WIFO
ÖSTERREICHISCHES INSTITUT
FÜR WIRTSCHAFTSFORSCHUNG

WORKING PAPERS

Cycles Inside Cycles

Spanish Regional Aggregation

**Maria Dolores Gadea, Ana Gomez Loscos,
Antonio Montañes (University of Zaragoza)**

390/2011

Cycles Inside Cycles

Spanish Regional Aggregation

Maria Dolores Gadea, Ana Gomez Loscos, Antonio Montañes (University of Zaragoza)

WIFO Working Papers, No. 390
February 2011

E-mail address: lgadea@unizar.es

2011/045/W/0

Medieninhaber (Verleger), Hersteller: Österreichisches Institut für Wirtschaftsforschung • 1030 Wien, Arsenal, Objekt 20 • Tel. (43 1) 798 26 01-0 • Fax (43 1) 798 93 86 • http://www.wifo.ac.at/ • Verlags- und Herstellungsort: Wien

Cycles inside cycles. Spanish regional aggregation*

By Maria Dolores Gadea[a]†, Ana Gomez Loscos[b] and Antonio Montañes[c]

[a]Dept. of Applied Economics, U. de Zaragoza ; [b]Fundear and Dept. of Applied Economics, U. de Zaragoza ; [c]Dept. of Economic Analysis, U. de Zaragoza

June 2010

Abstract

This paper sets out a comprehensive framework to identify regional business cycles within Spain and analyses their stylised features and the degree of synchronization present among them and the Spanish economy. We show that the regional cycles are quite heterogeneous although they display some degree of synchronization that can be partially explained using macroeconomic variables. We also propose a dynamic factor model to cluster the regional comovements and find out if the country cycle is simply the aggregation of the regional ones. We find that the Spanish business cycle is not shared by the seventeen regions, but is the sum of the different regional behaviours. The implications derived from our results are useful both for policy makers and analysts.

Keywords: Business Cycle. Synchronization measures. Dynamic factor models. Regional policy.

JEL Classification: C22,C32,E32,R11

*The authors are grateful to two anonymous referees for their useful suggestions and the editor for his encouragement. We would also like to thank M. Camacho, J. R. Cancelo, B. Candelon, J. M. Casado, E. M. Quilis and G. Pérez-Quirós, for their helpful comments on previous versions of this paper. We acknowledge financial support from the Ministerio de Ciencia y Tecnología through grants ECO2008/03040 and ECO2009/13085. The usual disclaimer applies.

†Corresponding Author: M. Dolores Gadea, Dept. of Applied Economics, Universidad de Zaragoza. Gran Vía 2, 50005 Zaragoza, Spain. E-mail: lgadea@unizar.es

"Aggregative analysis ... not only does not tell the whole tale but necessarily obliterates the main... point of the tale"

(J. A. Schumpeter, Business Cycles, New York, McGraw-Hill, 1939, Vol. I, p.134)

1. INTRODUCTION

In a context of growing economic and monetary integration such as the recent creation of the EMU in Europe (1999), there is still controversy as to whether comovements within economies are high enough for these processes to be carried out successfully. This paper focuses on a previous step, employing a lower aggregation level, to find the way in which comovements within a country determine its business cycle, that is, the cycles inside the cycles usually considered in the literature. We employ two different approaches. Firstly, we set out a framework to identify the regional business cycles and their stylized features, compare them to the national cycle and analyze their synchronization. Secondly, we estimate a Dynamic Factor Model to detect common and idiosyncratic factors in the regional cycles and to determine whether the country cycle is shared by all the regional business cycles or whether, on the contrary and as we suspect, it is the consequence of aggregating different regional business cycles.

The usual practice of considering a country's business cycle as an aggregation of the regional ones may mask very different activity rhythms. The loss of regional detail would be negligible if the divergence between regional and national cycles were small. If, on the other hand, the divergence were large, it would make it difficult to apply policies satisfactorily in all parts of the country and would have important implications, not considered up to now, for integration processes. In short, knowing the regional cycle path should be a key question in the design of the economic policy.

This paper aims to determine the pattern of regional business cycles within Spain, to check which peculiarities are shared by the regions that are more synchronized or coordinated with the rest, to provide empirical evidence for the existence of different common regional business cycles and to analyze their synchronization with the Spanish aggregate cycle. This approach makes sense in industrialized countries where lower aggregation levels are significant (regions or counties), and with federal fiscal systems that allow differential economic policies to be implemented. We focus on Spain –a country divided into seventeen NUTS-2 regions with a high degree of fiscal federalism- and we test if the Spanish cycle is really unique and its path is shared by the seventeen regional business cycles. If not, a centralized economic policy may not be the most suitable alternative and would contribute to intensifying regional inequalities. So, it would be important to establish a comprehensive framework of the regional cycles in the Spanish economy in order to design policies for the less coordinated regions to improve the specific factors that create the differences.

The studies that identify cyclical patterns have mainly been applied to countries and there are few concerning a lower level mostly because of the absence of adequate data[1]. Nevertheless, recently, two papers have studied similarities and differences across US states during the different phases of their business cycles. The first, by Hamilton and Owyang (2009), uses common Markov-switching components in a panel data set. The second, by Owyang *et al.* (2005), applies a regime-switching model to state-level coincident indices[2]. On the whole, when GDP was not available, the attempts to investigate regional business cycles have used employment variables as a proxy for economic indicators of activity and hardly any of them use industrial production indexes as we do in this paper[3].

In spite of having suitable characteristics for this type of analysis, the previous literature on Spain is scarce. The Spanish cycles have been studied without going into regional behaviours and using GDP or employment variables, the latter being a less accurate in-

[1] However, the study of cyclical patterns for countries has a long tradition, as is detailed in Section 3.1.

[2] Three other main lines in the field of regional cycles have received attention from economists and mainly focus on eight major US regions. The first one considers the regional transmission of cyclical impulses (see Metzler (1950), Airov (1963), Carlino and Defina (1995), Carlino and Sill (1997) and Kouparitsas (2002)). The second focuses on the effect of determined shocks or policies on the economy (see Carlino and Defina (1998), Kozlowski (1995) and Garrison and Chang (1979)). The last tries to explain the regional cycles relating them to their growth patterns (see Borts (1960) and Carlino and Sill (2001)).

[3] We use the series of industrial production indexes that better (than employment variables) fit the economic fluctuations in the Spanish regions. As far as we know, the only paper applied to regions to use this index is Rodríguez and Villemaire (2004) for Canada.

dicator of economic activity than the industrial production index (Dolado *et al.* (1993), Dolado and María-Dolores (2001) and Doménech and Gómez (2005))[4]. The only papers that try to characterize Spanish regional business cycles are Cancelo and Uriz (2003) and Cancelo (2004). They use employment data and analyze turning points, comovements and bidirectional causality.

Never before has such a comprehensive study of the regional business cycles in Spain been carried out. Our analysis can be divided into two large blocks. In the first block, we first identify Spanish regional business cycles through the Bry-Boschan non-parametric technique (1971), defining turning points in a way quite close to the one used by the National Bureau of Economic Research (NBER), which allows us to determine the different phases of the business cycle that, in most cases, closely follow the general path of the Spanish economy during the same period. Following Harding and Pagan (2002a), we also present some key features for describing the business cycles, such as their amplitude, cumulation and excess of recessions and expansions. Second, we explore their synchronization using different measures that consider the degree of comovement between each region and the others and with Spain as a whole. Third, the role of some macroeconomic variables that could explain the synchronization across regional economies is analyzed. These variables are the industrial composition, the per capita income level, human capital and the unemployment rate. The results obtained allow us to define some key lines for the future implementation of any measure of economic policy that tries to increase intracountry synchronization. We also consider another dimension that may influence the explanation of the similarities and differences in regional comovements, namely, neighbourhood.

In the second block, we first carry out a preliminary analysis of comovements by using the coherence and cohesion measures that allow us to know the degree of synchronization at different frequencies. Secondly, although there is a wide variety of clustering techniques that could have been applied to the basic features of the cycle to form clusters of regions, we complement our study with the use of Dynamic Factor Models to investigate whether

[4] There are also the papers that analyse the European business cycles and, so, include Spain (see Camacho *et. al* (2008 and 2006), Artis *et. al* (2004) and Croux *et. al* (2001), amongst others).

some common factors could be driving the regional business cycles. Then, we compare these results with the Spanish cycle and identify the national component. Based on the results obtained, we carry out the cluster analysis using the idiosyncratic regional components.

Our contribution is twofold. Firstly, we find a high degree of heterogeneity in the basic features of the regions' cycle dating. When we test for synchronization, we obtain that, although the results are not symmetrical across regions, the regional cycles are sufficiently correlated to consider the possibility of the existence of common cycles. The most synchronized regions are characterised by an important industrial weight, per capita income and human capital and a low unemployment rate. One outstanding result is that there is an inverse relationship between economic growth and regional comovements or synchronization. Furthermore, most regions have a high coherence/cohesion in the long run that decreases dramatically in the short run and they do not show signs of convergence. Secondly, we find a common dynamic factor associated with Spain. After removing this factor, we use regional idiosyncratic components to cluster regions and we can confirm the idea that the Spanish business cycle is a result of aggregating regional ones that do not share identical patterns. This finding should be taken into account either to study and forecast the Spanish business cycle or to implement economic policy.

The paper is organized as follows. Section 2 describes the data, presenting the stylized facts in Spain as a whole and in its regions. Section 3 defines the regional cycles and the basic features that characterize them as well as their mutual synchronization and their synchronization with Spain. It also proposes some variables that may have a role in explaining these results. Section 4 presents a preliminary analysis of comovements and then investigates whether common driving forces appear in the regional business cycles. From these results, we identify clusters of regions. Finally, Section 5 concludes.

2. STYLIZED FACTS ABOUT SPAIN AND SPANISH REGIONS.

In this paper, we consider the 17 Spanish Autonomous Communities that correspond to NUTS-2 in the EUROSTAT nomenclature. Each region is denoted by its acronym: Andalucia (AND), Aragón (ARA), Asturias (AST), Baleares (BAL), Canarias (CAN), Cantabria (CANT), Castilla y León (CYL), Castilla-La Mancha (CLM), Cataluña (CAT), Comunidad Valenciana (CVAL), Extremadura (EXT), Galicia (GAL), Madrid (MAD), Murcia (MUR), Navarra (NAV), País Vasco (PVAS) and La Rioja (LAR).

We concentrate on the analysis of the monthly industrial production index (IPI), extracted from the Instituto Nacional de Estadística (Spanish Statistical Institute, INE), for Spain as a whole and for its 17 regions. Our working sample spans from 1991:10 to 2009:09, and we have linked two different series. With base year 1990, we have data from 1991:10 to 2002:12 while, with base year 2005, the available data are from 2002:01 to 2009:09. Thereby, we obtain 216 observations for each region and for Spain as a whole.

This is the first time this index has been used to measure regional business cycles in Spain and it could be controversial but we can not use the regional accounts series, such as the GDP, which are more comprehensive measures of aggregate activity, because they are not sufficiently long and only have an annual frequency. The IPI is monthly, better than the annual or quarterly frequency of other variables (production or employment). Furthermore, the IPI is one of the main series when estimating regional GDP so, the manufacturing sector reflects the rhythm of the business cycle more accurately than employment data, which are those most commonly used in regional studies. The GDP shows a cyclical path similar to the industrial sector, but smoothed by the more stable behaviour of the tertiary sector.

In fact, a glance at the growth rates of GDP and industrial production since the 90s shows that they have similar profiles, although they are more pronounced in the case of the industrial sector because it is more affected by business cycle fluctuations. In the period analyzed, we can distinguish four main phases in the Spanish economic cycle[5]: the end of the expansion of the 80s, the profound crisis of 1992-1993, the dilated recovery that began in 1994 and included years of high growth and periods of slower growth, more marked in industrial activity and, finally, the current recession that started in 2008[6].

[5] Although our data source begins in October 1991, we capture the end of the expansion of the 80s.

[6] Taking quarterly data for both Spanish GDP and IPI from 1991.4 onwards, the correlation between the two variables is really high: 0.81 with annual growth rates and 0.66 with quarterly growth rates.

The degree of heterogeneity between sectors in the Spanish regions is high. In the almost twenty years considered in this study, the services sector, followed by construction, are the ones that have increased their weight in the total production at the expense of the industrial, energy and agricultural sectors, in that order. In 2008, the last year with available regional accounts data, 69% of the total Spanish GDP is generated by the services sector as in the most developed economies; 14.3% is industrial, 11.4% comes from construction and the remaining 2.6% and 2.7% belong to the agriculture and energy sectors.

The most industrialised regions in the whole period (those with an industrial weight clearly above the Spanish average, which is 17.18%) are NAV, LAR, PVAS, CAT, ARA CVAL and CANT. There are two groups with an industrial average similar to the Spanish one: AST and CYL (a little higher) and CLM and GAL (slightly below). Finally, the less industrialised regions are MUR, MAD, AND, EXT, BAL and CAN, although the distance between the first two and the last two is of almost 10 pp.

The IPI series have previously been seasonally adjusted. A preliminary analysis is carried out by applying the MZt-GLS unit root tests proposed by Ng and Perron (2001), which are modified forms of the Phillips-Perron test [Phillips and Perron (1988)] and based on the detrended GLS data. We also use the KPSS of Kwiatkowski *et al.* (1992) that tests for the null hypothesis of stationarity. Both have been applied to two different specifications, one that includes an intercept and one that has an intercept and a trend. The results lead us to a very robust conclusion: we cannot reject the presence of a unit root in the series, whilst we can reject the null of stationarity[7]. Consequently, in the rest of the analysis, we have to take these properties of the series into account. Figures 1 and 2 show the levels of seasonally adjusted series and the first difference of their logs, respectively.

3. BUSINESS CYCLE DATING AND SYNCHRONIZATION

This section provides a complete framework for the analysis of the Spanish regional business cycles. Firstly, we select an appropriate turning points dating method to identify

[7]Detailed results of these tests are available from the authors upon request.

regional business cycles. Secondly, we illustrate the key features that describe these business cycles. Thirdly, we examine the degree of similarity among the business cycles identified in Spain and its regions. Finally, we study both the role of different macroeconomic variables in explaining the degree of business cycle synchronization and whether neighborhood presents any effects.

3.1 Cycle dating and basic features

The seminal work of Burns and Mitchell (1946) paved the way for methods to measure the business cycle. These authors define the cycle as a pattern in the level of aggregate economic activity, and describe it through a two-stage methodology. First, turning points are located in the series by using graphical methods, thereby defining specific cycles. Second, the specific cycle information is distilled into a single set of turning points that identify the reference cycle. These authors also define concepts such as peak (the high point of an expansion) and trough (the worst moment in a recession period) to determine the cycle length. These terms became standard in any work about business cycles undertaken after the publication of that work.

Their approach has important advantages for academics and politicians because of the ease of computing algorithms to establish the dates at which there were turning points in the business cycle, and because of the intuitive interpretation of the results. Their aggregate cycle was called the business cycle, and their tools were immediately used by the NBER to study US business cycles in greater depth and, afterwards, became a reference for the study of business cycles in other economies. Nowadays, NBER continues to publish a single set of turning points for the US economy.

This pioneering work generated a great deal of literature in which the level of sophistication of the statistical tools evolves more than the definition of the business cycles. Bry and Boschan (1971) (BB) developed the most popular non-parametric method to determine when the peaks and troughs, which frame economic recessions or expansions, appear. In the last few decades, many alternative procedures have been suggested. Among them, the

Markov-switching (MS) approach proposed by Hamilton (1989) stands out[8]. Unlike to the BB method, the MS first fits a statistical model to the data and then uses the estimated parameters to determine the turning points. Since the well-known paper of Hamilton (1989), there has been a rebirth of interest in this method as an alternative to classical business cycle measures[9]. The MS models try to characterize the evolution of a variable through a process of conditioned mean to a state of a specific nature. The changes in value in this dynamic process will allow us to differentiate periods of expansions and contractions. Regime shifts are governed by a stochastic and unobservable variable which follows a Markov chain.

Except for the US, for which the NBER Business Cycle Dating Committee establishes the official chronology or turning points, there are no widely accepted reference chronologies of the classical business cycle for other countries. So, the examination of the synchronization of Spain and the 17 Spanish regions will have to rely on dating algorithms that can be either non-parametric (Bry-Boschan type methods) or parametric (Markov-switching models).

The selection of the most suitable cycle dating algorithm for Spain and its regions is very conditioned by the data and its sample size. In contrast to BB type methods, which are valid for the sample size used in this analysis, the Markov-switching method requires a longer sequence of business cycle states to estimate the transition probability matrix coefficients with a reasonable degree of confidence. We only have monthly observations from 1991:10 to 2009:9 and a glance at the growth rates of GDP and industrial added value shows that we can distinguish four large phases in the Spanish economic cycle. So, we have reasonable doubts about the ability of the Markov-switching procedure to adequately estimate the probability of staying in recession or in expansion or the transition probabilities between regimes. Although non-parametric procedures are based on some assumptions about the duration of business cycles and about the detection of false signals, they can be much more

---

[8] See Harding and Pagan (2002b and 2003) and Hamilton (2001) for a debate about the two business cycle dating methods.

[9] Krolzig (1997), Artis *et al.* (2004) and Krolzig and Toro (2005), amongst others, have highlighted the ability of this parametric approach to capture stylized business cycle features. MS-VAR models offer more robust statistical tools.

appropriate for short samples and for a small number of changes in regime transitions[10].

Furthermore, an eye-ball examination of Figures 1 and 2 suggests that the series in levels would create important problems in computing turning points with BB techniques. On the contrary, the growth rates of the series seem too noisy and the algorithm could produce false signals. Figure 3 shows the original level series together with the recessions (represented by bars) identified with the BB methods[11]. They clearly coincide with periods of falls in industrial production and reproduce the "known-knowns" recent aggregate Spanish business cycle. Consequently, we can trust BB techniques to find the chronology of turning points at regional level. With respect to the regional cycles, in most cases, the BB method locates four recession periods which coincide with two well-known crises, the beginning of the 90's and the current one, and two deceleration episodes during the long expansion. Nevertheless, some regions present a more turbulent chronology. AST very frequently alternates expansion and recession periods throughout the sample; CVAL has spent almost all the noughties in recession; and, in AND and EXT, the algorithm does not yet detect the deep recession that started in 2008. So, we should analyze the basic characteristics of the regional cycles in greater detail.

Following Harding and Pagan (2002a), we dissect the business cycle and calculate some outcomes, such as the probability of recessions measured as the number of months in recession over the total, the mean duration, amplitude, cumulation and excess of recessions and expansions[12]. All these results appear in Figure 4. The probability of recessions is 0.38 for Spain (0.39, on average for the regions), the mean duration of the recessions is 17 months (14) and the mean duration of the expansions is 45 months (35)[13]. These results are plau-

[10] An aplication of the MS method to Spanish regional cycles can be found in Gadea *et al.* (2006).

[11] We have used the code written by Watson and Denson in Watson (1994).

[12] Harding and Pagan (2002a) propose a graphical representation of the cyclical phase as a triangle whose height is the amplitude and whose base is the duration. The area of triangle is an approximation to the cumulated gains or losses in output from trough to peak and peak to trough, respectively. In these calculations, we have used logs of the series to obtain more representative figures, such as growth rates.

[13] Just to put these figures in context, they closely agree with the estimated duration of business cycle phases proposed by the NBER for the 32 cycles in the recent history of the US (1854-2001), which is 17 and 38 months for recessions and expansions, respectively. According to Camacho *et al.* (2006), European

sible and agree with the stylized fact that expansion periods are longer than recessions[14]. It is noteworthy that, while recessions last a similar number of months to those in Europe, expansions last more than two years in Spain and half a year for the regions, on average, which could explain the convergence process attained during the sample period.

However, we find some heterogeneity in the probability of recession and the duration of the cycles across regions. As we said previously, in the seventeen regions (and in Spain), the average probability of a recession is nearly 40% but only seven of these eighteen geographical units are above the mean. CVAL and AST are the regions with the highest probability, more than 50%, and the lowest probabilities, around 30%, are found in MUR, MAD and ARA. With respect to the duration, MUR, MAD and ARA stand out as the regions with the longest expansions (around 50 months) and the shortest-lived recessions (only a fifth of the cycle duration is spent in a recession). Nevertheless, CLM and CVAL present long-lasting recessions (more than 20 months). Furthermore, we can appreciate big asymmetries in duration in the cases of MUR and ARA; while, on average, expansions last 2.5 times as long as recessions, in these two regions, the ratio is around 5 times. On the contrary, the highly symmetrical cycles found in CVAL (and, to a lesser extent, in AST and CLM) are noticeable, as the time spent in expansion is 23 months while recessions last 21.

Clear asymmetries between the amplitude of the phases of the cycle are also observed. This measure, expressed in percentage, shows the gains or losses in industrial production as a result of expansions or recessions. It is clear that, on the whole, expansions are wider than recessions; on average, there is a difference between the two phases of about ten percentage points and, in all regions, the amplitude of expansions is bigger than that of recessions. Of special interest is the case of EXT, where the amplitude across the two phases is almost identical because the region presents the highest loss during recessions, and the cases of NAV and ARA, two regions that show very pronounced asymmetries between the amplitude of

---

expansions last about 30 months, while recessions last 15 months. This means that a cycle spends 67% of its duration in expansion.

[14] However, if we work with growth rates, the algorithm identifies more turning points and the duration of recessions and expansions are 22 and 27 months, respectively, for the case of Spain, and 24 and 23 for the regions on average.

their business cycle phases, the expansions being more than four times greater than the recessions.

MAD, ARA, LAR and NAV have the clearest cumulative gains during expansions, and CLM and EXT, stand out for the severity of their recessions. Cumulation is a measure used to identify the cumulated gain or loss, calculated as the sum of the amplitudes for each period of the phase. It is very useful as it can be interpreted as the gain or loss in wealth in the economy, and sums up the previous ones by combining the duration, amplitude and shape of the business cycle. However, it is normally calculated by the triangle approximation and differs from the actual cumulation because the path through the phase may not be well estimated by a triangle. Harding and Pagan (2002a) propose adding up the area of rectangles and removing the bias as a more accurate measure of the area. Nevertheless, we have calculated the area associated with the phase exactly by using methods of numerical integration. The difference between the actual shape and its triangle approximation is known as excess.

Negative excess dominates during expansions, so the shape of the wealth gain is mainly concave. This means that the path of this phase begins with steep changes and ends smoothly. That is, an expansion is more commonly characterized by a high growth period that ends in normal growth period. During recessions, the concave shape of the phase dominates even more than during expansions. Consequently, the paths exhibit gradual changes at the beginning of the phase that become sharp at the end. Both features are positive because they mean that the wealth losses in recessions are lower and the gains in expansions higher than in a linear behaviour. So, regions with convex expansions (GAL, CAN, CYL, LAR, EXT, PVAS) and SP do not benefit as much as the others from an expansion, while regions with convex recessions (MUR, CYL, ARA, LAR and AST) have a bigger wealth loss than the others when they are in recession.

Summing up, the BB method has allowed us to obtain the cycle dating and a first picture of the Spanish and regional business cycles, showing that they seem to have important disparities. We can affirm the existence of 17 non-identical regional business cycles in Spain. Nevertheless, it is possible that these cycles exhibit some synchronization, which could be interpreted as a sign that regional economies move together.

3.2 Measures of synchronization

We have proved the existence of different patterns in the regional business cycles. However, they could be coordinated between them so, in this section, we focus on the study of the possible relationships between the cyclical patterns of industrial activity in the different regions. In particular, we want to explore their possible synchronization in depth. To that end, we use different measures of the synchronization of cycles such as Pearson's coefficient and its independence test, both based on a contingency table, and the index and test statistic of concordance proposed by Harding and Pagan (2006), both in their bivariate and multivariate forms. Finally, if we find some degree of synchronization, we will apply multidimensional scaling techniques to represent the different regions on a map, looking for groups of regions with similarities in their cyclical path. In addition, we explore the evolution of synchronization over time and its relationship with important regional characteristics such as, industrial weight, unemployment rate, income level, human capital and geographical position.

The well-known independence test for regions i and j, is based on a contingency table where the frequencies of expansion and recession observations are shown for the two regions. This statistic has the following expression:

$$Q_{ij} = \sum_{u=1}^{s} \sum_{v=1}^{s} \frac{(n_{uv} - \widehat{m}_{uv})^2}{\widehat{m}_{uv}},$$

where s is the number of regimes, n denotes the joint observed frequencies and m the estimated marginal frequencies . The statistic is distributed under the null of independence as a χ^2 with (s-1)x(s-1) degrees of freedom. We can also compute the contingency coefficient which lies within the range [0,1] from lower to higher cycle commonality.

$$C_{ij} = \frac{1}{\sqrt{2}} \sqrt{\frac{Q_{ij}^2}{Q_{ij}^2 + T}},$$

According to Harding and Pagan (2004), for each $i-$region we can build a binary random variable S_{it}, taking value 1 when the $i-$region is in an expansion phase and zero when it is in a recession phase. The concordance index for two regions i, j is defined as follows:

$$I_{ij} = T^{-1}[\sum_{t=1}^{T}(S_{it}S_{jt}) + \sum_{t=1}^{T}(1 - S_{it})(1 - S_{jt})],$$

where T is the sample size. I_{ij} measures the proportion of time that the two regions are in the same phase. Notice that this index only shows similarities in the periodicity of regional cycles, independently of the length of the expansion and recession phases. Although this measure is very easy to interpret and offers a first picture of synchronization in regional cycles, it has the disadvantage that it does not provide a statistical way of knowing whether the comovements are significant or not. To solve this problem, Harding and Pagan (2004) suggest an alternative method based on the correlation between S_{jt} and S_{it}. They recommend estimating the coefficient which reflects the correlation between $S_{it}$and S_{jt} by using the generalized method of moments.

Starting with the following moment condition:

$$E[\sigma_{S_{it}}^{-1}(S_{it} - \mu_{S_{it}})\sigma_{S_{jt}}^{-1}(S_{jt} - \mu_{S_{jt}}) - \rho_{S_{ij}}],$$

where μ_{S_t} and $\sigma_{S_t}^{-1}$ are, respectively, the mean and standard deviation of the time series S_t, we can estimate the value of $\rho_{S_{ij}}$ and test if $\rho_{S_{ij}} = 0$ using the t-test in its implicit estimator equation:

$$T^{-1}\sum_{t=1}^{T}\widehat{\sigma}_{S_{it}}^{-1}(S_{it} - \mu_{S_{it}})\widehat{\sigma}_{S_{jt}}^{-1}(S_{jt} - \mu_{S_{jt}}) - \widehat{\rho}_{S_{ij}} = 0,$$

As Harding and Pagan (2006) recognize, $\widehat{\rho}_{S_{ij}}$ can be found from this regression:

$$\sigma_{S_{it}}^{-1}\sigma_{S_{jt}}^{-1}S_{it} = \alpha + \rho_{S_{ij}}\sigma_{S_{it}}^{-1}\sigma_{S_{jt}}^{-1}S_{jt} + \varepsilon_t,$$

The interpretation of the regression has advantages over the method of moments estimator because it allows us to analyze whether the degree of synchronization has changed over time. However, our inference has to be robust to the serial correlation as well as to any heteroskedasticity in the errors. We use the Newey-West autocorrelation-consistent covariance with Barlett weights and we also build confidence intervals following the stationary bootstrap techniques proposed by Politis and Romano (1994). This procedure is based on resampling blocks of random length, where the length of each block has a geometric distribution[15].

The multivariate version for n regions of this test is based on the following $n(n+1)/2$ moment conditions:

$$E[\frac{(S_{jt}-\mu_{S_j})(S_{it}-\mu_{S_i})}{\sqrt{\mu_{S_j}(1-\mu_{S_j})\mu_{S_i}(1-\mu_{S_i})}}-\rho_{S_{ij}}]=0,\ j=1,...,n,\ i>j$$

and the test has this expression:

$$W=\sqrt{T}g(\widehat{\theta}_0^{-1},\{S\}_{t=1}^{T})'\widehat{V}^{-1}\sqrt{T}g(\widehat{\theta}_0^{-1},\{S\}_{t=1}^{T})$$

where $\widehat{V}$ is a consistent estimate of the covariance matrix for $g(\widehat{\theta}_0^{-1},\{S\}_{t=1}^{T})$ and

$$g(\widehat{\theta}_0^{-1},\{S\}_{t=1}^{T})=\frac{1}{T}\sum_{t=1}^{T}h_t(\theta,S_t)$$

$$h_t(\theta,S_t)=\begin{bmatrix} S_{1t}-\mu_{S_1} \\ ... \\ S_{nt}-\mu_{S_n} \\ \frac{(S_{1t}-\mu_{S_1})(S_{2t}-\mu_{S_2})}{\sqrt{\mu_{S_1}(1-\mu_{S_1})\mu_{S_2}(1-\mu_{S_2})}}-\rho_{S_{12}} \\ ... \\ \frac{(S_{(n-1)}-\mu_{S_{(n-1)}})(S_{nt}-\mu_{S_n})}{\sqrt{\mu_{S_{(n-1)}}(1-\mu_{S_{(n-1)}})\mu_{S_n}(1-\mu_{S_n})}}-\rho_{S_{(n-1)n}} \end{bmatrix}$$

The vector $\widehat{\theta}'=[\widehat{\mu}_{S_1},...,\widehat{\mu}_{S_n},\widehat{\rho}_{S_{12}},...,\widehat{\rho}_{S_{(n-1)n}}]$ contains sample means and sample pairwise correlations and, under the null, has different expressions depending on the hypothesis. In the case of SMNS (strong multivariate non-synchronization), it is $[\widehat{\mu}_{S_1},...,\widehat{\mu}_{S_n},0,...0]$

[15] Following Camacho *et al.* (2006), we select the probability of the geometric distribution so its expected value is equal to the average duration of expansions.

or, if we want to test the hypothesis of SMS (strong multivariate synchronization), for instance $\rho_{S_{12}} = \rho$, $\forall i \neq j$ with $\rho \in (0,1)$, it is $[\widehat{\mu}_{S_1}, ..., \widehat{\mu}_{S_n}, \rho, ..., \rho]$. The W statistic has an asymptotic $\chi^2_{n(n-1)/2}$ distribution for $\rho \in [0,1)$. However, under the null of PS (perfect synchronization) the distribution is more complex, being a Brownian motion, or applying the Cramer-VonMises equivalent, a weighted average of χ^2 densities. In this case, Harding and Pagan (2006) propose an alternative statistic whose asymptotic density is a $\chi^2_{(n-1)}$:

$$W = \sqrt{T} g(\{S\}_{t=1}^{T})' \widehat{V}^{-1} \sqrt{T} g(\{S\}_{t=1}^{T})$$

where

$$g(\{S\}_{t=1}^{T}) = \frac{1}{T} \sum_{t=1}^{T} h_t(S_t)$$

and

$$h_t(S_t) = [-i_n - 1 I_{n-1}] \begin{bmatrix} S_{1t} \\ ... \\ S_{nt} \end{bmatrix}$$

where i_{n-1} is an $(n-1,1)$ vector of ones and I_{n-1} is an $(n-1, n-1)$ identity matrix.

Candelon *et al.* (2009) show that this multivariate test performs badly and has important size distortion when the number of individuals increases. So, they propose a block bootstrapped version of the test, instead of the asymptotic one, when the number of regions is sufficiently large (more than 5). However, the bootstrapped version can produce low power and the appearance of a trade-off between size and power that rises with the number of regions. In addition, the estimated sample value of covariance matrix V tends to be singular when n is big, even when the true covariance matrix is known to be non-singular. To solve this problem, we have used the shrinkage estimator proposed by Ledoit and Wolf (2003). The principle of shrinkage is that by properly combining two extreme estimators, one, the simple covariance matrix, unbiased but with a large estimation error, and the other, with structure and a relatively small estimation error, we obtain an estimator that performs better than either of the two extremes.

Figures 5 summarize all the synchronization measures, contingency, concordance and correlation. Instead of showing the three 18x18 matrices, we have calculated the regional averages of contingency, concordance and correlation and have displayed each in a graph together with the measure of each region with respect to Spain. In addition, we have included a multidimensional scaling map below the figures. This technique allows us to visualize similarities or dissimilarities and to produce a representation of the synchronization of the regions in a small number of dimensions. In the three cases, the corresponding synchronization index is used as the distance matrix; then, we transform the similarity matrix into a dissimilarity and reproduce its Euclidean distances[16]. A preliminary examination of the eigenvalues of this matrix shows that two dimensions are not enough to represent the points suitably, and we need at least 3 or 4 dimensions. Because of the impossibility of drawing graphs in four dimensions, we display them in three.

Different measures obtain very different ranges of values but nearly the same ranking of regions. We observe that the pair-wise correlations $\rho_{S_{ij}}$ (0.51, on average, for regions and 0.67 for Spain) are typically smaller than those obtained with the concordance index, which are around 0.76 and 0.84, suggesting that the stronger correlation between industrial regional cycles detected with I_{ij} is biased by the values of the mean. The pair-wise values obtained with the contingency analysis are also relatively big (0.61 and 0.76, respectively). Nonetheless, the evidence for rejecting the null hypothesis of no association is very strong between regions, around 88% on average when we consider the Harding-Pagan $t-test$ and 91% with the Pearson independence test (see Table I). With both tests, most non-rejections correspond to EXT, which is also the only region that is not synchronized with SP.

Contingency, concordance, and correlation are nearly identical with respect to their classification of regions[17]. We can see that PVAS, MAD and ARA are the regions most connected with the rest, while EXT, AND, CVAL and AST arethe most isolated. Notice that both groups of regions have atypical business cycle features. CVAL, AST and EXT are regions with low asymmetry between the duration of expansions and recessions, presenting brief

[16] A detailed explanation of this technique can been found in Timm (2002).

[17] In all cases the rank test of Spearman is over 0.95.

growth periods. On the contrary, the more synchronized regions are characterized by a great asymmetry in favour of expansions and a high cumulation of wealth gains during them. We also find another stylized fact: the degree of regional synchronization is directly related with the degree of synchronization with Spain.

The results of such a high correlation are confirmed by the bootstrap exercise for $\rho_{S_{ij}}$ because the zero is out of the confidence interval at 95% in all cases[18]. Nevertheless, as we can see in Figure 6, where we present the density of correlation coefficients, their variability is quite important in several regions. The correlation coefficient has also been estimated recursively throughout the sample to capture changes in synchronization over time. Figure 7 shows the evolution of the regional average of $\rho_{S_{ij}}$ and its value for each region and SP. The business cycle similarity of EXT and AND has gone down dramatically during the last decade both with respect to other regions and SP, indicating a lack of convergence. Other regions, such as CAT, NAV, MUR or CVAL also show a loss of synchronization during the long expansion phase (more pronounced in the last two) and seem to recover during the current recession. Finally, other regions, AST, ARA, MAD, PVAS and LAR show a remarkable stability.

In spite of the differences, we can conclude that the regional cycles are sufficiently correlated to explore the existence of some common cycles across Spanish regions. Firstly, we test different degrees of multivariate synchronization between the seventeen regions by using the statistic previously described. As we suspected, the test rejects all hypotheses from non-synchronization to perfect synchronization, passing through intermediate degrees (ρ =0.1,0.2,...,0.9), and the matrix is nearly singular even when there is a great intensity of shrinkage (see Table II) . However, if we apply the bootstraped version, we are not able to reject the null in any case, reflecting the loss of power.

As we can not obtain any conclusion from the multivariate synchronization test when we apply it for all the regions, we will explore the possibility of finding groups or clusters of regions with similar business cycle features. The natural way, using clustering techniques, does not seem the best way in this case. An eye-ball examination of the three multidimen-

[18] We do not include these results to save space.

sional scaling maps demostrates that, although some regions, such as EXT, CVAL, AND, MUR and AST, appear isolated from the rest, it is not easy to establish a clear pattern of clusterization. So, we will deal with the regional clustering using an alternative approach in Section 4 after trying to identify some variables that could explain synchronization.

3.3 Some clues to explain similarities or discrepancies among regions

Although there are a few regions that appear to be more isolated from the rest, we have shown that there is a certain degree of regional synchronization in terms of business cycle features. In this section, we will try to find the sources of the similarities as well as the discrepancies among regions. Disparities in regional business cycles have often been attributed either to idiosyncratic shocks or to differences in characteristics such as their sectorial composition. In the literature, there are some attempts to explain correlations across economies, countries or regions, and that, basically, use macroeconomic variables[19]. We have selected four representative macroeconomic and structural variables, namely, industrial weight, unemployment rate, per capita income level and human capital[20].

The results found in the study of the correlation between the cycle comovements across regions and the variables selected are presented in Table III. They show that all the selected variables are significantly related to synchronization, measured in terms of contingency, concordance and correlation. This relation is positive for industrial weight, per capita income and human capital and negative for unemployment. The coefficients are especially high for unemployment rates and higher for regional than for national synchronization[21].

[19] Clark and van Wincoop (2001) introduce the importance of growth rates and Bordo and Helbling (2003) try to measure the effect of the exchange rate regime; Camacho *et al.* (2006) and Owyang *et al.* (2005) use a wide set of variables.

[20] Average industrial weight over total output (1991-2008) and average unemployment rate (1991:III-2009:III) from the INE; average per capita income in PPP at current prices (1991-2008) from Funcas and average years of education for employees from Ivie (1991-2007).

[21] Camacho *et al.* (2006) find that the specialization of the economy explains differences across European and some other industrialised countries. Furthermore, Owyang *et al.* (2005) find that the share of manufacturing, as well as some human capital variables, are important to explain differences, in this case, between

Therefore, a region with a high unemployment rate, on average, is expected to be more isolated from the rest, while a region with high industrial weight and per capita income levels or where the labour force is more educated will be more synchronized. So, economic policies that focus on the improvement of these variables may enable a higher degree of regional convergence, measured through the level of synchronization.

In addition, we also test the influence of the geographical situation, namely, the border effect. Following Croux *et al.* (2001), this effect can be easily measured through the following ratio:

$$bi = \frac{\frac{1}{nb}\sum_{j=1}^{nb} \rho_{S_{ij}}}{\frac{i}{n}\sum_{j=1}^{n} \rho_{S_{ij}}},$$

where nb is the number of neighbouring regions and n the total number of regions. A ratio above 1 indicates that regions have more cyclical similarities with their neighbours than with the other regions. The non-parametric Wilcoxon-rank test is used to check the statistical significance of this effect.

The results show that, in general, neighbourhood matters, which is confirmed by the rejection of the null hypothesis of the Wilcoxon test at 1% of significance (Table IV). EXT, NAV, LAR and CANT are the regions with the highest border effects, while MUR, CLM and BAL present values under one[22]. Thus, for most regions, the political boundaries influence their degree of synchronization.

4. IS THERE A COMMON CYCLE? DYNAMIC FACTOR MODELS

In the previous section, we have only used categorical variables which describe the path of the cycle. These categorical variables are obtained by filtering the original data using non-parametric cycle dating techniques and finding the turning points that represent the business cycle. These techniques summarize all the characteristics of the series in a dichotomous

the cycle regimes.

[22] A similar approach is found in Croux *et al.* (2001) with a measure of dynamic correlation. They conclude that, for most US states and, to a lesser extent, for European countries, borders matter. Their ratios are higher than ours for the regions in both areas.

variable that represents the position of the unit (region or country) in the business cycle. Although it is very intuitive and useful, it loses some information which could be taken into account for comparing the behaviour of the different regions. So, in this section, we are going to use original series that contain different and complementary information.

As well as cycle dating approaches (parametric and non-parametric) to characterize the business cycle, there are other techniques that analyze comovements of economic variables using original data and extracting common factors[23]. One basic feature of the business cycle is, precisely, the presence of comovements across economic variables. Comovement measures constructed in the frequency domain, principal components and dynamic factor models are the main branches of this approach that deals with the original information. Of them, dynamic factor models (DFM henceforth) have recently emerged as a powerful tool to analyze shocks in large databases. The idea underlying DFM is simple: movements in a large number of economic series can be modelled through a small number of reference series or common factors. The DFM allow us to "let the data speak" without imposing a priori restrictions as in other approaches. We explore the concept of comovements in regional IPIs to assess to what extent the Spanish business cycle is shared by the regional ones[24].

In this section, we first carry out a preliminary analysis of comovements by using some measures in the frequency domain, such as coherence and cohesion. Second, we estimate the optimal number of factors, both static and dynamic, present in the seventeen Spanish regions and identify the national component. Then, we apply a DFM to test whether the seventeen Spanish regions move according to common driving forces. Finally, we subtract the common component associated with the Spanish cycle and form clusters using the idiosyncratic regional components.

[23] As Harding and Pagan (2002a) point out, these "can be thought of as a hybrid scheme" when standard dating methods are applied to the common factor obtained by using dynamic principal components (Forni *et al.*, 2000).

[24] There are also approaches in the literature that deal simultaneously with comomovements and business cycle dating. For a summary, see Camacho *et al.* (2010).

4.1 Some stylized comovements

In our preliminary analysis of comovements for regional output fluctuations, we first calculate the index used by Stock and Watson (2010) which summarizes the possible time-varying comovements among the regional IPIs and, then, compute spectral measures such as the coherence index and the modification proposed by Croux *et al.* (2001), which they call dynamic correlation. From here to the end of the section, we use first logarithmic differences of the regional IPI series y_{ti} with $i = 1...17$. They are equivalent to monthly growth rates, which conserve business cycle signals better than other alternatives such as interannual growth rates. Figure 8 shows the evolution of the regional industrial index (dashed lines) in levels (top figure) and growth ratios (middle figure). It also shows (solid lines) the median and 25% and 75% percentiles. Both in levels and in growth rates, there is a considerable dispersion of the regional business cycles, which increases dramatically after 1999, just after joining the EMU, when Spain underwent a long period of prosperity before the current crisis.

The measure proposed by Stock and Watson (2010) is based on Moran's spatial correlation index and captures the comovements over time across all regions through the rolling cross-correlation in differences. It has the following expression:

$$\widehat{I}_i = \frac{\sum_{j=1}^{N}\sum_{j=1}^{i-1} \widehat{cov(\Delta y_{it}, \Delta y_{jt})}/N(N-1)/2}{\sum_{i=1}^{N} \widehat{var(\Delta y_{it})}/N},$$

where,

$$\widehat{cov(\Delta y_{it}, \Delta y_{jt})} = \frac{1}{25}\sum_{s=t-12}^{t+10}(\Delta y_{is} - \overline{\Delta y_{it}})(\Delta y_{js} - \overline{\Delta y_{jt}}),$$

$$\widehat{var(\Delta y_{it})} = \frac{1}{25}\sum_{s=t-12}^{t+10}(\Delta y_{is} - \overline{\Delta y_{it}})^2,$$

$$\overline{\Delta y_{it}} = \frac{1}{25}\sum_{s=t-12}^{t+10}\Delta y_{is},$$

$$N = 17$$

The outcome, time series $\widehat{I}_t$, is plotted at the bottom of Figure 8. We observe that the synchronization of comovements is 0.49, on average, during all the period, reaching its minimum value (around 0.25) in 2007 and its maximum at the end of 2008 (more than 0.8)[25]. So, it seems there is an inverse relationship between economic growth and regional comovements. This result is very similar to that obtained with the correlation index from the BB cycle dating, whose regional average was 0.51. Furthermore, some regions showed a loss of synchronization during the expansion phase that began to recover in the current recession.

Spectral measures are very useful for analyzing comovements in time series and have several advantages over other binary concepts. Among them, coherence is the most popular measure in the literature. The coherence index between two processes, x_t and y_t is defined as:

$$C_{xy}(\lambda) = \frac{|S_{xy}(\lambda)|^2}{S_x(\lambda)S_y(\lambda)},$$

where S_{xy} is the cross-spectral density and S_x and S_y, the power spectral densities for each frequency λ. The results in Figure 9 show this measure between each region and Spain. In addition, in the bottom right-hand corner, we present the density functions of the regional coherence index for three selected frequencies, long-run (frequency 0), medium-run (3 years) and short-run (1 year). In general, most regions have a high coherence in the very long run, with the outstanding exceptions of EXT and CAN (the rest of the regions present very similar data at 0 frequency). However, comovements among regions decrease dramatically in the low frequencies, indicating that regional business cycles have great discrepancies in the short run. This conclusion is confirmed by the density functions, which demonstrate that the heterogeneity in comovements decreases with the frequency. Considering higher frequencies (from 0 frequency to 3 years), the regions that show the greatest comovements are CAT, PVAS, MAD, CANT and ARA (their synchronization measures showed that most of these regions are also the most connected with the rest). CVAL and CYL present more

[25] Notice that the sample is not able to evaluate the comovements in the 90s crisis with accuracy because the rolling procedure loses most of the observations in this period.

atypical patterns; the coherence begins to decrease at still high frequencies, then falls more sharply at about 9 months and increases again afterwards. At high frequencies, there also appear other regions with large values of coherence, but their comovements decrease more sharply when the frequencies begin to lower.

Although coherence is the most popular comovement measure in the time series literature, Croux *et al.* (2001) highlight that it is not appropriate as a comovement index and propose a modification, namely, dynamic correlation. Starting from the spectral decomposition of the processes x_t and y_t,

$$x_t = \int_{-\pi}^{\pi} e^{i\lambda t} dZ_x(\lambda),$$
$$y_t = \int_{-\pi}^{\pi} ie^{i\lambda t} dZ_y(\lambda),$$

where $dZ_x(\lambda)$ and $dZ_y(\lambda)$ are (complex) orthogonal increment processes, we can obtain the spectral and cross-spectral density functions as follows:

$$S_x(\lambda) = var(dZ_x(\lambda)),$$
$$S_y(\lambda) = var(dZ_x(\lambda)),$$
$$S_{xy}(\lambda) = var(dZ_x(\lambda), dZ_y(\lambda)),$$
$$S_{yx}(\lambda) = var(dZ_y(\lambda), dZ_x(\lambda)),$$

The dynamic correlation is

$$\rho_{xy}(\lambda) = \frac{c_{xy}}{S_x(\lambda)S_y(\lambda)},$$

where

$$c_{xy}(\lambda) = \frac{S_{xy}(\lambda)}{\sqrt{S_x(\lambda)S_y(\lambda)}},$$

is the coherence index, which is complex and, in general, non-symmetrical and $c_{xy}(\lambda)$ and $c_{yx}(\lambda)$ are conjugates. Nevertheless, coherence is real and symmetrical and, as Croux *et* *al.* (2001) point out, it does not measure dynamic correlation because it is invariant with respect to the lags of the processes. In other words, these authors demonstrate that the coherence between x_t and y_{t-k} is equal to that between x_t and y_t, with the simple example of two white noise processes. While coherence is equal to 1 at all frequencies, dynamic correlation ranges from 1, at frequency 0, to -1 at frequency π . Furthermore, dynamic correlation can be decomposed by frequency bands, which can be very useful for studying business cycle comovements. Finally, they propose a multivariate index of comovements, namely, cohesion, which has the following expression for a set of units X=[x_{1t},...x_{nt}]:

$$coh_X(\lambda) = \frac{\sum_{i \neq j} w_i w_j \rho_{xy}}{\sum_{i \neq j} w_i w_j}$$

The result of applying this cohesion index to the Spanish regions appears in Figure 10 for different frequencies[26]. We have also used this measure as a metric to construct a multivariate map at frequencies 0 and π. The conclusions are similar to those obtained with the coherence index. While regional comovements exhibit a high cohesion at frequency 0 (long-run), their dynamic correlation decreases greatly at lower frequencies. Indeed, an eye-ball examination of the maps suggests a strong grouping of regions at frequency 0, showing that their business cycles tend to converge in the long run. We can clearly distinguish two outliers, EXT and CLM, and two big clusters, one formed by CAN, AST, LAR and GAL and the other by the remaining regions. Nevertheless, the map corresponding to a low frequency, such as π, shows a high degree of dispersion and, consequently, indicates strong dissimilarities in short-run regional business cycles. This difference between the long and the short run has not been sufficiently emphasized in the literature, but is important to take this asymmetry into account when we analyze the comovements of business cycles. Although regions are very cohesive at frequency 0, with a value of 0.71, in other more typical business cycle frequencies of 1, 2 and 4 years (corresponding to $\pi/6$, $\pi/12, \pi/24$), the cohesion index has values of 0.67, 0.51 and 0.25, respectively[27].

[26] We have used the average GDP weight of each region over the total Spanish GDP (1991-2009) for weighting.

[27] A similar pattern is found by Croux *et al.* (2009) for the EMU countries, while the US regions and

4.2 Factor models

The seminal work of Geweke (1977) proposed a DFM as a time series extension of factor models previously developed for cross-sectional data. The main empirical finding that a few factors are able to explain a large fraction of the variance of many macroeconomic series was first reported by Sargent and Sims (1977), and confirmed afterwards by Giannone *et al.* (2004) and Watson (2004). The underlying idea of a DFM is that a few latent dynamic factors, f_t, drive the comovements of a high-dimensional vector of time series variables, X_t, which is also affected by a vector of mean-zero idiosyncratic disturbances, e_t. These idiosyncratic disturbances arise from measurement error and from special features that are specific to an individual series. The latent factors follow a time series process, which is commonly taken to be a vector autoregression (VAR)[28].

In this section, firstly, we estimate the minimum number of factors that explain the maximum variation of the regional industrial production index. Secondly, we identify the common factor as the national component and remove it. Finally, we look for idiosyncratic sources of variations that decrease when aggregating the data and construct clusters with them.

4.2a How many factors are there in the Spanish regions?.—

Bai and Ng (2002) developed a formal statistical procedure that consistently estimates the number of factors in a set of data with cross-sectional and temporal dimensions, N and T, respectively. Let Y_{it} be the observed data for region i in time t. The factor model then has the following expression:

$$Y_{it} = \lambda_i' F_t + \varepsilon_{it},$$

where F_t is a vector of r common factors, λ_i' is a vector of factor loadings associated with F_t, and ε_{it} is the idiosyncratic component of Y_{it}. The product $\lambda_i' F_t$ is called the common

states are more cohesive at 1.5 frequencies (even more than in the longer run).

[28] See Stock and Watson (2010) for a comprehensive survey of this literature.

component of X_{it}. For all the units and using matrix notation

$$Y_t = \Lambda F_t + \varepsilon_t$$

where the dimensions of Y_t, Λ, F_t, and ε_t are $Nx1$, Nxr, $rx1$ and $Nx1$, respectively. They propose estimating the common factors by minimizing the following expression:

$$V(k, F^k) = \min_{\Lambda} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \lambda_i k \widehat{F_t^k})^2$$

and consider the estimation of the number of factors r as a model selection problem, constructing different information criteria. When this procedure is applied to regional industrial index growth, all the criteria coincide in estimating 4 factors[29].

Having established the number of factors, we re-estimate the factor model with 4 common factors in order to obtain the idiosyncratic component of each region and we calculate the specific variance as $\Psi = \Sigma_y - \Lambda\Lambda'$. We have also included SP in this exercise to study its behaviour and compare it with the regional ones. Figure 11 shows these values. We observe that AND has the largest idiosyncratic component and BAL, AST, CAN and EXT also have large variances, while the lowest variances are found in PVAS, CAT, MAD and CYL. These results confirm those obtained in Section 3, where EXT, AND, AST and CVAL, in most cases, and the islands, in some cases, presented very peculiar behaviours, while PVAS and MAD were the regions most connected with the rest (CAT and CYL were closer to the first group than to the last two regions). However, the idiosyncratic factor of SP is nearly zero. This result is not surprising because, if we decompose the error term of each region into two components $\varepsilon_{it} = u_t + \xi_{it}$, one common and the other idiosyncratic, and we calculate their mean, we obtain $\epsilon_t = u_t + \sum_{i=1}^{N} \xi_{it}/T$ and, applying the Law of Large Numbers, the last term tends to zero. Furthermore, if we estimate the common factor, its correlation with Spanish data is 0.93. So, we confirm that the Spanish IPI is a good

[29] This result is obtained with a maximum value of 5 factors that explain 60% of the variation in the data. If we increase the maximum, we can find up to 18 factors. This result points to certain weaknesses in the estimation of the number of common factors.

representation of the aggregated regional ones and, consequently, suitably represents the common Spanish business cycle[30]. Nevertheless, although the Spanish business cycle is the result of aggregating regional ones, it does not mean that it is identical to any of the regional cycles, because within this common cycle there are seventeen different behaviours.

In the previous approach, we have only considered static common factors. But, as Bai and Ng (2007) point out, these may be dynamically related and, consequently, the spectrum of r has a reduced rank. The rank is actually q, the number of dynamic factors or primitive shocks. These authors propose a method for estimating the minimum number of primitive shocks based on the eigenvalues of the correlation or covariance matrix of a set of innovations of dimension rxr. They define two statistics as a sequential sum of the eigenvalues and estimate $\widehat{q}_3$ and $\widehat{q}_4$ as the minimum values that allow us to bound this sum. The value of q is estimated using the correlation matrix for r=4, which explain around 60% of the data variability, and we find that $\widehat{q}_3$ =1 and $\widehat{q}_4$=2. If we increase the number of static factors to r=12 (which explain 85% of the variation in the data), $\widehat{q}_3$ maintains 1 dynamic estimated factor and $\widehat{q}_4$ is now 3. When the covariance matrix is used, both tests estimate q=1 for r=1...7. Therefore, the results suggest that a common dynamic factor or primitive shock is the most plausible conclusion of our regional database and, consequently, that Spanish regions have a dynamic common factor that can be identified with the national component. In the next section, we will extract the common factor of the regions, which can be associated with Spain, and we will construct clusters with the residuals of the regional idiosyncratic factors.

4.2b Cluster analysis.—

Up to this point, we have documented a certain degree of synchronization in regional cycles but, also, a high heterogeneity. In this section, we focus on this idiosyncratic component and look for clusters in the Spanish business cycles. We follow the approach of Stock and Watson (2010) who suppose that regional variations are independent of the national

[30] This result seems very robust against other methodologies. Taking the average of regional cycle dating obtained previously with BB techniques, we find a correlation of 0.92 with respect to SP cycle dating.

one. So, we remove a single common factor associated with the national component to make the comovements more visible between regions, using this model:

$$
\begin{aligned}
Y_{it} &= \alpha_i + \lambda_i' F_t + \varepsilon_{it}, \\
F_t &= F_{t-1} + \eta_t \\
\varepsilon_{it} &= \sum_{j=1}^{k} \rho_{ij} \varepsilon_{it-j} + u_{it}
\end{aligned}
$$

where ε_{it} are the idiosyncratic terms or the contribution of regional factors to the total variation of the data and k=12 and $(\eta_t, \varepsilon_{it-j})$ are i.i.d. N(0,σ_η) and N(0,σ_ε), respectively. This model has been estimated by maximum likelihood, using least-squares estimates of the coefficients as starting values and the first principal component as an estimator of F_t. Then, we have removed the common factors and obtain the residuals $\varepsilon_{it} = Y_{it} - \alpha_i - \lambda_i' F_t + \varepsilon_{it}$, which are used in the cluster analysis.

The goal of the cluster analysis is to identify groups of regions. Regions in the same cluster will be more synchronized and share more similar business cycle features than regions in other groups. Two types of clustering methods have been used: the hierarchical and the partitioning algorithms. The first starts by forming a group for each individual. New items are then added employing some criterion of similarity, in our case minimizing the increase of the Euclidean square distance within clusters. The process goes on until all the individuals are in a single cluster. The sequence of clustering is displayed in a typical plot called a tree diagram or dendrogram where we can see the detailed process. Looking at these results, 3 or 4 clusters seems to be the most suitable decision. This method offers us a first approximation of the number of clusters present in our set of regional business cycles. In a second step, we apply a non-hierarchical clustering method called $k - means$ that requires previously deciding the number of groups. Furthermore, through the method of Bai and Ng (2002), 4 common factors have been indentified, which gives us a clue about the number of regional clusters.

The $k - means$ clustering creates a single level of clusters and assigns each region to a specific cluster. The algorithm finds a partition in which regions within each cluster are as close to each other as possible and as far from the regions in other clusters as possible. Each cluster is defined by its centroid, or centre, which is the point at which the sum of the distances from all the objects in the cluster is minimized. The iterative algorithm minimizes these distances within all the clusters, but its final results depend on the first random assignation. To overcome the two disadvantages of the $k-means$ method (the selection of the number of clusters and the dependence of initial partition), Stock and Watson (2010) propose a modified algorithm that repeats the procedure for multiple starting values and analyzes the value of the minimized objective function as a proxy of the most suitable number of clusters.

We apply this method to the idiosyncratic regional factors[31]. Increasing from 2 to 3 clusters reduces the value of the minimized objective function by approximately 10%, and increasing from 3 to 4 and 4 to 5 reduces the value by a 7 and 8%, respectively. After 5 clusters the value of the objective function is not reduced. Therefore, taking these results into account along with the findings from the hierarchical method and the optimal number of common factors, we could select from 3 to 5 clusters. The result with k=4 groups determines that AND, BAL and GAL are included in the first cluster, CLM, CVAL, EXT and MUR in the second, ARA, AST, CAN, CYL, NAV and LAR in the third and the rest of the regions, CANT, CAT, MAD and PVAS in the fourth. The results with k=3 basically join groups 1 and 2. The new first group is formed by AND, CLM, CVAL, EXT and MUR, the second by ARA, AST, BAL, CAN, CYL, GAL, NAV and LAR, and the third by CANT, CAT, MAD and PVAS. Finally, if we select k=5, we obtain the following clusters: group 1 AND, EXT, GAL; group 2 BAL, CLM; group 3 CVAL, MUR, LAR; group 4 ARA, AST, CAN, CANT, CYL, NAV and group 5 CAT, MAD and PVAS. Figure 12 illustrates these results.

Any of these 3 sets of groups fits the regional features identified in the previous sections

[31] We have used an adapted version, both for the DFM and the cluster analysis, of the original Gauss code of Stock and Watson (2010). The other procedures have been computed with Matlab. Codes are available upon request.

of this paper quite well. On the one hand, there are the regions most isolated from the rest and with the largest idiosyncratic components (more or less, groups 1 and 2 when k=4, group 1 when k=3 and groups 1, 2 and 3 when k=5) and, on the other hand, there are the most synchronized regions where the patterns are more stable throughout the sample (the other groups of the three distributions)[32].

Focusing on k=5 and analysing these groups in more depth, we can observe that group 1 is made up of the Spanish regions with the lowest per capita income levels and with a low presence of industry in their productive structure. The characteristics of group 2 are very similar to the previous group. In fact, and as we pointed out in Section 2, they form a group in which the industrial average is just below the national one. Group 3 brings together two regions that have presented atypical behaviours in the study (CVAL and, to a lesser extent, MUR) along with LAR, which seems to fit better in the large Group 4. This group includes regions with an important industrial weight (and CAN) and presents a high synchronization level. Finally, group 5 contains the most developed regions of the country with a high connection to the rest.

These results should be taken into account by policy-makers and academics when dealing with the Spanish business cycles. Policy-makers should take advantage of devolution and the degree of freedom that the Spanish fiscal system offers to design specific regional measures. Technicians that work in the field of cycle dating and that forecast recessions should be conscious of the differences across regions and use regional comovements to analyse Spanish economic developments.

5. CONCLUSIONS

We have carried out a comprehensive analysis of Spanish regional business cycles by using different approaches. In sum, this paper has found severe dissimilarities in regional business

[32] We have computed cohesion for each group of k=5. The results show that regions are more cohesive inside each cluster and, in general, throughout different frequencies. Cohesion clearly increases in the long run for all the groups but group 1 while, in the short run, the increase is not so clear in groups 4 and 5. Details are available upon request.

cycles, highlighting the importance of considering business cycles from a regional point of view. The most isolated regions are characterised by low asymmetries in their business cycle phases while more synchronized regions show high asymmetries in favour of expansions and cumulation during them. Evidence of an inverse relationship between economic growth and regional comovements is found, which hinders the convergence. As well as considering the previous results, it is necessary for economic policy design to take into account that the convergence process of the Spanish regions is a long run economic growth process more than a process related to fluctuations in the economy, which is confirmed by the strong asymmetries we find in the regional cohesion between the long and the short run.

On the whole, although there are many papers on intra-country linkages in an international context, this paper is the first to propose a systematic analysis of these linkages in Spain. Although policy-makers may not consider this analysis to be a previous step to the decision of whether or not to belong to more integrated economic areas, it is surely fundamental for reducing economic intra-country heterogeneity through the design of adequate economic policies. We do this for Spain, adding to its scarce regional business cycle literature. Given that the cycle of some regions is not similar to that of the country as a whole and that the Spanish path is not shared by the seventeen regional business cycles but is merely a consequence of aggregating them, carrying out economic policy measures at national level could bring about unwanted distortions in some regions and slow down their convergence processes, which would be further evidence of the need to apply specific economic measures. Macroeconomic stabilization policies, which are primarily related to the cyclical evolution of the economy, are very constrained in the European Union by the common monetary policy and the Stability and Growth Pact. Therefore, fiscal policy and devolution should be used to reduce regional disparities because, if their regional cyclical shapes are different, policy measures to fight recessions could be too accommodative for some regions and too tight for others.

REFERENCES

Airov, J. (1963). “The construction of interregional business cycle models”, *Journal of Regional Science*, 5, 1, 1-20.

Artis, M., H. M. Krolzig and J. Toro (2004). “The European business cycle”, *Oxford Economic Papers*, 56 (1), 1-44.

Bai, J., and S. Ng (2002), “Determining the Number of Factors in Approximate Factor Models”, *Econometrica*, 70, 191-221.

Bai, J., and S. Ng (2007), “Determining the Number of Primitive Shocks in Factor Models”, *Journal of Business and Economic Statistics,* 25, 52-60.

Bordo, M. D. and T. Helbling (2003). *Have national business cycles become more synchronized?*, NBER Working Paper No. W10130.

Borts, G. H. (1960). “Regional cycles of manufacturing employment in the United States, 1914-1953”, *Journal of the American Statistical Association*, 55 (289), 151-211.

Bry, G. and C. Boschan (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs.* New York, NBER.

Burns, A.F. and W. C. Mitchell (1946). *Measuring Business Cycles.* New York, NBER.

Camacho, M., G. Pérez-Quirós and P. Poncela (2010). *Green shoots in the Euro area. A real time measure.* Mimeo.

Camacho, M., G. Pérez-Quirós and L. Saiz (2006). "Are European business cycles close enough to be just one?", *Journal of Economic Dynamics and Control*, 30, 1687-1706.

Camacho, M., G. Pérez-Quirós and L. Saiz (2008). "Do European business cycles look like one?". *Journal of Economic Dynamics and Control,* 32, 2165-2190.

Cancelo, J.R. (2004). “El ciclo del empleo en España. Un análisis regional”, *Revista económica de Castilla La Mancha*, 4, 107-138.

Cancelo, J. R. and P. Uriz (2003). *Cycles in Spanish regional employment*, Seminar in Fundear.

Candelon, B., Piplack, J. and S. Straetmans (2009). "Multivariate Business Cycle Syncrhonization in Small Samples", *Oxford Bulleting of Economics and Statistics*, 71(5), 715-

737.

Carlino, G. and R. DeFina (1995). “Regional income dynamics”, *Journal of Urban Economics*, 37, 88-106.

Carlino, G. and R. DeFina (1998). *The differential regional effects of monetary policy: evidence from the U. S. States*, Federal Reserve Bank of Philadelphia Working Paper 3291.

Carlino, G. and K. Sill (1997). “Regional economies: separating trends from cycles”, *Federal Reserve Bank of Philadelphia: Business Review*, 1-13.

Carlino, G. and K. Sill (2001). “Regional income fluctuations: common trends and common cycles”, *The Review of Economics and Statistics*, 83(3), 46-456.

Clark, T. O. and E. van Wincoop (2001). "Borders and business cycles", *Journal of International Economics*, 55(1), 59-85.

Croux, C., M. Forni and L. Reichlin (2001). "A measure of comovement for economic variables: theory and empirics", *The Review of Economics and Statistics*, 83(2), 232-241.

Dolado, J. J. and R. María-Dolores (2001). “An empirical study of the cyclical effects of monetary policy in Spain (1977-1997)”, *Investigaciones Económicas*, XXV (1), 3-30.

Dolado, J. J., M. Sebastián and M. Vallés (1993). “Cyclical patterns of the Spanish economy”, *Investigaciones Económicas*I, XVII (3), 445-473.

Doménech, R. and V. Gómez (2005). “Ciclo económico y desempleo estructural en la economía española”, *Investigaciones Económicas*, XXIX (2), 259-288.

Forni, M. Haillin, M. Lippi, M. and L. Reichlin (2004). "The generalized dynamic-factor model: identification and estimation", *The review of Econmics and Satistics* 82 (4), 540-554.

Gadea, M. D., Gómez-Loscos, A. and A. Montañés (2006). *How many regional business cycles are there in Spain? A MS-VAR approach*, Documento de trabajo Fundear (Serie Investigación), 27.

Garrison, C. B. and H. S. Chang (1979). “The effect of monetary and fiscal policies on regional business cycles”, *International Regional Science Review*, 4(2), 167-180.

Geweke, J. (1977). “The Dynamic Factor Analysis of Economic Time Series,” in *Latent Variables in Socio-Economic Models*, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam: North-Holland.

Giannone, D., L. Reichlin and L. Sala (2004). “Monetary Policy in Real Time,” *NBER Macroeconomics Annual*, 161-200.

Hamilton, J. D. (1989). “A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica*, 57, 357-384.

Hamilton, J. D. (2001). “Comment on "A comparison of two business cycle dating methods"”, Mimeo.

Hamilton, J. D. and M. T. Owyang (2009). "The propagation of regional reccesions", Mimeo.

Harding, D. and A. Pagan (2002a). "Dissecting the cycle: a methodological investigation", *Journal of Monetary Economics*, 49(2), 365-381.

Harding, D. and A. Pagan (2002b). "A comparison of two business cycle dating methods", *Journal of Economic Dynamics and Control*, 27, pages 1681-1690.

Harding, D. and A. Pagan (2003). "Rejoinder to James Hamilton", *Journal of Economic Dynamics and Control*, 27, pages 1695-1698.

Harding D. and K.D. Pagan (2006). "Synchronization of cycles", *Journal of Econometrics*, 132(1), 59-79

Kouparitsas, M. A. (2002). “Understanding US regional cyclical comovement: how important are spillovers and common shocks?”, *Federal Reserve Bank of Chicago Economic Perspectives*, 24 (4), 30-41.

Kozlowski, P. J. (1995). “Money and interest rates as predictors of regional economic activity”, *The Review of Regional Studies*, 25(2), 143-157.Krolzig, H. M. (1997). *International business cycles: regime shifts in the stochastic process of economic growth*, Applied Economics Discussion Paper 194, Institute of Economics and Statistics, Oxford.

Krolzig, H. M. and J. Toro (2005). “Classical and modern business cycle measurement: The European case”, *Spanish Economic Review*, 7, 1-21.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992). “Testing the Null Hypothesis 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.

Ledoit, O. and M. Wolf (2003). "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection", *Journal of Empirical Finance*, 10(5), 603-621).

Metzler, L. A. (1950). “A multiple-region theory of income and trade”, *Econometrica*, 18 (4), 329-354.

Ng S. and P. Perron (2001). “Lag length selection and the construction of unit root tests with good size and power”, *Econometrica*, 69, 1519-1554.

Owyang, M. T., Piger, J. and H. J. Wall (2005). “Business Cycle Phases in U.S.States”, *Review of Economics and Statistics*, 87(4), 604-616.

Phillips, P. C.B. and P. Perron (1988). "Testing for a unit root in time series regression", *Biometrika*, 75, 335-346.

Politis, D. N. and J. Romano (1994). "The stationary bootstrap", *Journal of the American Statistical Associatio*n, 89, 1303-1313.

Rodríguez, G. and J. Villemaire (2004). “Identifying and measuring Canadian regional business cycles using MS-VAR models”, Mimeo, University of Ottawa.

Sargent, T.J. and C.A. Sims (1977). “Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory,” in *New Methods in Business Cycle Research*, ed. by C. Sims *et al.*, Minneapolis: Federal Reserve Bank of Minneapolis.

Schumpeter, J. A. (1939). *Business Cycles: A theoretical, historical and statistical analysis of the Capitalist process*, New Edition Porcupine Pr (November 1989).

Stock, J. H. and M. Watson (2010, forthcoming). "Dynamic Factor Models", in *Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements and D. F. Hendry, Oxford University Press.

Timm, N. H. (2002). *Applied Multivariate Analysis*, Springer Text in Statistics, Springer-Verlag, New York.

Watson, M. (1994). "Business Cycle Durations and Postwar Stabilization of the U.S. Economy", *American Economic Review*, 84(1), 24-46.

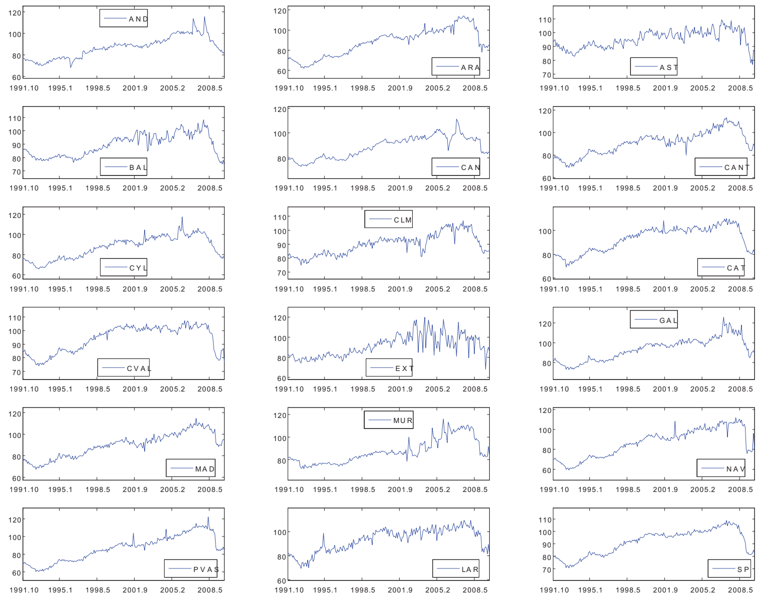


FIG 1. Seasonally adjusted Spanish and regional industrial production indexes (levels).

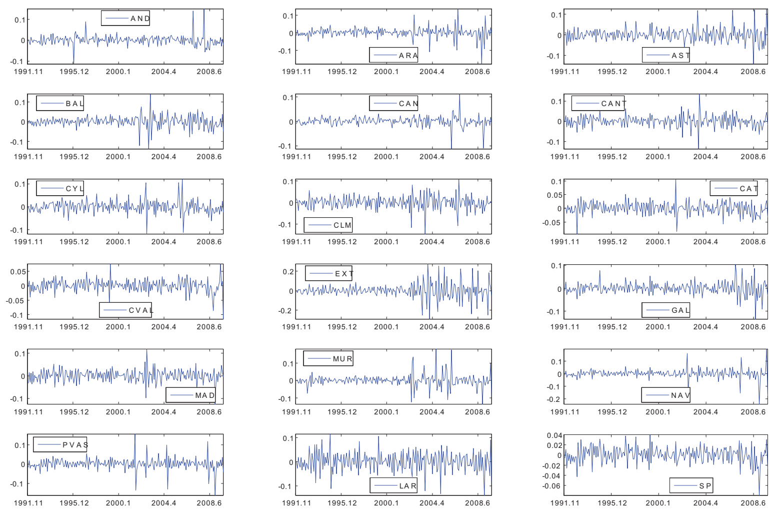


FIG 2. Seasonally adjusted Spanish and regional industrial production indexes (growth rates).

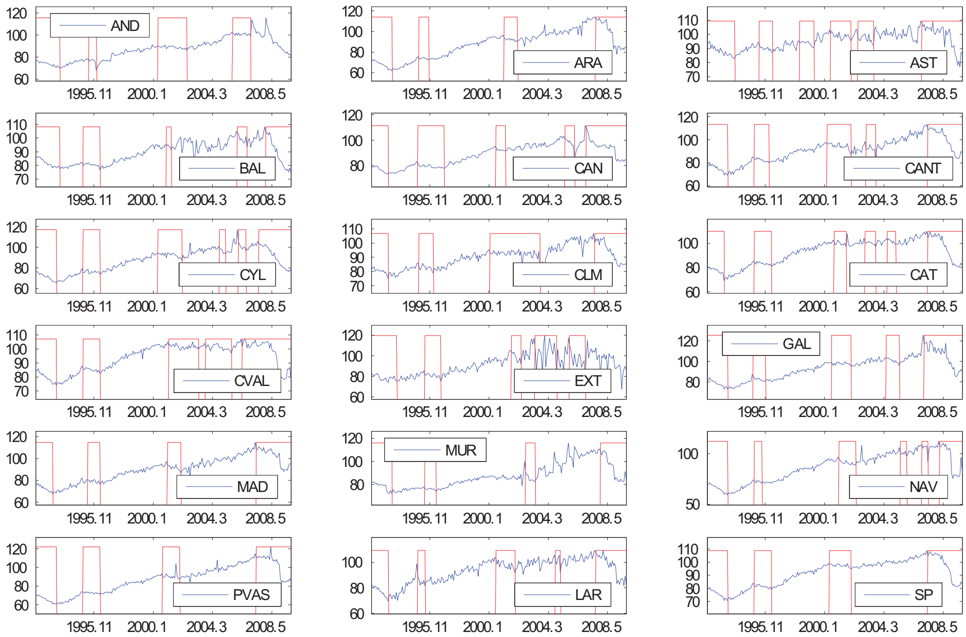


Fig 3. Bry-Boschan cycle dating.

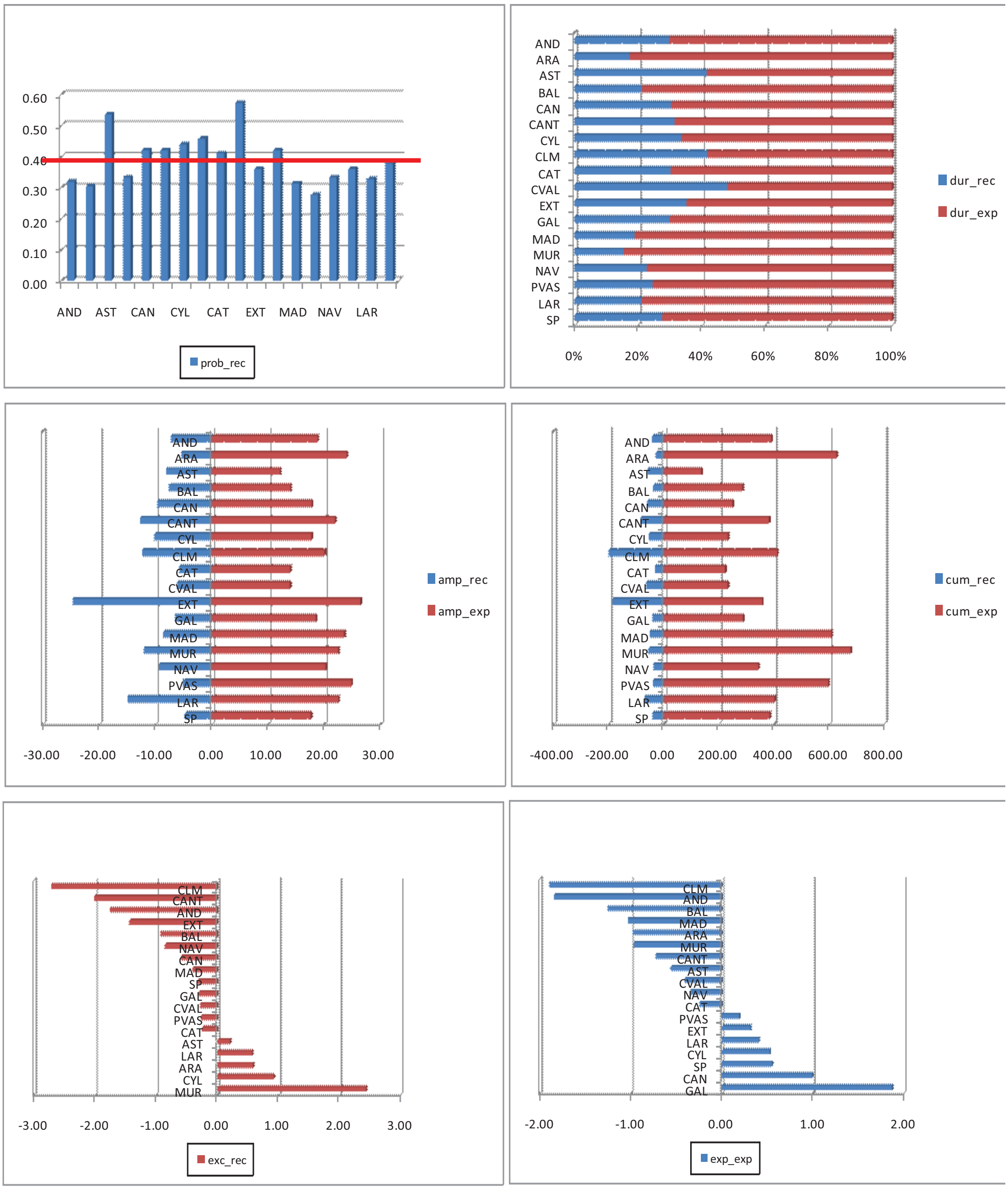
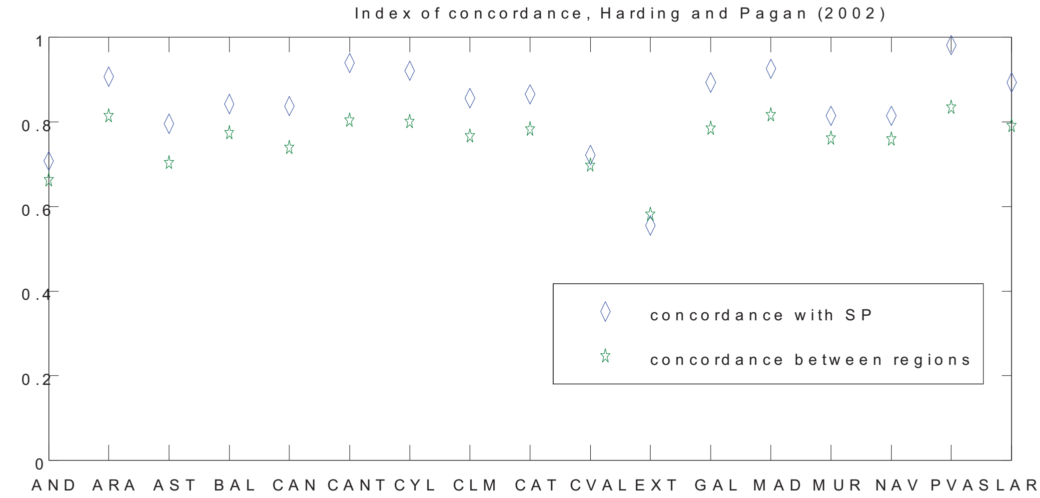


FIG 4. Dissecting the cycles of Spain and its regions.



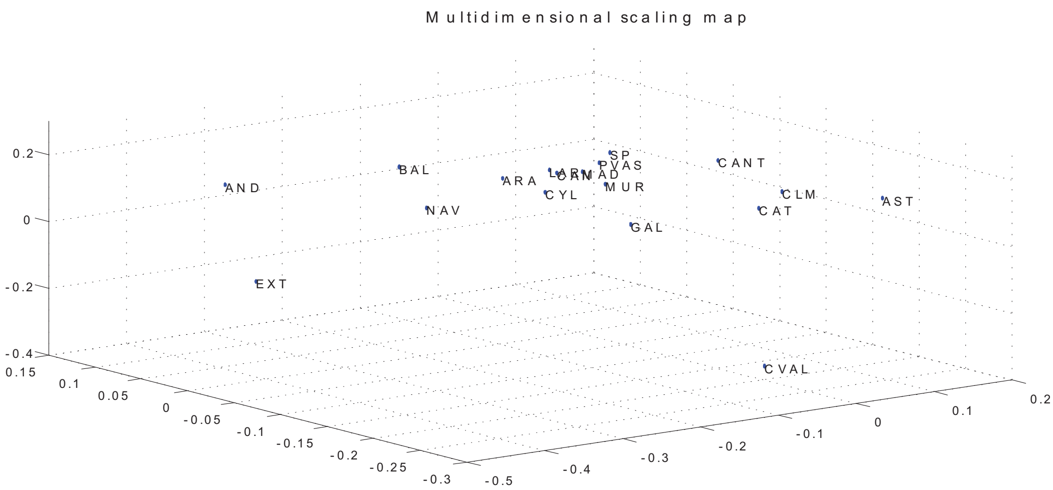


FIG 5A. Synchronization measures.

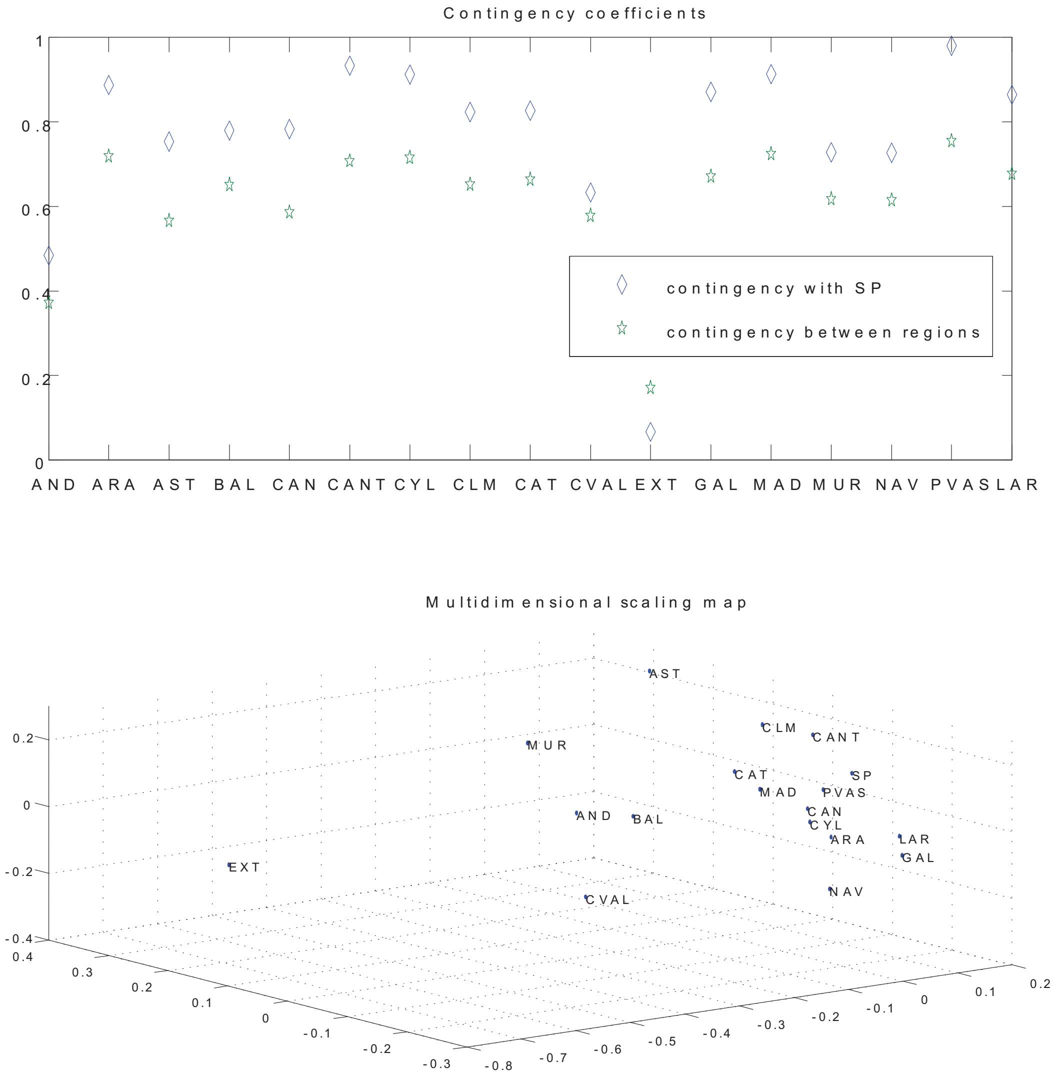
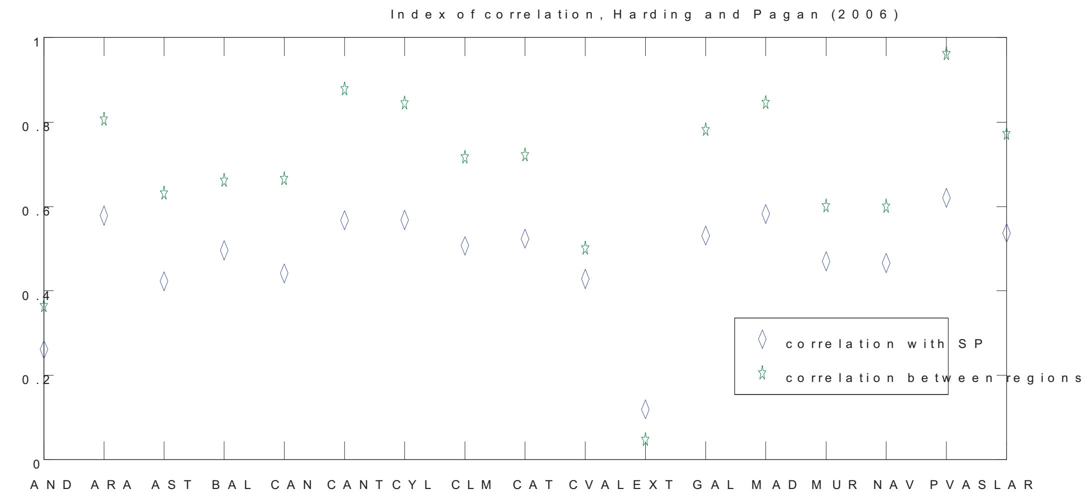


Fig 5b. Synchronization measures.



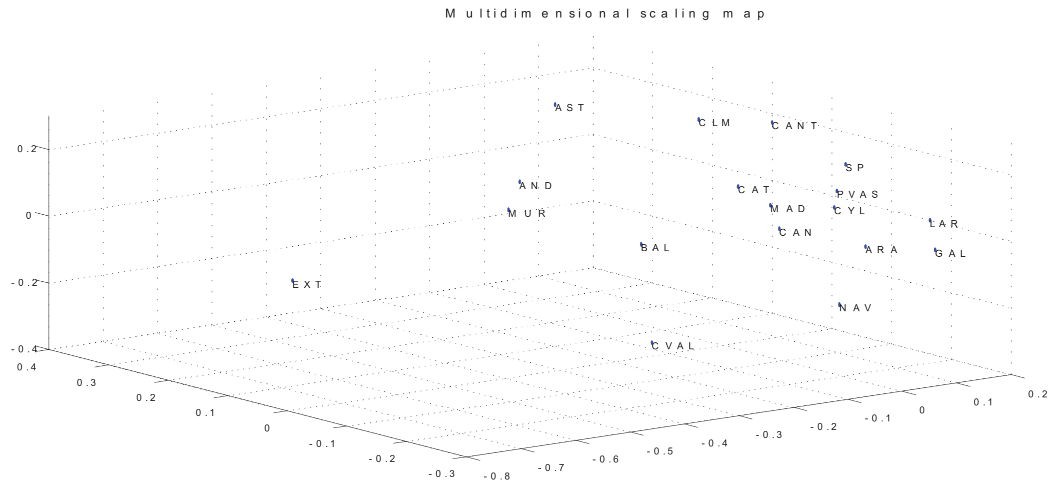


FIG 5C. Synchronization measures.

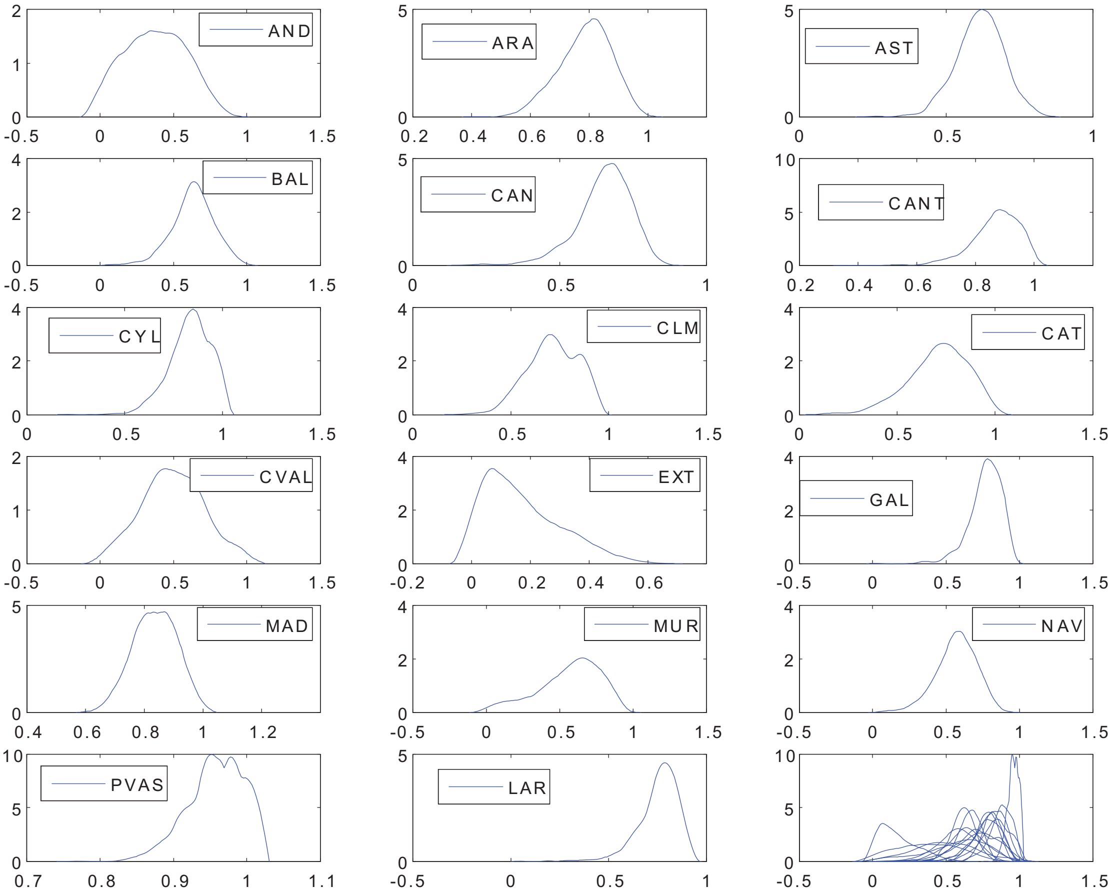


Fig 6. Density of correlation coefficients.

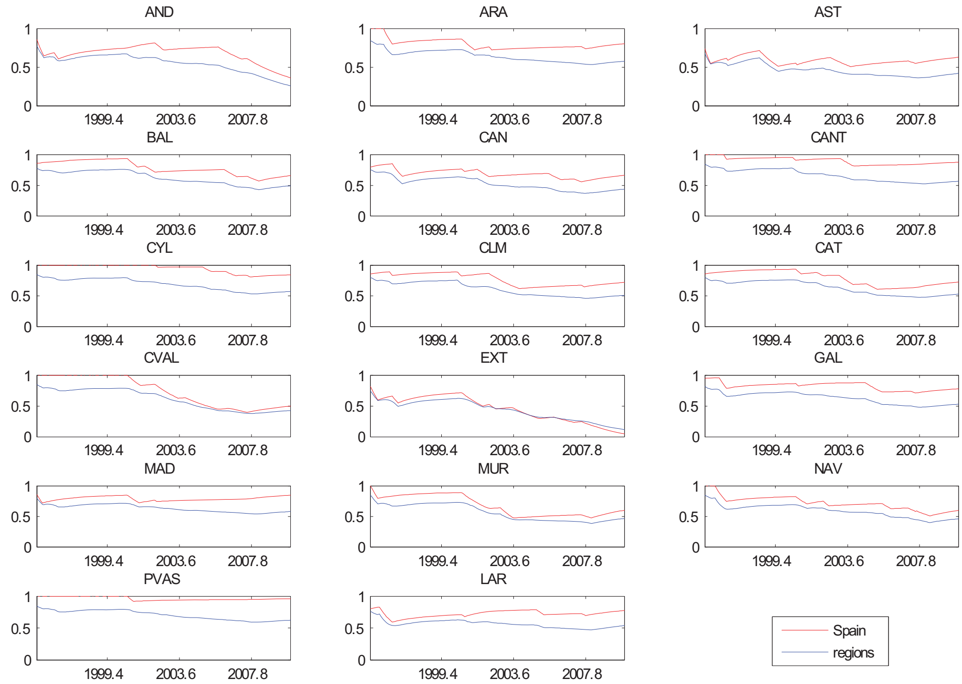


FIG 7. Recursive correlation coefficients.

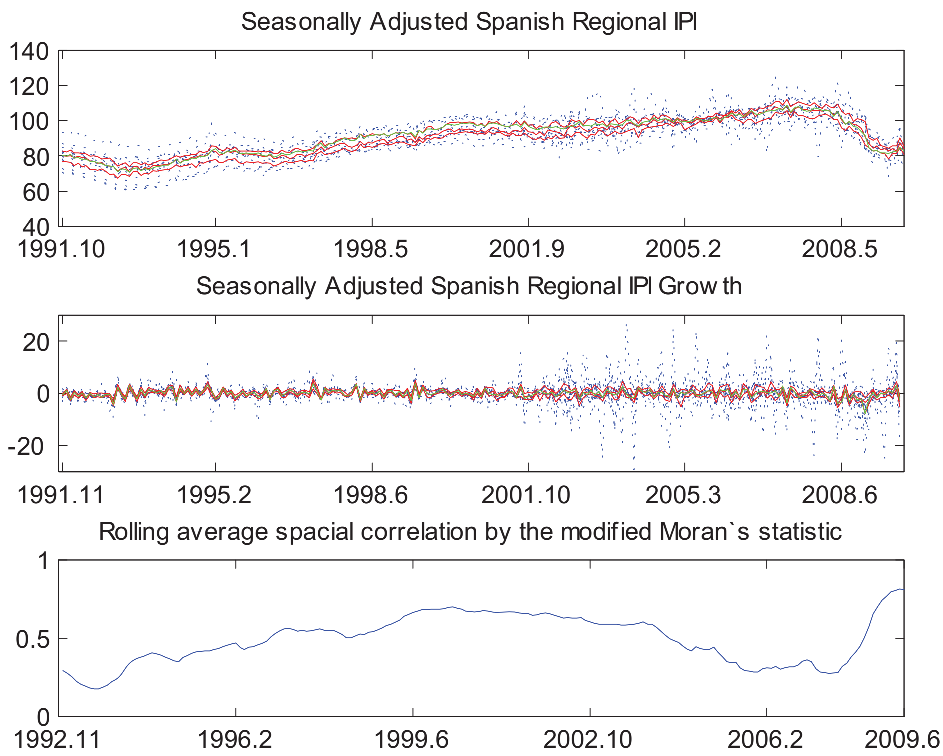


Fig 8. Comovements.

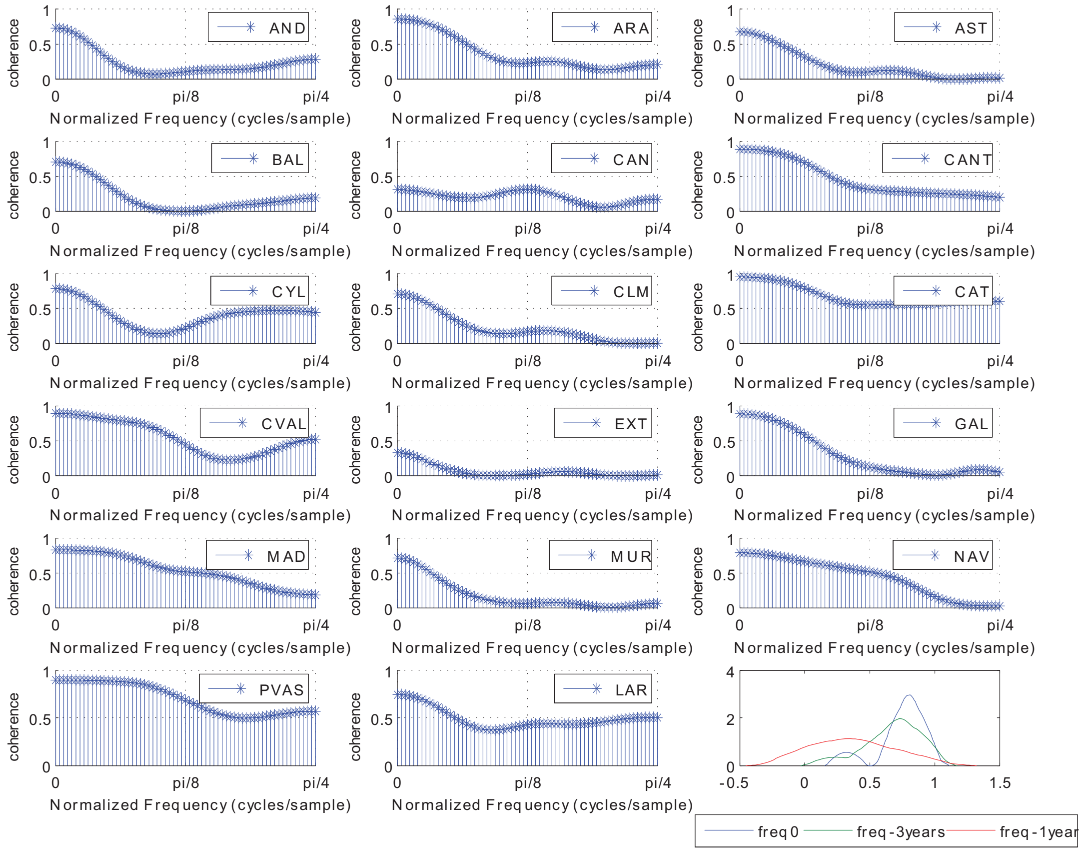


Fig 9. Coherence.

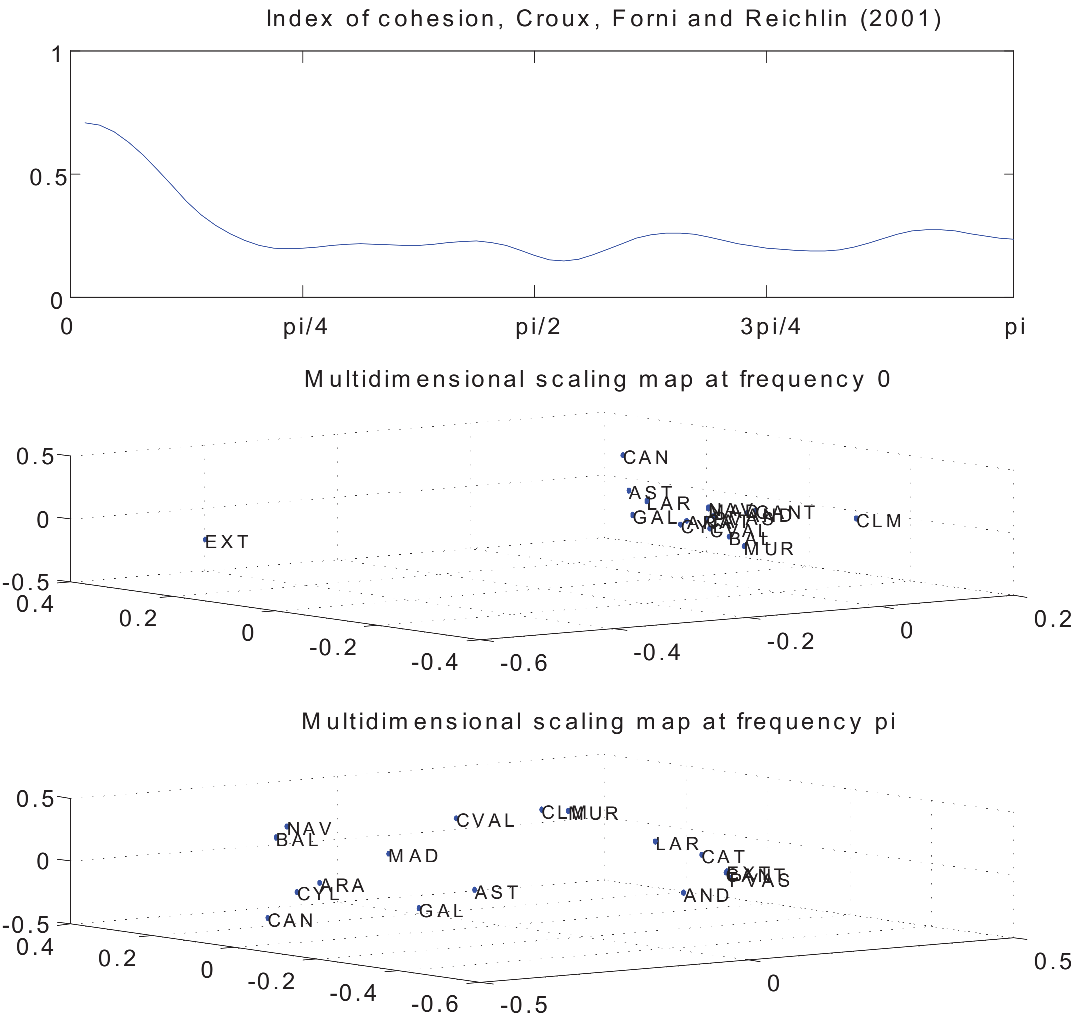


FIG 10. Cohesion.

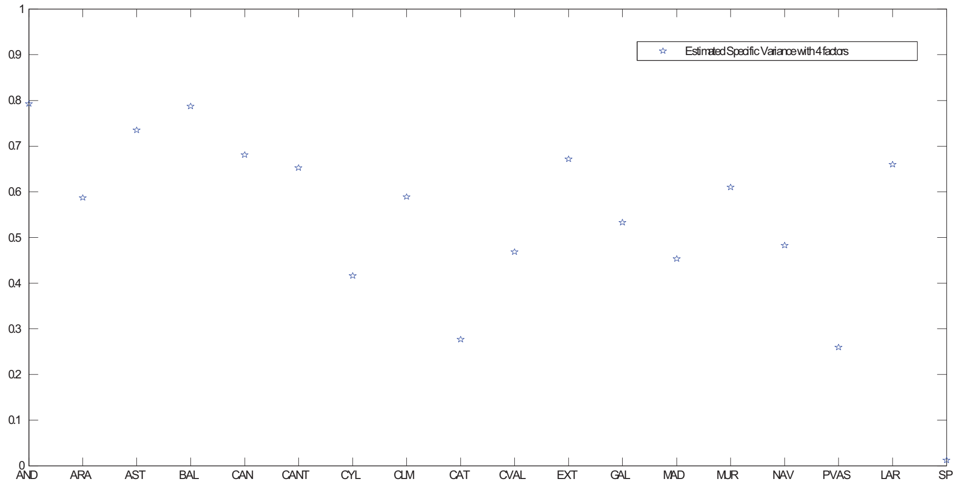


FIG 11. Regional specific variance with 4 factors.

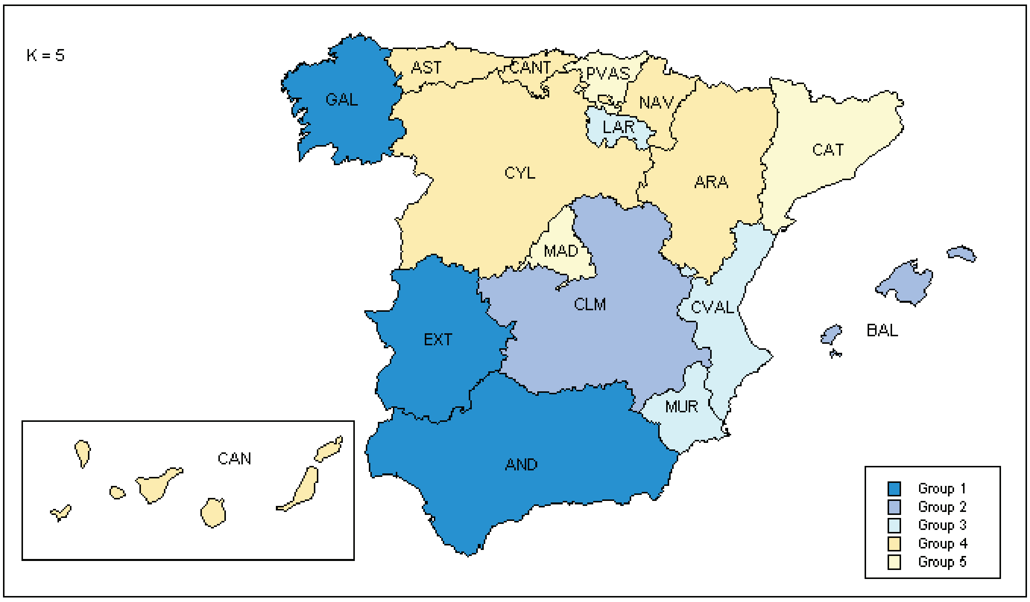


FIG 12. Cluster map

TABLE I

T-RATIO FOR ρ AND INDEPENDENCE TEST

	AND	ARA	AST	BAL	CAN	CANT	CYL	CLM	CAT	CVAL	EXT	GAL	MAD	MUR	NAV	PVAS	LAR	SP
AND	-	2.46	2.08	3.70	1.66	2.76	4.51	2.83	0.84	1.80	5.32	2.00	2.32	1.11	3.14	2.81	1.93	3.31
ARA	16.74	-	5.57	6.87	5.12	8.79	9.13	7.50	6.47	7.11	0.78	8.87	11.57	6.13	9.19	11.77	10.06	10.38
AST	10.26	52.91	-	4.39	4.32	8.19	4.88	6.79	5.98	3.41	-0.33	5.09	10.24	10.38	4.17	8.83	5.29	8.86
BAL	34.59	88.36	34.64	-	5.81	5.71	8.52	4.25	5.60	4.79	2.62	4.94	6.72	7.88	5.20	8.07	5.17	7.31
CAN	6.97	56.80	37.40	65.41	-	5.96	5.77	4.34	6.21	3.39	0.20	6.04	7.60	5.25	5.86	8.64	6.34	8.55
CANT	19.47	117.24	83.83	65.41	63.98	-	9.67	13.70	11.59	4.44	0.38	8.87	13.54	8.48	5.93	17.81	10.63	20.63
CYL	44.35	121.05	43.47	105.57	60.33	111.16	-	7.30	8.42	6.11	1.60	10.89	13.64	5.96	5.93	16.52	16.60	18.14
CLM	20.28	94.26	71.31	40.62	41.49	150.05	84.73	-	8.91	4.94	0.22	5.82	10.38	8.31	5.18	9.92	9.09	10.74
CAT	1.83	80.00	56.88	64.25	68.23	128.60	94.24	100.92	-	6.10	0.51	9.50	10.70	8.82	4.12	11.62	10.53	9.89
CVAL	7.68	70.51	25.80	40.00	24.49	40.23	62.26	48.30	60.87	-	2.56	7.17	8.95	8.41	7.37	7.00	5.32	5.67
EXT	62.80	1.64	0.29	17.70	0.11	0.38	6.16	0.13	0.68	14.35	-	0.03	1.01	1.38	0.75	0.69	-0.71	0.43
GAL	10.44	117.23	48.23	51.99	68.52	104.69	117.09	61.22	104.45	73.48	0.00	-	9.96	4.52	7.14	12.99	19.67	12.50
MAD	14.88	140.14	80.21	83.10	81.10	116.33	113.48	93.20	108.41	68.64	2.76	97.92	-	7.23	6.41	14.30	7.99	12.06
MUR	3.61	77.34	71.62	93.46	48.87	78.09	56.73	70.30	81.65	56.91	5.38	40.65	90.66	-	3.95	6.20	3.80	5.82
NAV	24.53	120.27	31.32	58.59	60.76	60.76	58.64	48.34	35.55	60.59	1.44	75.21	77.54	41.54	-	6.92	6.46	6.33
PVAS	21.00	152.58	83.22	104.39	95.92	153.17	141.59	100.43	125.09	65.37	1.28	132.61	167.59	86.07	86.78	-	11.81	26.09
LAR	10.28	130.33	44.14	55.90	67.89	105.95	128.02	89.04	104.55	42.44	1.20	145.28	103.68	38.87	75.79	126.94	-	10.21
SP	28.67	140.51	85.91	94.40	95.71	166.58	154.01	110.85	112.37	54.04	0.49	131.95	154.58	78.05	77.85	199.51	128.96	-

HAC t-Student for correlation coefficient in the upper triangle and χ^2contingency test in the lower triangle.

TABLE II

Multivariate Synchronization test

	test	asymptotic critical value	bootstrap critical value
SPPS	3671.99	164.22	1.4255
SMS			
$\rho = 0.1$	2959.57	164.22	$1.1465x10^5$
$\rho = 0.2$	2363.13	164.22	$0.8984x10^5$
$\rho = 0.3$	1882.67	164.22	$0.7131x10^5$
$\rho = 0.4$	1518.19	164.22	$0.5264x10^5$
$\rho = 0.5$	1269.70	164.22	$0.3760x10^5$
$\rho = 0.6$	1137.19	164.22	$0.2830x10^5$
$\rho = 0.7$	1120.66	164.22	$0.2256x10^5$
$\rho = 0.8$	1120.11	164.22	$0.2493x10^5$
$\rho = 0.9$	1435.55	164.22	$0.2933x10^5$
PS	234.66	27.59	721.4989

TABLE III

Correlations with Structural Variables

	industrial weight	unemployment ratio	per capita income	human capital	border
concord_reg	0.48 (0.059)	−0.79 (0.000)	0.68 (0.004)	0.60 (0.014)	−0.60 (0.013)
concord_sp	0.47 (0.064)	−0.72 (0.001)	0.62 (0.010)	0.60 (0.014)	−0.57 (0.021)
conting_reg	0.51 (0.045)	−0.83 (0.000)	0.65 (0.006)	0.68 (0.019)	−0.68 (0.003)
conting_sp	0.50 (0.049)	−0.78 (0.000)	0.59 (0.016)	0.56 (0.022)	−0.69 (0.003)
corre_reg	0.51 (0.043)	−0.82 (0.000)	0.66 (0.005)	0.60 (0.014)	−0.63 (0.008)
corr_sp	0.49 (0.054)	−0.74 (0.001)	0.61 (0.001)	0.60 (0.013)	−0.59 (0.015)

p_values in brackets.

TABLE IV

Do Borders Matter?

	ratio	neighbour regions
AND	1.17	3
ARA	1.11	6
AST	1.15	3
BAL	0.93	2
CAN	–	0
CANT	1.21	3
CYL	1.01	9
CLM	0.88	7
CAT	1.01	3
CVAL	1.11	5
EXT	1.93	3
GAL	1.07	2
MAD	1.12	2
MUR	0.81	3
NAV	1.33	3
PVAS	1.16	4
LAR	1.27	4
Wilcoxon test	21	

The critical value of the Wilcoxon test at 1% is 23.