | 0.922 | -2.883 | 0.004 ** |
| d_II_2002 | 2.58 | 0.922 | 2.802 | 0.005 ** |
| d_I_2003 | -0.35 | 0.932 | -0.372 | 0.710 |
| d_II_2003 | -0.04 | 0.937 | -0.047 | 0.963 |
| d_I_2004 | 0.96 | 0.943 | 1.021 | 0.308 |
| d_II_2004 | 1.73 | 0.955 | 1.817 | 0.070 . |
| d_I_2005 | 1.44 | 0.972 | 1.478 | 0.140 |
| d_II_2005 | 1.39 | 0.988 | 1.405 | 0.161 |
| d_I_2006 | 2.43 | 1.004 | 2.423 | 0.016 * |
| d_II_2006 | 2.16 | 1.028 | 2.099 | 0.036 * |
| d_I_2007 | 2.39 | 1.049 | 2.280 | 0.023 * |
| d_II_2007 | 1.46 | 1.072 | 1.363 | 0.174 |
| d_I_2008 | -0.12 | 1.084 | -0.110 | 0.912 |
| d_II_2008 | -0.76 | 1.078 | -0.710 | 0.478 |
| d_I_2009 | 0.16 | 1.066 | 0.155 | 0.877 |
| d_II_2009 | -0.16 | 1.065 | -0.151 | 0.880 |
| d_I_2010 | -0.19 | 1.061 | -0.182 | 0.855 |
| d_II_2010 | -0.01 | 1.057 | -0.008 | 0.994 |
| d_I_2011 | -1.11 | 1.055 | -1.050 | 0.294 |
| d_II_2011 | -1.63 | 1.043 | -1.567 | 0.118 |

**Table 4** PC1 Weights for the LPC variables

| VARIABLE | PC1 |
|---|---|
| ABRUZZO | -0.225 |
| BASILICATA | 0.108 |
| CALABRIA | -0.113 |
| CAMPANIA | 0.171 |
| EMILIA ROMAGNA | -0.267 |
| FRIULI VENEZIA GIULIA | -0.306 |
| LAZIO | 0.280 |
| LIGURIA | 0.283 |
| LOMBARDIA | -0.305 |
| MARCHE | -0.168 |
| MOLISE | -0.301 |
| PIEMONTE | 0.062 |
| PUGLIA | 0.267 |
| SARDEGNA | 0.003 |
| SICILIA | 0.100 |
| TOSCANA | 0.181 |
| UMBRIA | -0.298 |
| VALLE D'AOSTA | -0.294 |
| VENETO | -0.259 |
| Eigenvalue | 10.421 |
| % Cumulative variance explained | 54.9% |

**Table 5** Stationarity test on LPC first component

| | ADF | ADF-GLS | KPSS |
|---|---|---|---|
| $H_0$ | I(1) | I(1) | I(0) |
| Test | -1.455 | -0.023 | 0.370 |
| *p-value* | 0.556 | 0.675 | 0.095 |

These tests verify the presence of a unit root in a time series. We implement the Augmented Dickey - Fuller (ADF) test and a its variant the ADF - GLS test, that performs a generalised least squares regression, under the null hypothesis of non stationarity of the series. On the contrary, the null hypothesis of the KPSS test (Kwiatkowski, Phillips, Schmidt and Shin, 1992) is the stationarity.