

## Cycles inside cycles: Spanish regional aggregation

Maria Dolores Gadea · Ana Gómez-Loscos ·  
Antonio Montañés

Received: 24 June 2010 / Accepted: 21 May 2011 / Published online: 10 June 2011  
© The Author(s) 2011. This article is published with open access at SpringerLink.com

**Abstract** This paper sets out a comprehensive framework to identify regional business cycles within Spain and analyses their stylised features and the degree of synchronisation both within them and between them and the Spanish economy. We show that the regional cycles are quite heterogeneous although they display some degree of synchronisation. We also propose a dynamic factor model to cluster the regional comovements. We find that the Spanish business cycle is not the same as those of the 17 regions, but is the sum of the different regional behaviours. Clusters with a high industrial weight, per capita income and human capital and a low unemployment rate are also more synchronized. The implications derived from our results are useful both for policy makers and analysts.

**Keywords** Business cycle · Synchronisation 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.

---

M. D. Gadea (✉) · A. Montañés  
Department of Applied Economics, Universidad de Zaragoza,  
Gran Vía 2, 50005 Saragossa, Spain  
e-mail: lgadea@unizar.es

A. Gómez-Loscos  
Fundear and Universidad de Zaragoza, Saragossa, Spain

Aggregative analysis ... not only does not tell the whole tale but necessarily obliterates the main... point of the tale  
(Schumpeter 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 analyse their synchronisation. 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 the same as 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 synchronised or coordinated with the rest, to provide empirical evidence for the existence of different common regional business cycles and to analyse their synchronisation with the Spanish aggregate cycle. This approach makes sense in industrialized countries where lower aggregation levels are significant (regions or counties) and which have federal fiscal systems that allow differential economic policies to be implemented. We focus on Spain, a country divided into 17 NUTS-2 regions with a high degree of fiscal federalism, and we test if the Spanish cycle is really unique and its path is the same in the 17 regional business cycles. If not, a centralized economic policy may not be the most suitable option 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 geographical level mostly because of the absence of adequate data.<sup>1</sup> Nevertheless, recently, two papers have studied similarities and differences across US states during the different phases of their business cycles. The

---

<sup>1</sup> The study of cyclical patterns for countries has a long tradition, as will be detailed in Sect. 3.1.

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.<sup>2</sup> 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.<sup>3</sup>

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 indicator of economic activity than the industrial production index ([Dolado et al. 1993](#); [Dolado and María-Dolores 2001](#) and [Doménech and Gómez 2005](#)).<sup>4</sup> The only paper that tries to characterise Spanish regional business cycles is [Cancelo \(2004\)](#). It uses employment data and analyse 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 synchronisation using different measures that consider the degree of comovements between each region and the others and with Spain as a whole.

In the second block, we first carry out a preliminary analysis of time-varying comovements in the original series through an Index proposed by [Stock and Watson \(2010\)](#). Secondly, we investigate whether some common factors could be driving the regional business cycles. Then, we compare these results with the Spanish cycle and identify the national component. After identifying the number of static and primitive factors, we estimate a Dynamic Factor Model for the regions. Based on the results obtained, we carry out the cluster analysis using the idiosyncratic regional components.

We look for an economic characterisation of the clusters obtained relying, on the one hand, on the key cycle features and the degree of synchronisation found in the first block and, on the other hand, on the role of some macroeconomic variables. These

---

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

<sup>3</sup> 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.

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

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 synchronisation. We also consider another dimension that may influence the explanation of the similarities and differences in regional comovements, namely, neighbourhood.

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 synchronisation, 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. One outstanding result is that there is an inverse relationship between economic growth and regional comovements or synchronisation. Furthermore, there are no signs of convergence over time in most regions. Secondly, we find a common dynamic factor, which represents the Spanish cycle. After removing this factor, we use regional idiosyncratic components to cluster regions and we confirm the idea that the Spanish business cycle is a result of aggregating regional ones that do not have the same patterns. The clusters with an important industrial weight, per capita income and human capital and a low unemployment rate also show a high level of synchronisation. These findings 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 characterise them as well as their mutual synchronisation and their synchronisation with Spain. 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, which are characterised in Sect. 5. Finally, Sect. 6 concludes.

## 2 Data and stylized facts

We consider the 17 Spanish Autonomous Communities that correspond to NUTS-2 in the EUROSTAT nomenclature. Each region is denoted by an acronym: Andalucía (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 2010:12, 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 2010:12. Thereby, we obtain 231 observations for each region and for Spain as a whole.

This is the first time the IPI has been used to measure regional business cycles in Spain and it could be controversial but we cannot 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 production or employment series, respectively. By using the more timely and monthly IPI series, we hope to find early signals of turning points of economic activity which is essential for policy makers and analysts to enable timely analysis of the current and short term economic framework and so, to design the most suitable economic policy.

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 analysed, we can distinguish four main phases in the Spanish economic cycle<sup>5</sup>: 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 which, however, shows some signals of recovery.<sup>6</sup>

These phases are similar to the ones obtained by [Cancelo \(2004\)](#) using quarterly employment data for the 1977–2003 sample. He finds four phases: the first one (1977–1984) is characterised by a high heterogeneity in the regional cycles; the second (1985–1991) captures the wide expansion of the 80s; the third shows the effect of the 1992–1993 crisis and the fourth (1994–2003) reflects an expansion.

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 2009, the last year with available regional accounts data, 71.3% of the total Spanish GDP is generated by the services sector as in the most developed economies; 12.4% is industrial, 10.8% comes from construction and the remaining 2.9 and 2.6% belong to the energy and agriculture sectors. The biggest changes in the productive structure appear in the tertiary sector, which has increased 9.4 percentage points with respect to the figures of 1991, and industry, which is 6.8 pp below.

The most industrialised regions in the whole period (those with an industrial weight clearly above the Spanish average, which is 16.9%) 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 CAN, BAL, EXT, AND, MAD and MUR, although the distance between the first two and the last two is of almost 10 pp.

### 3 Business cycle dating and synchronisation

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 iden-

<sup>5</sup> Although our data source begins in October 1991, we capture the end of the expansion of the 80s.

<sup>6</sup> Taking quarterly data for both Spanish GDP and IPI from 1991.4 onwards, the correlation between the two variables is really high: 0.74 with annual growth rates and 0.70 with quarterly growth rates.

tify regional business cycles and we illustrate the key features that describe these business cycles. Secondly, we examine the degree of similarity among the business cycles identified in Spain and its regions.

### 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 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, the 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 evolved 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.<sup>7</sup> 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. The MS models try to characterise 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 expansion and contraction. 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 synchronisation 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).

---

<sup>7</sup> See [Harding and Pagan \(2002b, 2003\)](#) and [Hamilton \(2001\)](#) for a debate about the two business cycle dating methods.

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 2010:12 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 appropriate for short samples and for a small number of changes in regime transitions.<sup>8</sup>

Before applying the BB methodology, we need to transform the original series properly.<sup>9</sup> As is well known, IPI series are characterised by a pronounced seasonal component, some outliers and a remarkable volatility that increases at the regional level. So, we have opted to remove them and use the cycle-trend component of the original series to carry out the dating of the regional business cycles, following the most habitual data transformations used in papers devoted to cycle-dating.<sup>10</sup> To do so, we have employed the TRAMO-SEATS method with options that include automatic model selection, seasonal adjustment and correction for outliers (additive and transitory), the trading effect day and the Easter effect.<sup>11</sup> Figure 1 shows the original data and the cycle-trend series for all the regions and Spain.<sup>12</sup>

In addition, we have taken some cautions about very short cycles and very flat phases. The original algorithm considers a minimum phase-length criterion of 5 months and cycle-length criterion of 15 months, but the amplitude of expansions and recessions is not taken into account. However, other works (Artis et al. 1997; Mönch and Uhlg 2005; McKay and Reis 2008) have included additional restrictions in accordance with the specific features of the data that can differ from the established by the NBER for US. So, taking into account the characteristics of our data, we have incorporated a minimum amplitude criterion that excludes phases (peak to trough or trough to peaks) that have an amplitude of less than one standard deviation of the corresponding growth rates (negative or positive) in the series to be dated.

Figure 2 shows the original level series together with the recessions (represented by bars) identified with the BB methods.<sup>13</sup> They clearly coincide with periods of falls

<sup>8</sup> An application of the MS method to Spanish regional cycles can be found in Gadea et al. (2006).

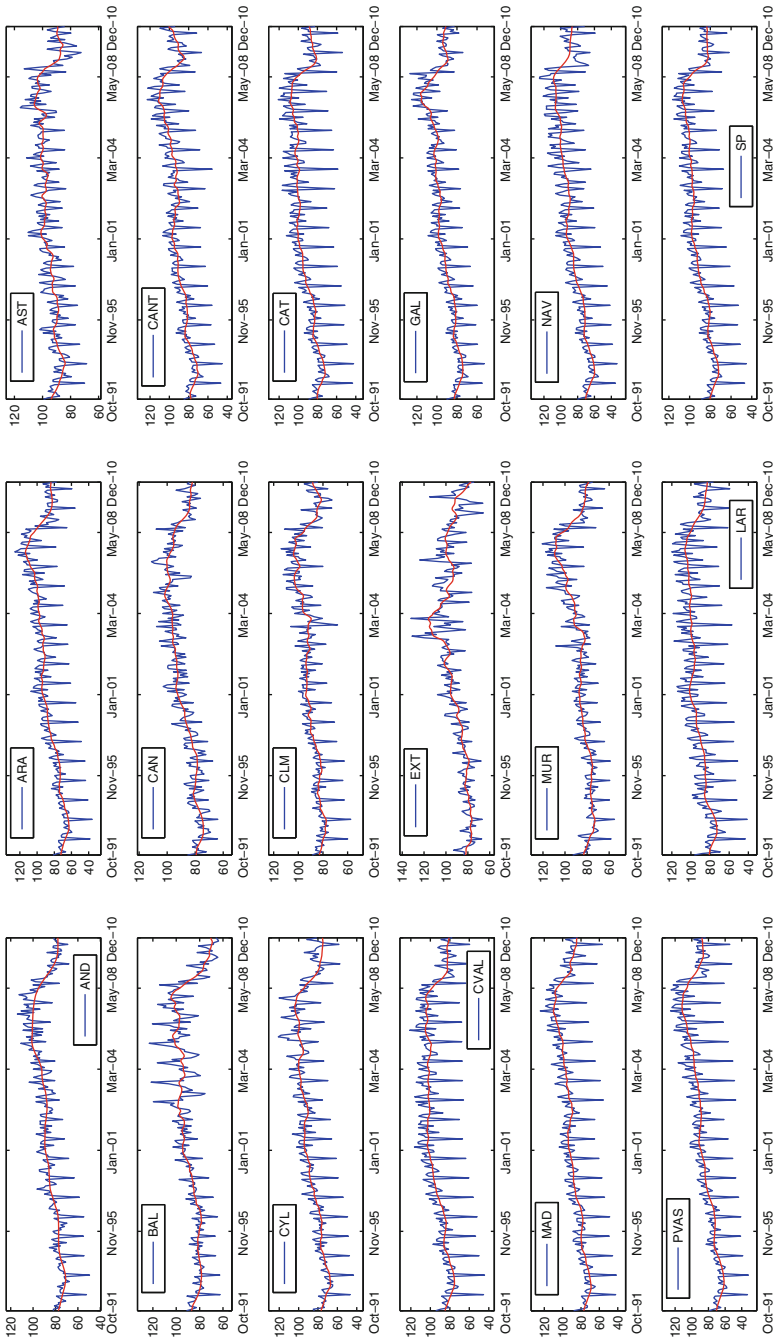
<sup>9</sup> See Baxter and King (1999) or Kaiser and Maravall (2005).

<sup>10</sup> See, among others, Mönch and Uhlg (2005), Bengoechea et al. (2006) and Camacho et al. (2008).

<sup>11</sup> This method was developed by Gómez and Maravall (1996). A Matlab routine is available in the website of the Bank of Spain.

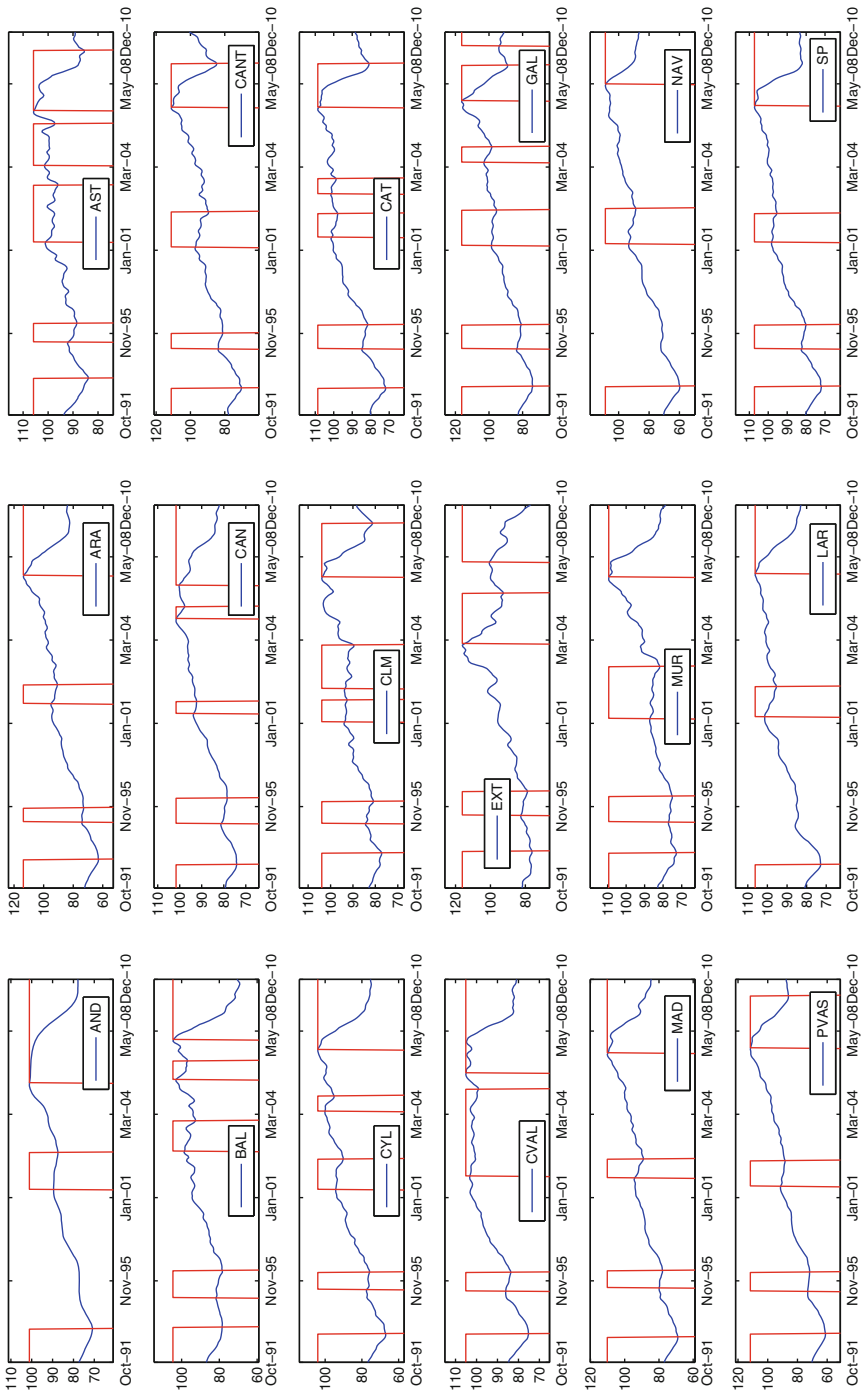
<sup>12</sup> In the case of Spain it makes no difference whether we work with the cycle-trend component or the seasonally-adjusted series without outliers and calendar effects. However, in some regions the use of the smoother cycle-trend series produces better results, avoiding the identification of false signals.

<sup>13</sup> As a starting point, we have used the code written by Watson and Denson in Watson (1994) into which we have incorporated our own specifications.



**Fig. 1** Cycle-trend adjusted Spanish and regional industrial production indexes (levels)





**Fig. 2** Bry-Boschan cycle dating

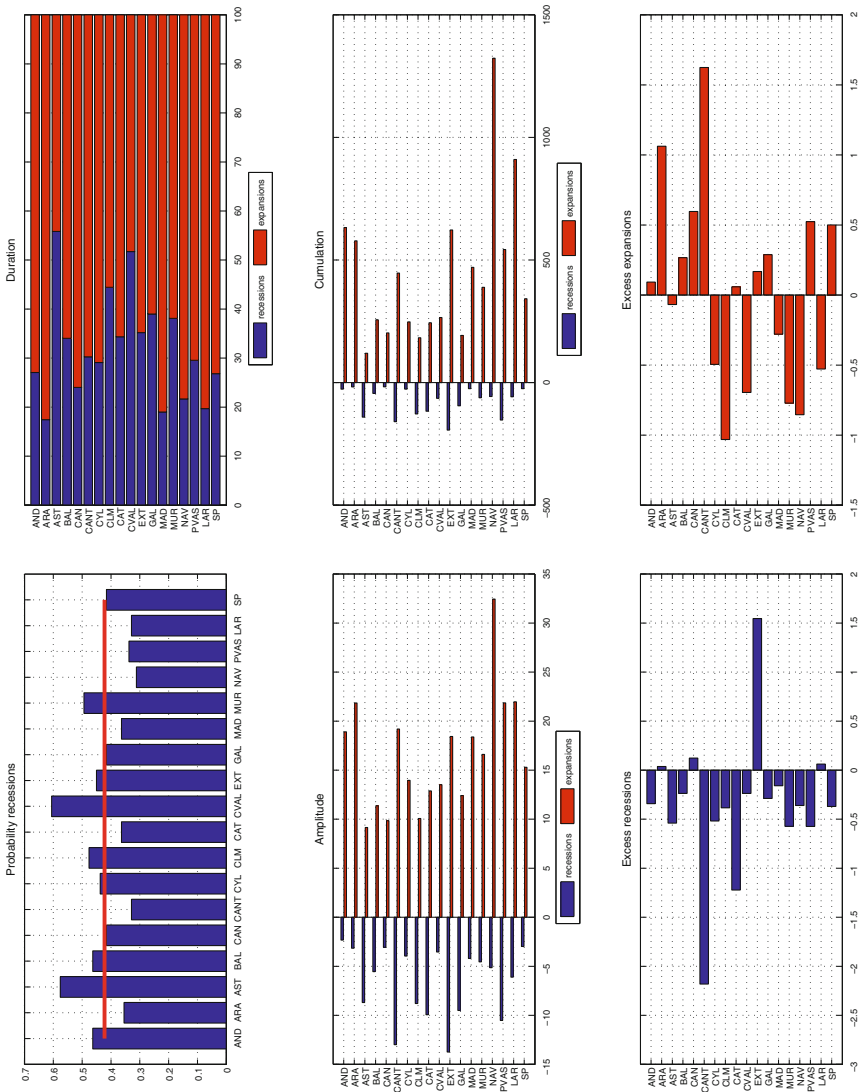
in industrial production and reproduce the “known-knowns” recent aggregate Spanish business cycle. Consequently, we can trust BB techniques to find the chronology of turning points at regional level. With respect to the regional cycles, in most cases, the BB method locates four recession periods which coincide with two well-known crises, the beginning of the 90s and the current one, and two deceleration episodes during the long expansion. Nevertheless, some regions present a different chronology. AST and CVAL have spent almost all the noughties in recession. In some other regions, the algorithm detects a deceleration period between the deceleration of the early 2000s and the deep recession that started in 2008. In five regions, some positive data in the latest part of the sample suggest that this last recession period is ended. However, more data would be needed to confirm this outcome. Thus, we should analyse 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, and the mean duration, amplitude, cumulation and excess of recessions and expansions.<sup>14</sup> All these results appear in [Fig. 3](#). The probability of recessions is 0.42 for Spain (and for the regions on average), the mean duration of the recessions is 17 months (19) and the mean duration of the expansions is 45 months (43).<sup>15</sup> These results are plausible and agree with the stylized fact that expansion periods are longer than recessions. It is noteworthy that, while recessions last a similar number of months to those in Europe, expansions last more than one year more in Spain and the regions on average, which could explain the convergence 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 17 regions (and in Spain), the average probability of a recession is just above 40% but only eight of these eighteen geographical units are above the mean. CVAL and AST are the regions with the highest probability, around 60%, and the lowest probabilities, just above 30%, are found in NAV, CANT and LAR. With respect to the duration, NAV, LAR and AND stand out as the regions with the longest expansions (more than 50 months) and the shortest-lived recessions are found in ARA, CAN and MAD. In fact, in ARA, LAR and MAD, less than a fifth of the cycle duration is spent in a recession. Nevertheless, CVAL and AST present long-lasting recessions (around 30 months). Furthermore, we can appreciate big asymmetries in duration in the case of ARA (and, to a lesser extent, MAD and LAR); while, on average, regional expansions last 2.2 times as long as recessions, in ARA, the ratio is almost 5 times (and just above 4 in MAD and LAR). On the contrary, in AST and CVAL recessions last 5 and 2 months more than expansions, respectively. Highly symmetrical cycles are found in CLM and GAL.

<sup>14</sup> [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 the 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.

<sup>15</sup> 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 expansions last about 30 months, while recessions last 15 months.



**Fig. 3** Dissecting the cycles of Spain and its regions

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 pp and, in all regions, the amplitude of expansions is bigger than that of recessions. Of special interest are the cases of AST and CLM, where the amplitude across the two phases is almost identical. EXT presents the highest loss during recessions. In AND, ARA and NAV, three regions that show very pronounced asymmetries between the amplitude of their business cycle phases, the expansions are about seven times greater than the recessions.

NAV and LAR have the clearest cumulative gains during expansions and EXT and CANT 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. It combines the duration, amplitude and shape of the business cycle. However, it is normally calculated by the triangle approximation which does not capture the actual cumulation precisely 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.

Neither a concave nor a convex shape of regional business cycle dominate during expansions. When the excess is negative, it means that 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 characterised by a high growth period that ends in a normal growth period. During recessions, the concave shape of the phase clearly dominates (in 13 cases). Consequently, the paths exhibit gradual changes at the beginning of the phase that become sharp at the end. The fact that both shapes are concave is positive because it means that the wealth losses in recessions are lower and the gains in expansions higher than in a linear behaviour. So, regions with convex expansions (CANT, ARA, CAN, PVAS, GAL, BAL, EXT, AND and CAT) and SP do not benefit as much as the others from an expansion, while regions with convex recessions (EXT, CAN, LAR and ARA) have a bigger wealth loss than the others when they are in recession. Both negative features occur in EXT, ARA and CAN.

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 synchronisation, which could be interpreted as a sign that regional economies move together.

### 3.2 Measures of synchronisation

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 synchronisation in depth. To that end, we use different measures of the synchronisation 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 synchronisation, 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.

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  $\chi^2$  with  $(s - 1) \times (s - 1)$  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 \(2006\)](#), 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:

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

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 synchronisation 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 \(2006\)](#) 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  $\mu_{S_t}$  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 analyse whether the degree of synchronisation 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.<sup>16</sup>

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

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

and the test has this expression:

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

<sup>16</sup> Following [Camacho et al. \(2006\)](#), we select the probability of the geometric distribution so that its expected value is equal to the average duration of expansions.

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

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

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

The vector  $\widehat{\theta}' = [\widehat{\mu}_{S_1}, \dots, \widehat{\mu}_{S_n}, \widehat{\rho}_{S_{12}}, \dots, \widehat{\rho}_{S_{(n-1)n}}]$  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-synchronisation), it is  $[\widehat{\mu}_{S_1}, \dots, \widehat{\mu}_{S_n}, 0, \dots, 0]$  or, if we want to test the hypothesis of SMS (strong multivariate synchronisation), for instance  $\rho_{S_{12}} = \rho, \forall i \neq j$  with  $\rho \in (0, 1)$ , it is  $[\widehat{\mu}_{S_1}, \dots, \widehat{\mu}_{S_n}, \rho, \dots, \rho]$ . 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 synchronisation) the distribution is more complex, being a Brownian motion, or applying the Cramer–VonMises equivalent, a weighted average of  $\chi^2$  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

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

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 increases 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 \(2003\)](#).<sup>17</sup>

Figure 4 summarizes all the synchronisation measures, contingency, concordance and correlation. Instead of showing the three  $18 \times 18$  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 synchronisation of the regions in a small number of dimensions. In the three cases, the corresponding synchronisation index is used as the distance matrix; then, we transform the similarity matrix into a dissimilarity and reproduce its Euclidean distances.<sup>18</sup> A preliminary examination of the eigenvalues of this matrix shows that two dimensions are not enough to represent the points suitably and that 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  $\rho_{S_{ij}}$  (0.53, on average, for regions and 0.69 for Spain) are typically smaller than those obtained with the concordance index, which are 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 (0.65 and 0.79, respectively). Nonetheless, the evidence for rejecting the null hypothesis of no association is very strong between regions, 97% on average when we consider the Harding–Pagan  $t$  test and 98% with the Pearson independence test (see Table 1). With both tests, all non-rejections correspond to EXT, which is also the only region that is not synchronised with SP.

Contingency, concordance, and correlation are nearly identical with respect to their classification of regions.<sup>19</sup> We can see that MAD, ARA, PVAS and CYL, are the regions most connected with the rest, while EXT, AST, CVAL and BAL are the most isolated. Notice that both groups of regions have atypical business cycle features. EXT, AST, CVAL and BAL are regions with low asymmetry between the duration and amplitude of expansions and recessions, presenting brief growth periods. On the contrary, the more synchronised regions are characterised by a great asymmetry in favour of expansions and a cumulation of wealth gains during them. We also find another stylized fact: the degree of regional synchronisation is directly related to the degree of synchronisation with Spain.

The results of such a high correlation are confirmed by the bootstrap exercise for  $\rho_{S_{ij}}$  because the zero is outside the confidence interval in all cases.<sup>20</sup> Nevertheless, as we can see in Fig. 5, where we present the density of correlation coefficients, their

<sup>17</sup> 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.

<sup>18</sup> A detailed explanation of this technique can be found in [Timm \(2002\)](#).

<sup>19</sup> In all cases the rank test of Spearman is about 0.98.

<sup>20</sup> We do not include these results to save space.



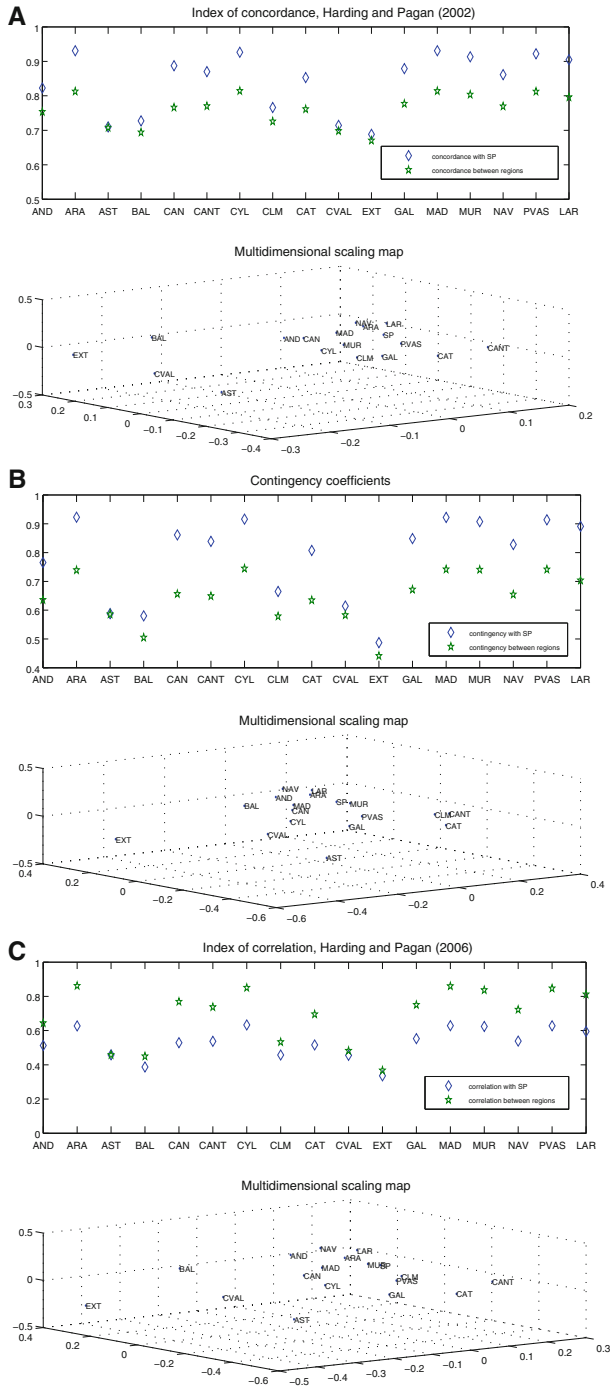


Fig. 4 Synchronisation measures

**Table 1** T-Ratio for  $\rho$  and independence test

	AND	ARA	AST	BAL	CAN	CANT	CYL	CLM	CAT	CVAL	EXT	GAL	MAD	MUR	NAV	PVAS	LAR	SP
AND	–	8.34	5.25	3.47	5.69	5.99	6.63	3.45	3.63	4.88	2.99	4.50	4.69	7.15	13.39	6.49	14.12	7.84
ARA	93.24	–	3.83	4.88	7.53	8.33	10.35	4.98	6.50	6.81	3.76	7.78	14.53	10.62	9.56	10.59	12.02	14.31
AST	53.49	30.31	–	2.84	4.60	6.16	6.65	5.91	6.44	5.83	4.74	5.24	5.12	6.25	3.90	9.19	4.23	5.04
BAL	29.25	51.47	19.16	–	4.17	1.70	6.26	5.74	3.33	3.90	5.44	4.94	6.72	7.88	5.20	8.07	5.17	7.31
CAN	62.52	94.94	41.08	39.70	–	6.15	4.24	4.42	7.37	5.64	4.74	5.64	7.27	5.20	7.91	7.94	9.13	11.51
CANT	60.94	105.04	51.15	7.57	65.19	–	6.87	6.83	10.62	3.16	0.82	9.05	8.17	5.33	9.39	6.82	6.41	9.30
CYL	46.13	63.44	64.14	61.97	35.54	85.58	–	4.37	6.40	7.07	5.35	15.05	16.49	10.00	12.85	17.32	16.72	16.52
CLM	28.05	51.05	62.53	63.02	42.15	75.95	40.31	–	8.40	4.45	0.88	4.34	9.07	4.11	7.93	4.53	4.45	6.11
CAT	78.07	136.49	73.06	26.13	88.87	144.90	69.68	86.70	–	5.56	1.04	7.23	6.89	6.51	3.55	10.33	4.92	8.45
CVAL	30.37	79.44	58.50	33.47	60.78	20.87	74.46	39.58	49.33	–	4.28	4.93	8.48	5.11	4.24	5.77	4.56	5.46
EXT	22.35	33.95	45.19	58.45	47.86	1.81	57.85	2.10	2.89	38.65	–	3.50	4.53	2.97	3.21	3.48	2.80	3.63
GAL	43.14	100.45	48.30	15.14	62.16	119.15	153.45	38.75	89.52	45.98	28.17	–	7.79	8.74	6.85	8.65	7.82	10.65
MAD	86.28	130.47	69.75	59.37	85.48	83.36	149.02	39.68	85.15	75.74	44.20	89.52	–	10.59	8.39	12.41	11.56	14.81
MUR	115.03	117.03	28.41	22.50	52.25	98.83	119.22	109.51	74.48	52.53	21.86	106.38	129.18	–	12.84	14.04	13.41	17.67
NAV	69.46	130.65	76.50	33.91	89.89	63.28	115.44	34.71	30.88	31.69	25.21	75.16	99.73	170.36	–	6.64	12.28	8.51
PVAS	124.90	144.12	32.90	22.25	90.15	122.24	144.75	78.76	124.87	49.57	27.88	100.92	138.13	120.86	90.60	–	9.24	13.16
LAR	87.44	166.79	46.49	47.37	105.97	90.93	139.15	40.90	54.52	36.01	19.74	90.15	138.06	116.25	163.64	122.24	–	11.25
SP	95.67	171.39	48.30	46.73	136.35	125.44	167.08	65.54	111.76	53.69	31.09	130.10	170.81	161.73	120.46	165.61	1152.19	–

HAC  $t$ -Student for correlation coefficient in the upper triangle and  $\chi^2$  contingency test in the lower triangle

variability is quite important in several regions. The correlation coefficient has also been estimated recursively throughout the sample to capture changes in synchronisation over time. Figure 6 shows the evolution of the regional average of  $\rho_{S_{ij}}$  and its value for each region and SP. The similarity of the business cycles of EXT and CVAL with respect to both other regions and Spain has reduced dramatically during the last decade, indicating a lack of convergence. This similarity has also decreased in BAL, AST and CLM but, to a lesser extent. Other regions, such as LAR, NAV, CAN and CYL also show a loss of synchronisation during the long expansion phase but seem to recover it during the current recession. Finally, ARA, MAD and PVAS 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 synchronisation between the 17 regions by using the statistic previously described. As we suspected, the test rejects all hypotheses from non-synchronisation to perfect synchronisation, passing through intermediate degrees ( $\rho = 0.1, 0.2, \dots, 0.9$ ), and the matrix is nearly singular even when a great intensity of shrinkage is considered (see Table 2). However, if we apply the bootstrapped version, we are not able to reject the null in any case, reflecting the loss of power.

As we cannot obtain any conclusion from the multivariate synchronisation test when we apply it to 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 multidimensional scaling maps demonstrates that, although some regions, such as EXT, CVAL, AST and BAL, appear isolated from the rest, it is not easy to establish a clear pattern of clustering. So, we will deal with the regional clustering using an alternative approach in Sect. 4.

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

As well as cycle dating approaches (parametric and non-parametric) to characterise the business cycle, there are other techniques that analyse comovements of economic variables using original data and extracting common factors.<sup>21</sup> One basic feature of the

<sup>21</sup> 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. See [Forni et al. \(2004\)](#).

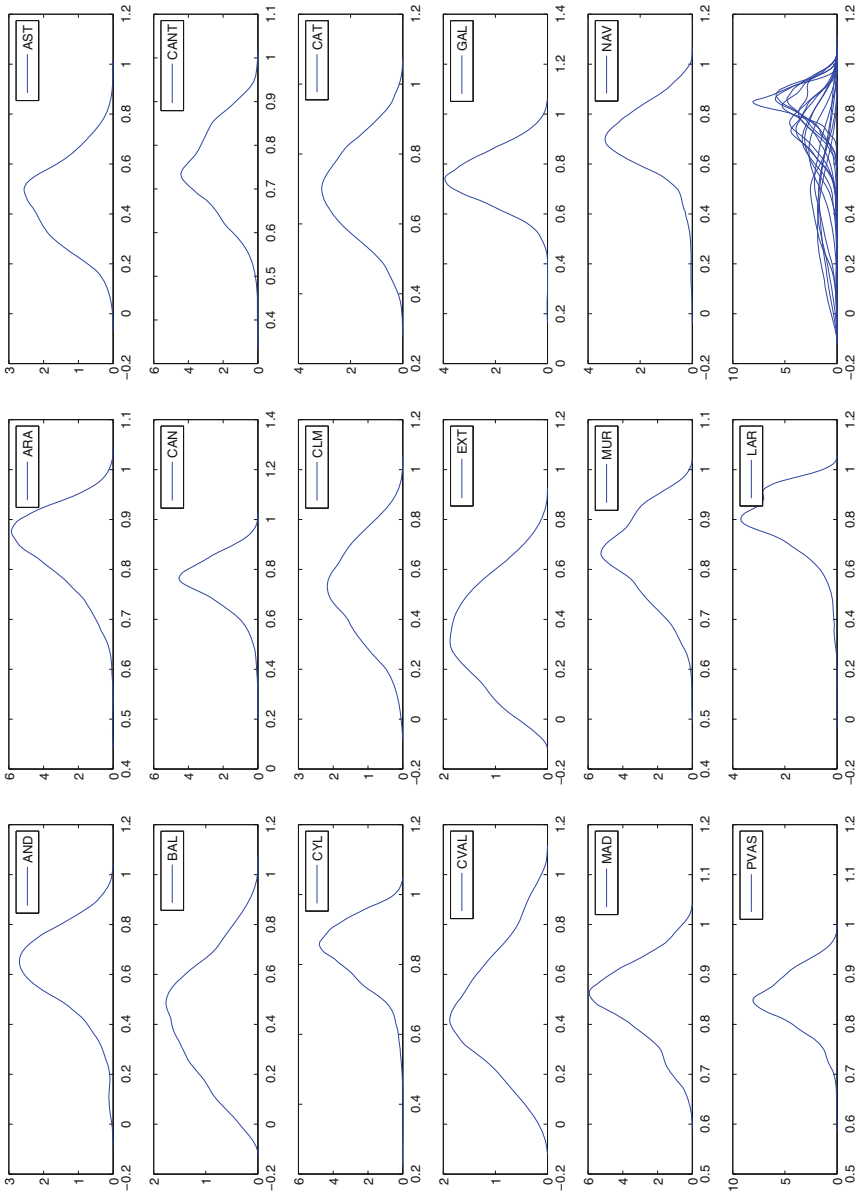


Fig. 5 Density of correlation coefficients

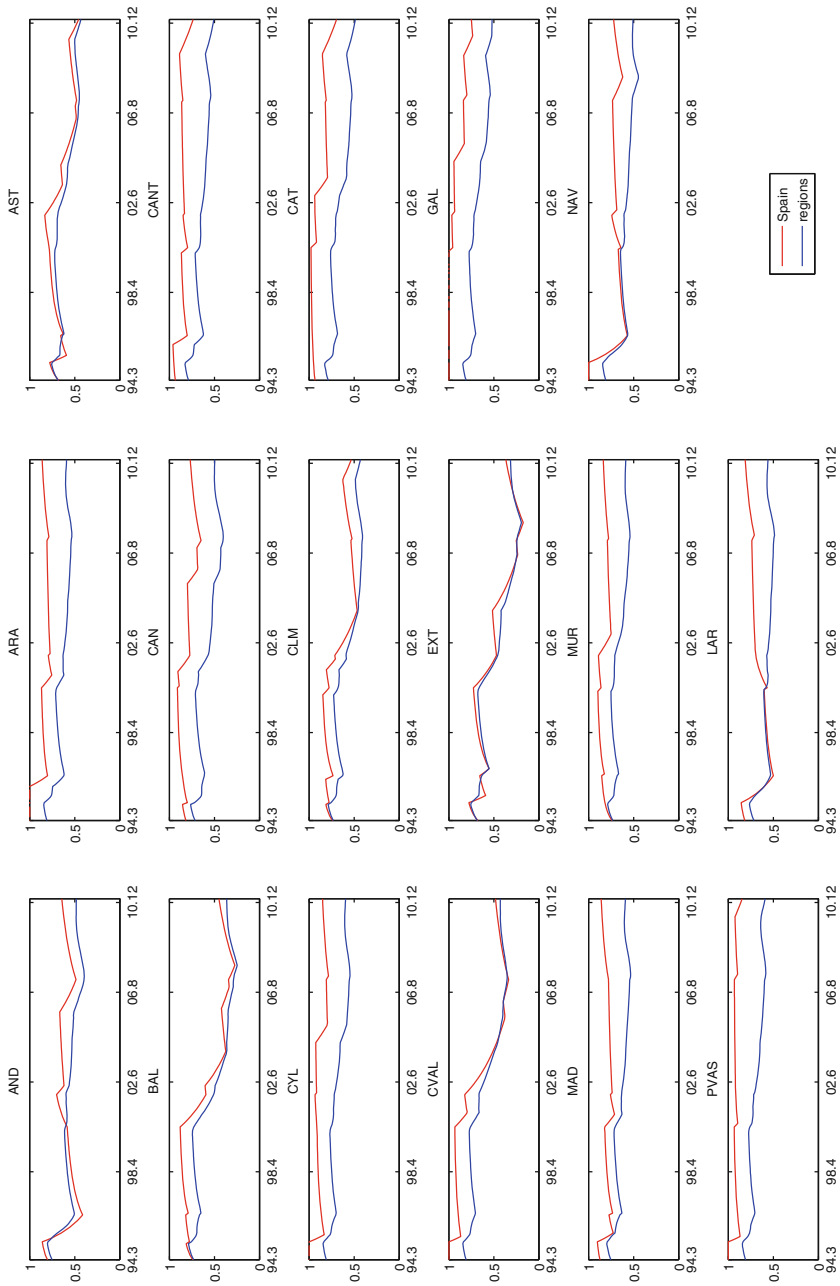


Fig. 6 Recursive correlation coefficients

**Table 2** Multivariate synchronisation test

	Test	Asymptotic critical value	Bootstrap critical value
SPPS	2312.62	164.22	$7.67 \times 10^4$
SMS			
$\rho = 0.1$	1897.07	164.22	$6.385 \times 10^4$
$\rho = 0.2$	1553.17	164.22	$5.134 \times 10^4$
$\rho = 0.3$	1280.91	164.22	$4.051 \times 10^4$
$\rho = 0.4$	1080.30	164.22	$3.25 \times 10^4$
$\rho = 0.5$	951.33	164.22	$2.84 \times 10^4$
$\rho = 0.6$	894.00	164.22	$2.71 \times 10^4$
$\rho = 0.7$	908.33	164.22	$2.98 \times 10^4$
$\rho = 0.8$	994.30	164.22	$3.45 \times 10^4$
$\rho = 0.9$	1151.91	164.22	$3.83 \times 10^4$
PS	267.24	27.59	26.30

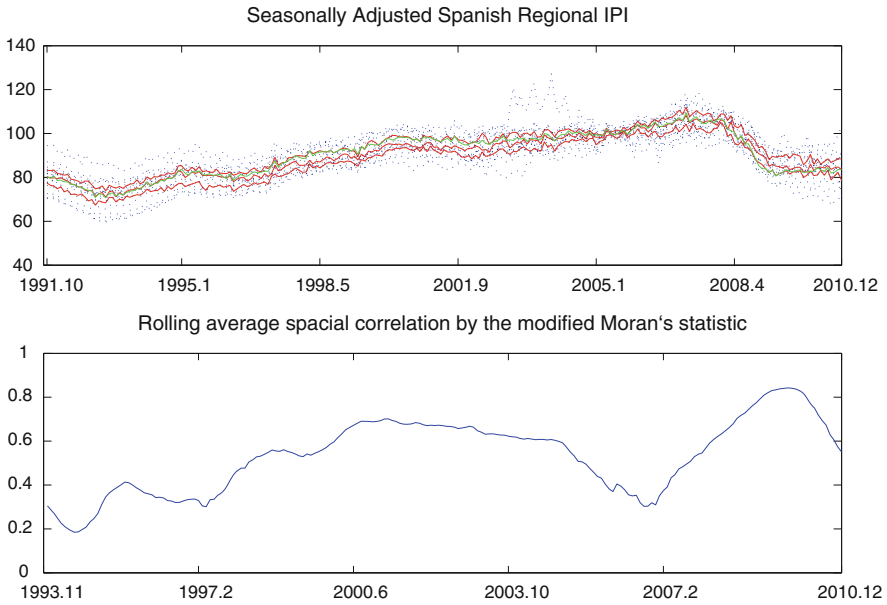
business cycle is, precisely, the presence of comovements across economic variables. Comovement measures constructed in the frequency domain, principal components and dynamic factor models are the main branches of this approach that deals with the original information. Of them, dynamic factor models (DFM henceforth) have recently emerged as a powerful tool to analyse 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.<sup>22</sup>

In this section, we first carry out a preliminary analysis of comovements by using an index proposed by [Stock and Watson \(2010\)](#). Second, we estimate the optimal number of factors, both static and dynamic, present in the 17 Spanish regions and identify the national component. Then, we apply a DFM to test whether the 17 Spanish regions move according to common driving forces. Finally, we subtract the component associated with the Spanish cycle and form clusters using the idiosyncratic regional components.

#### 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. From here to the end of the section, we use the original regional IPI series  $y_{ti}$  with  $i = 1 \dots 17$  seasonally adjusted

<sup>22</sup> There are also approaches in the literature that deal simultaneously with comovements and business cycle dating. For a summary, see [Camacho et al. \(2010\)](#).



**Fig. 7** Comovements

and cleaned of outliers. We have opted for this filter because it conserves business cycle signals better than other alternatives such as growth rates or cycle-trend components. Figure 7 shows the evolution of the regional industrial index (dashed lines) in levels (top figure). It also shows (solid lines) the median and 25 and 75% percentiles. There is a considerable dispersion of the regional business cycles, which increases dramatically after 1999, just after Spain joined the EMU and 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:

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

where,

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

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

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

$$N = 17$$

The outcome, time series  $\widehat{I}_t$ , is plotted at the bottom of Fig. 7. We observe that the synchronisation of comovements is 0.53, on average, during all the period, reaching its minimum value (around 0.18) in 1994.6 and its maximum in 2009.10 (more than 0.8).<sup>23</sup> So, it seems there is an inverse relationship between economic growth and regional comovements which has been especially intense in the recent big recession. This result is very similar to that obtained with the correlation index from the BB cycle dating whose regional average was 0.53. Furthermore, some regions showed a loss of synchronisation during the expansion phase but began to recover it in the current recession.

## 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).<sup>24</sup>

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

### 4.2.1 How many factors are there in the Spanish regions?

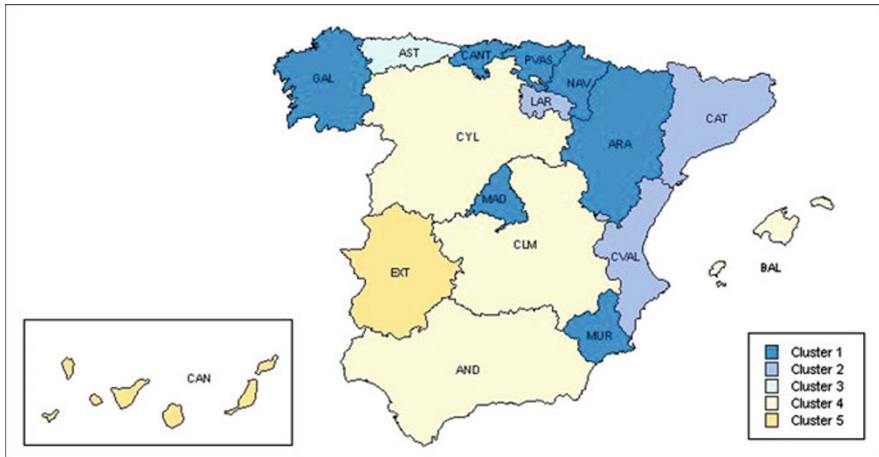
Bai and Ng (2002) developed a formal statistical procedure that consistently estimates the number of factors in a set of data with cross-sectional and temporal dimensions,  $N$  and  $T$ , respectively. Let  $Y_{it}$  be the observed data for region  $i$  in time  $t$ . The factor model then has the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

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

<sup>23</sup> 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.

<sup>24</sup> See Stock and Watson (2010) for a comprehensive survey of this literature.





**Fig. 8** Cluster map

where  $F_t$  is a vector of  $r$  common factors,  $\lambda'_i$  is a vector of factor loadings associated with  $F_t$  and  $\varepsilon_{it}$  is the idiosyncratic component of  $Y_{it}$ . The product  $\lambda'_i F_t$  is called the common component of  $X_{it}$ . For all the units and using matrix notation

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

where the dimensions of  $Y_t$ ,  $\Lambda$ ,  $F_t$ , and  $\varepsilon_t$  are  $N \times 1$ ,  $N \times r$ ,  $r \times 1$  and  $N \times 1$ , respectively. They propose estimating the common factors by minimizing the following expression:

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

and consider the estimation of the number of factors  $r$  as a model selection problem so they construct different information criteria. When this procedure is applied to regional IPI growth, all the criteria coincide in estimating 5 factors.<sup>25</sup>

Having established the number of factors, we re-estimate the factor model with 5 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 8 shows these values. We observe that AST and EXT have the largest idiosyncratic components and LAR, BAL, CAN and MUR also have large variances, while the lowest variances are found in AND, CAT, PVAS and NAV. These results confirm those obtained in Sect. 3, where EXT and AST, in most cases, and the islands, in some cases, presented very peculiar behaviours, while PVAS and MAD were the regions

<sup>25</sup> This result is obtained with a maximum value of 5 factors that explain 70% 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.

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  $\varepsilon_t = u_t + \sum_{i=1}^N \xi_{it}/T$  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 representation of the aggregated regional ones and, consequently, suitably represents the common Spanish business cycle.<sup>26</sup> 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 17 different behaviours. After confirming this fact, we have removed SP from the estimation of the DFM because it is redundant and may bias the rest of the results.

Up to now in this section, 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  $r$  will have 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  $r \times r$ . They define two statistics as the sequential sum of the eigenvalues and estimate  $\hat{q}_3$  and  $\hat{q}_4$  as the minimum values that allow us to bound this sum. The value of  $q$  is estimated using the correlation matrix for  $r = 5$ , which explain around 97% of the data variability, and we find that  $\hat{q}_3 = 1$  and  $\hat{q}_4 = 2$ . If we increase the number of static factors to  $r = 9$  (which explain 99% of the variation in the data),  $\hat{q}_3$  maintains 1 dynamic estimated factor and  $\hat{q}_4$  is now 2. When  $r$  is between 1 and 4, both tests estimate  $q = 1$ . Similar results are obtained when the covariance matrix is used; both tests estimate  $q = 1$  for  $r = 1 \dots 4$  and  $\hat{q}_3 = 1$  and  $\hat{q}_4 = 2$  for  $r = 5$ . Therefore, the results suggest that a common dynamic factor or primitive shock is the most plausible conclusion from our analysis and, consequently, that Spanish regions have a dynamic common factor that is identified with the national component. In the next section, we will extract this common factor and construct clusters with the residuals of the regional idiosyncratic factors.

#### 4.2.2 Cluster analysis

Up to this point, we have documented a certain degree of synchronisation in regional cycles but, also, a high degree of 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 one. So, we remove the common factor associated with the national component to make the comovements between regions more visible, using

<sup>26</sup> 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.

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  $\varepsilon_{it}$  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.  $N(0, \sigma_\eta)$  and  $N(0, \sigma_\varepsilon)$ , respectively.<sup>27</sup> 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 remove the common factor and obtain the residuals  $\widehat{\varepsilon}_{it}$  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 synchronised 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 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 deciding the number of groups beforehand. Furthermore, through the method of Bai and Ng (2002), four common factors have been identified, which gives us a clue about the number of regional clusters.

The  $k$ -means clustering method 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 its dependence on the initial partition), Stock and Watson (2010) propose a modified algorithm that repeats the procedure for multiple starting values and analyses the value of the minimized objective function as a proxy of the most suitable number of clusters.

<sup>27</sup> We should note that we are implicitly assuming that the IPI variables exhibit a unit root captured by the common factor  $F_t$ . To verify this hypothesis, we have applied a wide battery of unit root tests to both the raw and the transformed variables, given the distortion caused by the methods for removing the seasonal component on the unit root inference, as is reported in Maravall (1993). In any event, the results obtained confirm the presence of a unit root and are available upon request.

We apply this method to the idiosyncratic regional factors.<sup>28</sup> Increasing from 2 to 3 clusters reduces the value of the minimized objective function by approximately 20%, and increasing from 3 to 4 and 4 to 5 reduces the value by 21 and 27%, respectively. After five clusters, the value of the objective function is not reduced. Therefore, taking these results together with the findings from the hierarchical method and the optimal number of common factors, we select five clusters.

The result with  $k = 5$  groups determines that ARA, CANT, GAL, MAD, MUR, NAV and PVAS are included in the first cluster, CAT, CVAL and LAR in the second, AST in the third, AND, BAL, CYL and CLM in the fourth and the other regions, CAN and EXT, in the fifth. All of these five groups fit in with the regional features identified in the previous sections of this paper quite well. On the one hand, there are the regions most isolated from the rest and with the largest idiosyncratic components (groups 5, 4 and 3) and, on the other hand, there are the most synchronised regions where the patterns are more stable throughout the sample. The map of Fig. 8 illustrates these results.

## 5 An economic characterisation of cluster patterns

In this section, we will try to find the sources of the similarities as well as the discrepancies among the previously identified clusters. 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.<sup>29</sup>

However, as well as these structural variables, the business cycle features and the degree of regional synchronisation found in Sect. 3, could be important to understand and characterise the five clusters identified. We have selected the probability of being in a recession, the correlation coefficients with the other regions and with Spain and four representative structural variables, namely, industrial weight, unemployment rate, per capita income level and human capital.<sup>30</sup>

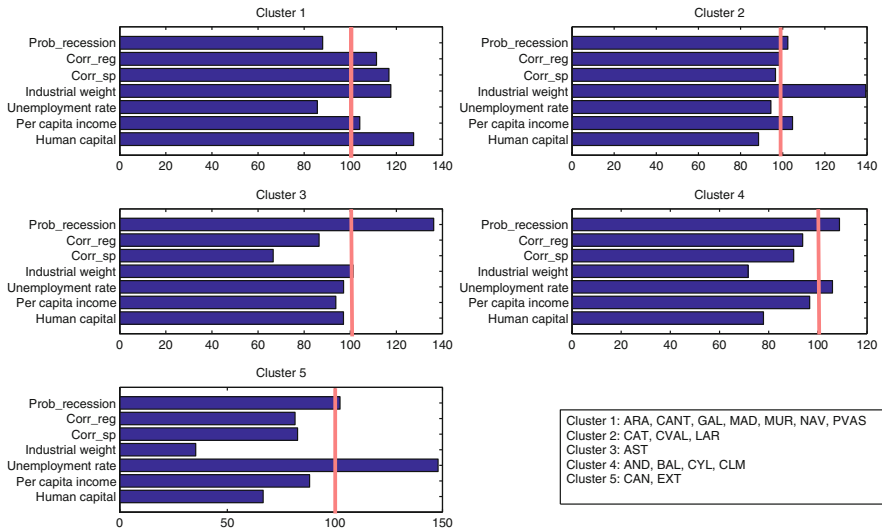
The characterisation of the five clusters is presented in Fig. 9.<sup>31</sup> The largest Group 1 (ARA, CANT, GAL, MAD, MUR, NAV and PVAS), contains the majority of the most developed regions of the country. It presents the highest level of human capital (22.5) and the lowest unemployment ratio (12.36). Furthermore, the per capita income level is the second highest of the clusters, while the average industrial weight is clearly

<sup>28</sup> 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.

<sup>29</sup> 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.

<sup>30</sup> 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 percentage of population with a university degree from Ivie (1991–2007).

<sup>31</sup> The values of the variables have been standardised, the regional average being 100.



**Fig. 9** Cluster features

above the Spanish mean. When we observe the basic features of the business cycle phases, this group clearly has the lowest probability of being in a recession (0.37). It is also the group that is most synchronised with the other regions (0.59) and, particularly, with Spain (0.8).

Group 2 (CAT, CVAL and LAR) presents similar characteristics to the previous one. This group includes three of the most industrialised regions in our sample (23.85 of industrial weight). It is the wealthiest group, as its average per capita income level is 12,041.28 euros. Its average unemployment rate (13.6) also confirms its similarity to Group 1. However, the average percentage of population with a university degree is 7pp lower than in Group 1. The probability of being in a recession is 0.43. The correlation, both with respect to the other regions and to Spain, is the second highest of the groups (0.52 and 0.66, respectively).

Group 3 shares some of the positive characteristics of the previous groups together with some of the negative characteristics of the last two groups. It is made up of just one region, AST. When we consider the human capital level, it occupies the second position of the ranking (17.13 years) and it is in an intermediate position when we focus on the other structural variables. Nevertheless, the recession phase lasts more than in any other group (0.58) and is the region that is least correlated with Spain (0.46). Although it is not the least correlated to the other regions, it is clear that it presents important cyclical discrepancies.

Group 4 brings together four regions (AND, BAL, CYL and CLM) where the unemployment ratio is high and where the level of population with a degree is almost 10pp less than in Group 1. The weight of the industrial sector is below the Spanish average; it would be lowest if CYL were not included in the group, as some of its specific features have more to do with those of Groups 1 and 2. Furthermore, the distortions of CYL and, especially, of BAL make this group the third wealthiest. However, the

cycle is fairly symmetric as the probability of being in a recession is 0.46. It is the third group in the ranking in terms of degree of synchronisation.

Finally, Group 5 is made up of EXT and CAN. It is the poorest group in per capita income levels (10,156.89 euros), the population with a degree is 11.77 and the unemployment ratio is the worst (21.33). The importance of the services sector in the islands and the productive structure of EXT which is more oriented towards agriculture and construction make this group the lowest in industrial weight. In spite of this, the probability of being in a recession is the same as in Group 2. Its connection to the other regions is the lowest (0.43), although it is much more correlated with Spain than Group 3.

In sum, the groups with a high unemployment rate, on average, are more isolated from the rest, while groups with high industrial weight and per capita income levels or where the population is more educated also show a higher degree of synchronisation. So, economic policies that focus on the improvement of these variables may enable a higher degree of regional convergence, measured through the level of synchronisation.

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.

In addition, to better understand the features of the clusters we also test the influence of the geographical situation, namely, the border effect, which has been calculated using the results from Sect. 3.2. Following [Croux et al. \(2001\)](#), this effect can be easily measured through the following ratio:

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

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 does not matter, which is confirmed by the non-rejection of the null hypothesis of the Wilcoxon test at 5% of significance (Table 3). NAV and LAR are the regions with the highest border effects, while AND, CLM and EXT present values under one.<sup>32</sup> These regions, that have been receiving structural and cohesion funds from the EU, continue to be the most isolated ones and, as we saw in the previous cluster characterisation, they also present some of the worst results in terms of macroeconomic variables. Other regions with an important border effect are GAL, PVAS, CANT and AST, which we could denominate the Cantabrian coast group. In the last three regions, the weight of the

<sup>32</sup> 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.

**Table 3** Do borders matter?

	Ratio	Neighbour regions
AND	0.83	3
ARA	1.02	6
AST	1.08	3
BAL	1.16	2
CAN	–	0
CANT	1.12	3
CYL	1.06	9
CLM	0.89	7
CAT	0.92	3
CVAL	1.05	5
EXT	0.90	3
GAL	1.15	2
MAD	0.97	2
MUR	0.95	3
NAV	1.34	3
PVAS	1.14	4
LAR	1.31	4
Wilcoxon test	53	

The critical value of the  
Wilcoxon test at 5% is 35

industrial sector in the 90s was really important and they share some characteristics with GAL. Indeed, all four regions, but AST, belong to Group 1. NAV, LAR and ARA (which also presents a significant border effect) are traditionally identified as the Ebro Valley regions. The high border effect of BAL, considering that it consists of a group of islands, is striking but we establish that its boundaries are CAT and CVAL. Thus, although, for whole sample, neighbourhood is non significant, we see that, for most regions, political boundaries influence their degree of synchronisation and help to give a complementary explanation to the clusterisation patterns previously found.

## 6 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 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 synchronised 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 convergence. We do not find signs of convergence in most regions, which is confirmed both by the recursive coefficients and the comovement index. This result should be taken into account for economic policy design, as convergence is a

long-run economic process more than a process related to fluctuations in the economy. When grouping regions, we find that clusters with an important industrial weight, per capita income and human capital and a low unemployment rate also show a high degree of synchronisation.

There are many papers on intra-country linkages in an international context, but 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 deciding whether or not to belong to more integrated economic areas, it is surely fundamental for reducing intra-country economic heterogeneity through the design of adequate economic policies. We carry out this analysis for Spain, contributing 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 the same as those of the 17 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.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Airov J (1963) The construction of interregional business cycle models. *J Reg Sci* 5(1):1–20
- Artis M, Kontolemis ZG, Osborne DR (1997) Business Cycles for G7 and European Countries. *J Bus* 70:249–279
- Artis M, Krolzig HM, Toro J (2004) The European business cycle. *Oxf Econ Pap* 56(1):1–44
- Bai J, Ng S (2002) Determining the number of factors in approximate factor models. *Econometrica* 70:191–221
- Bai J, Ng S (2007) Determining the number of primitive shocks in factor models. *J Bus Econ Stat* 25:52–60
- Baxter M, King RG (1999) Measuring business cycles: approximate band-pass filters for economic time series. *Rev Econ Stat* 81:575–593
- Bengoechea P, Camacho M, Pérez Quirós G (2006) A useful tool to forecast the Euro-area business cycle phases. *Int J Forecast* 22:735–749
- Bordo MD, Helbling T (2003) Have national business cycles become more synchronized? NBER working paper no. W10130
- Borts GH (1960) Regional cycles of manufacturing employment in the United States, 1914–1953. *J Am Stat Assoc* 55(289):151–211
- Bry G, Boschan C (1971) Cyclical analysis of time series: selected procedures and computer programs. NBER, New York
- Burns AF, Mitchell WC (1946) Measuring business cycles. NBER, New York
- Camacho M, Pérez-Quirós G, Saiz L (2006) Are European business cycles close enough to be just one? *J Econ Dyn Control* 30:1687–1706
- Camacho M, Pérez-Quirós G, Saiz L (2008) Do European business cycles look like one? *J Econ Dyn Control* 32:2165–2190



- Camacho M, Pérez-Quirós G, Poncela P (2010) Green shoots in the Euro area. A real time measure. Banco de España working paper, 1026
- Cancelo JR (2004) El ciclo del empleo en España. Un análisis regional. *Revista económica de Castilla La Mancha* 4:107–138
- Candelon B, Piplack J, Straetmans S (2009) Multivariate business cycle synchronization in small samples. *Oxf Bull Econ Stat* 71(5):715–737
- Carlino G, DeFina R (1995) Regional income dynamics. *J Urban Econ* 37:88–106
- Carlino G, DeFina R (1998) The differential regional effects of monetary policy: evidence from the U.S. States. Federal Reserve Bank of Philadelphia working paper 97-12
- Carlino G, Sill K (1997) Regional economies: separating trends from cycles. Federal Reserve Bank of Philadelphia: Business Review, pp 1–13
- Carlino G, Sill K (2001) Regional income fluctuations: common trends and common cycles. *Rev Econ Stat* 83(3):46–456
- Clark TO, van Wincoop E (2001) Borders and business cycles. *J Int Econ* 55(1):59–85
- Croux C, Forni M, Reichlin L (2001) A measure of comovements for economic variables: theory and empirics. *Rev Econ Stat* 83(2):232–241
- Dolado JJ, María-Dolores R (2001) An empirical study of the cyclical effects of monetary policy in Spain (1977–1997). *Investigaciones Económicas XXV(1)*:3–30
- Dolado JJ, Sebastián M, Vallés M (1993) Cyclical patterns of the Spanish economy. *Investigaciones Económicas I. XVII(3)*:445–473
- Doménech R, Gómez V (2005) Ciclo económico y desempleo estructural en la economía española. *Investigaciones Económicas XXIX(2)*:259–288
- Forni M, Haillin M, Lippi M, Reichlin L (2004) The generalized dynamic-factor model: identification and estimation. *Rev Econ Stat* 82(4):540–554
- Gadea MD, Gómez-Loscos A, Montaños A (2006) How many regional business cycles are there in Spain? A MS-VAR approach. Documento de trabajo Fundear (Serie Investigación), 27
- Garrison CB, Chang HS (1979) The effect of monetary and fiscal policies on regional business cycles. *Int Reg Sci Rev* 4(2):167–180
- Geweke J (1977) The dynamic factor analysis of economic time series. In: Aigner DJ, Goldberger AS (eds) *Latent variables in socio-economic models*. North-Holland, Amsterdam
- Giannone D, Reichlin L, Sala L (2004) Monetary policy in real time. NBER Macroeconomics Annual, pp 161–200
- Gómez V, Maravall A (1996) Programs TRAMO and SEATS. Instructions for the User. Working paper 9628, Research Department, Bank of Spain (with several updates)
- Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384
- Hamilton JD (2001) Comment on “A comparison of two business cycle dating methods”. Mimeo
- Hamilton JD, Owyang MT (2009) The propagation of regional recessions. Federal Reserve Bank of St. Louis working paper no. 2009-013
- Harding D, Pagan A (2002a) Dissecting the cycle: a methodological investigation. *J Monet Econ* 49(2):365–381
- Harding D, Pagan A (2002b) A comparison of two business cycle dating methods. *J Econ Dyn Control* 27:1681–1690
- Harding D, Pagan A (2003) Rejoinder to James Hamilton. *J Econ Dyn Control* 27:1695–1698
- Harding D, Pagan KD (2006) Synchronization of cycles. *J Econom* 132(1):59–79
- Kaiser R, Maravall A (2005) Combining filter design with model-based filtering (with an application to business-cycle estimation). *Int J Forecast* 21:691–710
- Kouparitsas MA (2002) Understanding US regional cyclical comovements: how important are spillovers and common shocks? Federal Reserve Bank Chicago *Econ Perspect* 24(4):30–41
- Kozlowski PJ (1995) Money and interest rates as predictors of regional economic activity. *Rev Reg Stud* 25(2):143–157
- Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J Empir Financ* 10(5):603–621
- Maravall A (1993) Stochastic Linear Trends: Models and Estimators. *J Econom* 56:5–37
- McKay A, Reis R (2008) The brevity and violence of contractions and expansions. *J Monet Econ* 55: 738–751
- Metzler LA (1950) A multiple-region theory of income and trade. *Econometrica* 18(4):329–354

- Mönch E, Uhlg H (2005) Towards a Monthly Business Cycle Chronology for the Euro Area. *J Bus Cycle Meas Anal* 2:43–69
- Owyang MT, Piger J, Wall HJ (2005) Business cycle phases in U.S. States. *Rev Econ Stat* 87(4):604–616
- Politis DN, Romano J (1994) The stationary bootstrap. *J Am Stat Assoc* 89:1303–1313
- Rodríguez G, VILLEMAIRE J (2004) Identifying and measuring Canadian regional business cycles using MS-VAR models. Mimeo, University of Ottawa
- Sargent TJ, Sims CA (1977) Business cycle modeling without pretending to have too much a-priori economic theory. In: Sims C et al (eds) *New methods in business cycle research*. Federal Reserve Bank of Minneapolis, Minneapolis
- Schumpeter JA (1939) *Business cycles: a theoretical, historical and statistical analysis of the capitalist process*, vol I. McGraw-Hill, New York. New Edition Porcupine Pr (November 1989)
- Stock JH, Watson M (2010, forthcoming). Dynamic factor models. In: Clements MP, Hendry DF (eds) *Oxford handbook of economic forecasting*. Oxford University Press
- Timm NH (2002) *Applied multivariate analysis*. Springer text in statistics. Springer, New York
- Watson M (1994) Business cycle durations and postwar stabilization of the U.S. Economy. *Am Econ Rev* 84(1):24–46
- Watson M (2004) Comment on Giannone, Reichlin and Sala. *NBER Macroeconomics Annual* 2004, pp 216–221