Discovering pervasive and non-pervasive common cycles

Guillermo Carlomagno*
Department of Statistics and Econometrics, University Carlos III of Madrid;
Centro de Investigaciones Económicas (CINVE); and Universidad ORT Uruguay
and
Antoni Espasa
Department of Statistics and Econometrics, Instituto Flores de Lemus,
U*niversity Carlos III of Madrid*

Abstract

The objective of this paper is to propose a strategy to exploit short-run commonalities in the sectoral components of macroeconomic variables to obtain better models and more accurate forecasts of the aggregate and the components. Our main contribution concerns cases in which the number of components is large so that traditional multivariate approaches are not feasible. We show analytically and by Monte Carlo that subsets of components in which all the elements share a single common cycle can be discovered by pairwise methods. As the procedure does not rely on any kind cross-sectional averaging strategy, it does not need to assume pervasiveness, it can deal with highly correlated idiosyncratic components and it does not need to assume that the size of the subsets goes to infinity. Nonetheless, the procedure works both, with fixed N and $T \to \infty$, and with $[T, N] \to \infty$.

Keywords: Common cycles, Factor Models, Disaggregation, Pairwise tests.

JEL: C01, C22, C32, C53.

1 Introduction

There is a clear tendency among statistical offices around the world to produce more disaggregated information, both at the regional and sectoral level. The reason for this must be that decision makers need to analyze the disaggregates to make better decisions. However, the usual macroeconometric analyses that focuses on modeling and forecasting economic aggregates (e.g. GDP, CPI, industrial production, employment, etc) do not make full use of the large amount of information contained in the disaggregates.

The objective of this paper is to propose a strategy to exploit short-run commonalities in the sectoral components of macroeconomic variables to model and forecast those components. Our main contribution concerns cases in which the number of components is large and traditional

*Correspondence to Guillermo Carlomagno. E-mail: gcarloma@est-econ.uc3m.es. Phone: (+34)916249674. Fax: (+34)916249847.

multivariate approaches are not feasible. As a byproduct, this approach allows to construct an indirect forecast of the aggregate, that may be more accurate than that obtained by direct procedures, as it uses more information and includes restrictions between the components which could palliate the curse of dimensionality.

The presence of commonalities in the short-run dynamics of macroeconomic variables has been extensively documented. Referring to the movements of macroeconomic variables around their trends, Lucas (1977) points out: '*Output movements across broadly defined sectors move together,... Prices generally are pro-cyclical*'. Long Jr and Plosser (1987) develop a theoretical model to explain commovments in sectoral output and Engle and Issler (1995) derive the reduced form of that model. Engle and Issler also state the conditions for the existence of common cycles. After Engle and Kozicki's (1993) seminal article on testing for common features, several empirical studies testing for short-run commonalities among macroeconomic series appeared in the literature. Engle and Kozicki (1993) themselves find international commonalities in GNP data of seven OECD counties; Vahid and Engle (1993) find common cycles among four regional per capita incomes in the US; and Engle and Issler (1995) find common cycles for eleven sectoral outputs of the US economy. Using techniques for detecting common cyclical features, Candelon et al. (2005) study financial contagion during the 1997 Hong Kong stock market crises. By estimating thirteen bi-variate VAR models they find evidence of contagion. Hecq et al. (2006) find common cyclical features among the GDP of five Latin American countries; and Cubadda (2007) finds commonalities in the short run movements of four monthly indicators that The Conference Board uses to build the composite coincident indicator of the business cycle in the US.

The study of common cyclical features in the components of macroeconomic series is relevant not only for understanding the interrelations among them, but also for constructing better empirical models and obtaining more accurate forecasts. Vahid and Issler (2002) analyze the importance of the restrictions implied by common cyclical features for forecasts, impulse-response functions, and variance-decomposition analysis of economic time series. As they argue, the reduction in the number of parameters of typical macroeconomic VAR models derived from the existence of common cycles can be substantial, and much larger than that implied by cointegration. Therefore, remarkable efficiency gains can be obtained by imposing — correct — common cycles restrictions, from which forecasting accuracy improvements could follow. In a Monte Carlo study, Vahid and Issler (2002) confirm that reduced rank models can lead to significant

forecasting accuracy improvements with respect to unrestricted models.

As Espasa and Mayo-Burgos (2013) show, an important characteristic of sectoral components of macro variables is that they can be grouped in relatively small subsets in which all of the series show the same short-run dyanimc behavior. The existence of those subsets translates into relevant short-run restrictions in the parameters of econometric models and can be exploited to reduce estimation uncertainty and, hopefully, to obtain more accurate forecasts. Espasa and Mayo-Burgos (2013) suggest a pairwise procedure to construct those subsets and **?** study the asymptotic properties of a similar strategy for discovering subsets of components with common trends. Castle and Hendry (2010) also point out the importance of including short-run common features restrictions in the individual models for the components in line with Mayo and Espasa (2009)[1].

In this paper we show that subsets of components that share single common cycle restrictions can be discovered by pairwise procedures, similar to those suggested by Espasa and Mayo-Burgos (2013). Our contributions concern the analysis of the procedure's asymptotic properties and a Monte Carlo study in which we confirm the large samples properties and study the short run behavior.

The pairwise strategy consists of testing for common cycles in all the $N(N-1)/2$ pairs that exist among the N components of an aggregate, and then, constructing subsets in which all of the pairs share a unique common cycle. Once these subsets are discovered, the restrictions that they imply for the short-run dynamic behavior of the components can be included in single-equation models for them. These models could be consistently estimated by OLS.

An alternative way to try to discover common cycles between the components of an aggregate could be the estimation of Dynamic Factor Models (DFM). However, when the cycles are non-pervasive (i.e., they are common to a reduced group of components only), one of the assumptions required by the usual estimation procedures (see e.g., Assumption B in Bai (2003) or Assumption A1 in Doz et al. (2012)) is violated, and therefore, these procedures are expected to perform poorly. Though several approaches to deal with this issue have been proposed in the DFM literature, most of them assume they know beforehand which series are affected by which factor (see, e.g., Karadimitropoulou and León-Ledesma (2013), Moench et al. (2013), Breitung and Eickmeier (2015)).

Bailey, Kapetanios, and Pesaran (2015) (BKP, hereafter), Bailey, Holly, and Pesaran (2015)

[1]This working paper was later published as Espasa and Mayo-Burgos (2013).

(BHP, hereafter), and Ando and Bai (2015), work with unknown non-pervasive structures. Ando and Bai deal with stationary series with a grouped factor structure and develop a procedure to determine the series in each group and estimate the factors. Every series is assumed to belong to some group, the size of the groups is assumed to go to infinity, and the usual restrictions of DFM on the cross-correlation of idiosyncratic components (see e.g., Assumption B in Bai (2003) or Assumption A2 in Doz et al. (2012)) are required. BKP propose a measure for the degree of non-pervasiveness of the factors, and BHP develop a two-stage procedure for dealing with pervasive and non-pervasive factors at the same time. Similarly to Ando and Bai (2015), these authors restrict their attention to stationary series, the cross-sectional dimension going to infinity, and also need the usual restrictions of DFM on the cross-correlation of idiosyncratic components.

Our procedure is more general than the previous ones in four aspects: First, we do not assume that all series belong to some group. Second, as we do not rely on any cross-sectional averaging method, we do not need to assume that the number of components (N) goes to infinity. Our theory only requires $T \to \infty$, N may be fixed or it may also go to infinity. Not relying on cross-sectional averaging methods gives our procedure a third advantage, namely, we do not need to restrict the cross-correlation of idiosyncratic components. Finally, though in this paper we focus on $I(0)$ series, as we argue in §6, a generalized version of our procedure is applicable both when series are $I(0)$ and $I(1)$. In the later case we do not need to differentiate.

These advantages are relevant when dealing with the components of macro variables. Working with the components of three different CPIs, Espasa and Mayo-Burgos (2013) find relatively small groups of components that share single common cycles. Therefore, the assumptions that the factors are pervasive, that every series belong to some group and the size of the groups goes to infinity, do not fit this framework.

There is, however, one aspect in which our procedure is lees general than the DFM alternatives described above. While in previous procedures the subsets may have any number of factors, our main focus is on subsets with single common cycles. Though this data structure may be reasonable when dealing with the components of a macro variable, it might be not when working with several variables which are not the components of the same aggregate. To palliate this issue, our approach can be generalized to cases in which the subsets may have two common cycles; one '*general*' and another '*sectorial*' one, in §6 we describe this generalization.

The rest of the paper is organized as follows. In §2 we give the precise definitions of *common*

cycles that are used along the paper. In §3 we state the required assumptions and study the properties of our proposal. In §4 we describe the algorithm for applying the procedure. §5 is devoted to the Monte Carlo experiments. Finally, in §6 we discuss some possible extensions and §7 concludes the paper.

2 Definitions

Before describing our proposal and its statistical properties in detail, in this section we give a more precise definition of the concepts of *common cycles* that will be used along the paper.

Engle and Kozicki (1993) is the seminal article in the literature of testing for common cycles. The authors generalize the concept of common trends to other possible common features (serial correlation, heteroscdasticity, excess kurtosis, etc), and state that a certain feature is said to be common if a non-zero linear combination of a multivariate time series fails to have the feature even though each of the series individually has it. Specifically, the authors define the presence of a *serial correlation common feature* (*SCCF*, hereafter), when a linear combination of serially correlated time series is an innovation with respect to the past of the series.

Assume that the data can be represented by a VAR:

$$X_t = \sum_{i=1}^{k} \Pi_i X_{t-i} + \epsilon_t, \tag{1}$$

where X_t is a $N \times 1$ vector and ϵ_t a *iid* N-dimensional process. The existence of a *SCCF* requires the existence of a $N \times s$ full column rank matrix δ such that $\delta' X_t$ does not present serial dependence on the past of X_t, what implies $\delta' \Pi_i = 0$, for all $1 \leq i \leq k$. Therefore, we can write $\Pi_i = \delta_\perp \psi_i'$ where $\delta_\perp$ is the orthogonal complement of δ (i.e, $\delta' \delta_\perp = 0$), and the VAR model can be rewritten as:

$$X_t = \delta_\perp \Psi' [X_{t-1}', ..., X_{t-k}']' + \epsilon_t, \tag{2}$$

where Ψ' is a full column rank matrix of dimension $N - s \times Nk$, such that $\delta_\perp \Psi' = [\Pi_1, ..., \Pi_k]$.

In this case, δ contains the *serial common correlation vectors*, and $\Psi' [X_{t-1}', ..., X_{t-k}']'$ are the common cycles. Since all the $\Pi_i' s$ have a left null space that includes δ, the rank of δ is the rank of the left null space of Π_i. Thus, Π_i has rank $N - s$, for $1 \leq i \leq k$.

The test for the existence of $N - s$ *SCCF* is a test for s zero canonical canonical correlations between X_t' and $[X_{t-1}', ..., X_{t-k}']'$:

$$C(p,s) = -(T-k-1)\sum_{i=1}^{s} ln(1-\lambda_i), \tag{3}$$

where λ_i $(i = 1, ..s)$ are the s smallest eigenvalues in the canonical correlation problem; T is the sample size; and k is the number of lags in the model. Under the null (the smallest s eigenvalues are zero), the statistic has a Chi-squared distribution with $(s^2 + sNk + sr - sN)$ degrees of freedom.

Engle and Kozicki (1993) develop their methodology for stationary variables, thus, in most of the cases, the analysis must be carried out for the differenced variables, leading to informational losses if cointegration relationships exist. Vahid and Engle (1993) extend the framework to I(1) cointegrated systems by proposing a procedure for estimating $SCCF$ vectors given the existence of common trends. Interestingly, the authors show that the presence of $SCCF$ among the first differences of I(1) cointegrated variables is equivalent to the existence of common cycles in the sense of Beveridge and Nelson (1981). This is the reason why the concepts of *common cycles* and *common serial correlation* can be used interchangeably.

A natural extension of $SCCF$ in cointegrated series is to allow the possibility that the $SCCF$ vectors cancel the short-run dynamics, but are not related in a particular way with the long-run pattern of the series. That is, there could exist a linear combination of the differenced series that is an innovation with respect to the past, but only after adjusting for the equilibrium deviations. This is the concept of *weak form of serial correlation common features* (WF) introduced by Hecq et al. (2006). As mentioned in the introduction, in this paper we focus on $I(0)$ variables, but our proposal can be generalized to the case of $I(1)$ with cointegration. In this case, we can deal both with with $SCCF$ and with WF structures.

Another interesting extension of the $SCCF$ concerns cases when the commonalities are not contemporaneous. In a comment to Engle and Kozicki (1993), Ericsson (1993) argues that a common correlation feature may exist in a multivariate time series, but it does not need to be contemporaneous as SCCF require. To deal with this possibility, Cubadda and Hecq (2001), introduce the concept of *polynomial serial correlation common feature* $(PSCCF)$. Though all the results of this paper can be generalized for the case of PSCCF (see §6), in order to keep things simple, we leave the implementation for future research.

3 Properties of the pairwise approach

The strategy of testing for common cycles between all possible pairs of components and then forming *single-cycle subsets* in which all pairs show a common cycle relies on the 'transitivity' of common cycles. That is, it should be the case that if series A_t and B_t share the cycle, and series A_t and C_t share the cycle, one can conclude that B_t and C_t also have the same cycle. In appendix A we show that $SCCF$ structures are transitive.

3.1 Assumptions

Our general framework can be summarized in four assumptions:

Assumption A *The N components are generated by the VAR in eq. (1), which may be generalized to include outliers and/or location shifts.*

Assumption B *The residuals of eq. (1) are iid and normally distributed.*

Assumption C *There is, at least, one subset of SC components (with $2 \leq SC \leq N$) that share a single common cycle.*

Assumption D *X_{it} is serially correlated for $i = 1, ..., N$.*

Assumption E *$N \times T^{-1/2} \to \leq c$, for some fixed constant c.*

Assumption B is necessary for the maximum likelihood procedures for testing for common features. Assumption C gives relevance to our objective of discovering single-cycle subsets, and assumption D rules out the 'trivial' common cycles that will appear if some components are white noise. Finally, as we argue below, assumption E is required to control false discoveries when we let N to go to infinity.

Remark 1 *Instead of assumption D we could require that at least sc components satisfy it. This flexibilization would require testing the significance of the estimated common cycles' coefficients.*

Remark 2 *Assumption B is related to the residuals, not to the components. Although this distinction is not relevant for the Monte Carlo experiments, it is important for empirical applications. By allowing for outliers and location shifts in the model, it would not be necessary to assume normality of the processes X_{it}. The only requirement is that normality can be achieved after correcting for a few outliers and location shifts which, as Juselius (2015) argues, is a quite general assumption in macro-economic VAR models.*

3.2 Statistical properties

Define SC_j as a subset in which all of the series share a $SCCF$ (we will use SC_j both, as the name of the subset, and to indicate the number of series inside it). Abusing notation, we will denote $\hat{SC}_j$ the subset estimated with the pairwise procedure.

The properties of the pairwise procedure for discovering *single-cycle* subsets must be evaluated in three dimensions: i) *Potency:* The proportion of correct series that are included in $\hat{SC}_j$. ii) *Gauge:* The proportion of wrong series that are included in $\hat{SC}_j$. iii) *False discovery:* The discovery of nonexistent *single-cycle* subsets[2].

3.2.1 Potency

In order to include all of the correct series in $\hat{SC}_j$ we should find a single cycle in all of the $SC_j(SC_j - 1)/1$ pairs that exist in the true subset. This implies not rejecting the hypothesis $s > 0$ *vs.* $s = 0$ for each pair.

If we were testing a single hypothesis, the probability of not false rejecting the null would be $1 - \varphi$ (with φ being the nominal size of the individual tests). When m tests are performed, if they are independent, the probability of not making any false rejection reduces to $(1-\varphi)^m$, and the probability of making at least one error is $1 - (1 - \varphi)^m$, which rapidly increases with m.

In our case of interest, $m = SC_j(SC_j-1)/2$ may be quite large. Thus, if tests are independent, the probability of including all of the correct series in $\hat{SC}_j$ will be close to zero. Simulation results (available upon request) show that, under some circumstances, common cycles tests between the series in SC_j may be independent. This means that the probability of including all of the correct series in $\hat{SC}_j$ may be a fast decreasing function of the number of series in the true subset. This is an undesirable property for our procedure.

To mitigate this problem we exploit two facts. First, since tests are transitive, and each series enter in several pairs, we could infer the correct result for one pair using the results of other ones — i.e., in the example at the beginning of this section, we could infer the existence of a common cycle between B_t and C_t, given that it exists between A_t and B_t and A_t and C_t. Our strategy for exploiting the transitivity consists of, instead of requiring that each series in $\hat{SC}_j$ passes a test for a common cycle with *all* of the other series in the subset, we let one series to enter in $\hat{SC}_j$ when it passes the test with *almost all* the other series in the subset. We call this strategy *relaxation strategy* (see step iii of the algorithm in §4).

[2]The terms *gauge* and *potency* are borrowed from Castle et al. (2011).

Second, since asymptotic power of common cycles tests is 1 (the probability of finding $s > 0$, when $s = 0$, goes to zero as T goes to infinity), for finite N, the *relaxation strategy* is asymptotically costless in terms of *gauge*.

The relevant question now is how this relaxed procedure is expected to perform in terms of *potency*. When performing the $N(N-1)/2$ tests for the whole set of components, the asymptotic probability of not rejecting the null of $s > 0$ for each individual pair formed by two series of SC_j, is $1 - \varphi$. For any other pair this probability is zero. Thus, a natural way to see the problem of finding the largest *single-cycle subset* is in terms of the random graph theory.

In simple words, a random graph can be seen as a square and symmetric matrix with zeros and ones in which each cell has probability p of having a one and $1-p$ of having a zero, independently of the other cells. When the cell (i, j) has a one, we say that there is an *edge* between units i and j and they are *connected* (for a detailed analysis of random graphs see Newman (2009)). In our case, the symmetric matrix is $N \times N$ and cell (i, j) corresponds to the pair formed by series i and j. Thus, ones would appear in pairs for which a common cycle was found. Though when testing for common cycles the edge probabilities may not be independent, the case of independence is the worst possibly one for having high potency, so there is no risk in keeping this assumption for analyzing the potency of our procedure.

Additionally, as the asymptotic probabilities of having an edge between pairs in which the series do not belong to the same SC_j are zero, we can focus on the sub-graphs formed by the series in a particular SC_j.

Finding the largest *single-cycle subset*, is equivalent to finding the largest *almost fully connected* subgraph — i.e., the largest subgraph in which *almost* all possible edges are present. This is closely reltated to the maximum *clique problem* described in the random graph literature (see, *interalia*, Matula (1976); Derényi et al. (2005); and Newman (2009)).

To get an intial idea about the potency of our procedure, we use the results in Derényi et al. (2005). The authors find the minimum edge probability for which all elements of a graph of size SC_j will be almost surely connected with, at least, $k - 1$ other elements. This probability is given by:

$$p_c(k) = \frac{1}{[(k-1)SC_j]^{\frac{1}{k-1}}} \tag{4}$$

Using eq. (4), we can fix a proportion ρ and chose $k^* = \rho(SC_j - 1) + 1$, so that eq. (4) will give

the edge probability required to find a subset of size SC_j, in which, for each series, a common cycle will be found with, at least, $\rho(SC_j - 1)$ of the remaining series.

Figure 1 shows this threshold probability for different alternatives of ρ and SC_j. These probabilities can be seen as the required magnitude for $(1 - \varphi)$ for finding the *almost fully connected* graph we are looking for (recall that φ is the nominal size of the individual tests). For instance, for $SC_j = 100$ and $\rho = 0.9$, we would need $(1 - \varphi) = 0.9$, meaning that with a φ smaller than or equal to 0.1 we would find the almost fully connected subset we are looking for almost surely.

Conversely, if we stick to the strict full connection criteria, under independence, the probability of including all the correct series in the estimated SC_j would be $(1 - \varphi)^{4950}$, which is virtually zero even for small values of φ [3]. This implies that relaxing the requirement from *full connection*, to *almost full connection* may lead to a great improvement in the probability of including all the correct series in the estimated SC_j.

As aforementioned, when N is finite the relaxation to almost full connection is asymptotically costless in terms of gauge. In short samples (short T) or when $N \to \infty$, larger values of ρ lead to higher gauges. The case of $N \to \infty$ is studied in §3.2.2 and the short samples behavior in §5. In the former case we need assumption E for the procedure to work properly and in the later we get some deterioration, but results are still good.

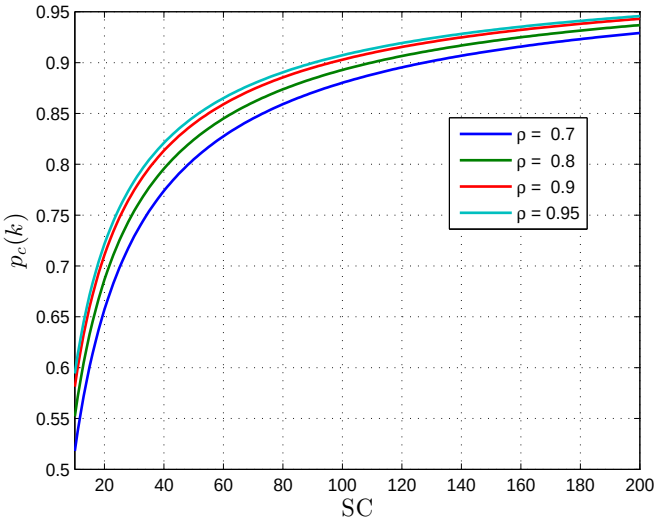

Figure 1: Percolation probabilities $p_c(k)$, for $k = \varphi(SC - 1) + 1$

Note however that, since Derényi et al. (2005) assume $SC_j \to \infty$ and k fixed, expression 4

[3] Equivalently we would need $\varphi = 1 \times 10^{-5}$ in other to have a 95% probability of finding the true SC.

is valid only asymptotically — large $(SC_j - k)$. Thus, as we are interested in large values of ρ, it may be a rough measure when SC_j is small. To better understand the properties of the relaxation procedure when SC_j is fixed and $T \to \infty$, we perform a small simulation study.

In each experiment we simulate a random graph of size SC_j with independent edge probability p. This is done by generating a square symmetric matrix with zeros and ones in which each cell has probability p of having a one and $1 - p$ of having a zero, independently of the other cells. The asymptotic probability of finding a common cycle between two series that truly have it is $1 - \varphi$. Hence, fixing p equal to $1 - \varphi$ replicates the random graph that would be obtained when testing for common cycles between all the pairs among SC_j series that share a unique common cycle, when $T \to \infty$, and using a nominal size of φ for each test.

We consider two values of p (0.95 and 0.99) and compute the average relative size of the estimated SC ($mean(\hat{SC}/SC)$) and the probability of including *all* the elements in $\hat{SC}$ ($mean(I_{\hat{SC}=SC})$), both when applying the strict full connection criteria and when relaxing it.

Then, we apply the algorithm described in §4 to find the largest *almost fully connected* subset (in that algorithm, the relaxation parameter λ is defined as the maximum number of tests that a series can fail and still enter in $\hat{SC}$). Results are included in Table 1. As a general conclusion: the relaxation procedure allows to reach acceptable results even when regular significance levels are used. For example, with a nominal significance level $\varphi = 0.05$, when the true dimension of the set is 50, the expected ratio *'size of the estimated subset* to *true size of the subset'* is 0.99. As this result is in line with the 0.49 probability of including *all the correct series* that the table shows, we have that while i 49% of the experiments we include the 50 series, in the remaining 51% we include 49.

In summary, when T goes to infinity, our procedure is expected to have high potency, regardless the size of SC_j.

3.2.2 Gauge and false discovery

To include wrong series in some $\hat{SC}_j$ or to discover nonexistent subsets, common cycles tests should lead to conclude $s = 1$, when the true hypothesis is $s = 0$, i.e., not to reject the false null hypothesis of $s > 0$.

Focus first on *gauge*. Let X_{out} be a series outside SC_j and let W_i be the event of wrongly not rejecting $s > 0$ with the i^{th} series in the estimated SC_j ($\hat{SC}_j$). As for wrongly including X_{out} in $\hat{SC}_j$ we need to find a common cycle with $\rho\hat{SC}_j$ series, we need to wrongly reject $\rho\hat{SC}_j$

Table 1: Full connection *vs.* almost full connection

		SC=5	**SC=10**	**SC=20**	**SC=25**	**SC=40**	**SC=50**
				$\boldsymbol{p = 0.95}$			
Full contection	mean($\hat{SC}/SC$)	0.92	0.83	0.73	0.69	0.61	0.56
	mean($I_{\hat{SC}=SC}$)	0.62	0.10	0.00	0.00	0.00	0.00
Almost Full conection	mean($\hat{SC}/SC$)	0.99	0.95	0.96	0.98	0.97	0.99
	mean($I_{\hat{SC}=SC}$)	0.94	0.56	0.40	0.56	0.28	0.49
				$\boldsymbol{p = 0.99}$			
Full contection	mean($\hat{SC}/SC$)	0.98	0.96	0.92	0.90	0.86	0.83
	mean($I_{\hat{SC}=SC}$)	0.90	0.61	0.16	0.04	0.00	0.00
Almost Full conection	mean($\hat{SC}/SC$)	1.00	1.00	1.00	1.00	1.00	1.00
	mean($I_{\hat{SC}=SC}$)	1.00	0.97	0.98	1.00	1.00	1.00

- The relaxation parameter (λ) is 1 for $SC = 5$ and $SC = 10$; 2 for $SC = 20$; and 5 for $SC = 50$.
- '*mean*' denotes the mean across experiments.
- The number of experiments is 1000.
- $\hat{SC}$ is the number of series included in the largest (almost) fully connected subset.
- $I_{\hat{SC}=SC}$ denotes the indicator function that takes the value 1 if $\hat{SC} = SC$ and 0 otherwise.

hypotheses. The probability of this event is $P(W_1 \cap ... \cap W_{\rho\hat{SC}_j})$, and can be factorized as:

$$P(W_1 \cap ... \cap W_{\rho\hat{SC}_j}) = P(W_1|W_2, ..., W_{\rho\hat{SC}_j}) \times ... \times \ P(W_{\rho\hat{SC}_j-1}|W_{\rho\hat{SC}_j}) \times P(W_{\rho\hat{SC}_j}), \quad (5)$$

where $P(W_i)$ is the marginal probability of not rejecting the false hypothesis $s > 0$, which, as asymptotic power is 1, converges in probability to zero as T goes to infinity. Using the extreme assumption that all the $\rho\hat{SC}_j - 1$ conditional probabilities in eq. (5) are equal to 1, the probability of wrongly including X_{out} in $\hat{SC}_j$ would be equal to $P(W_i)$ which goes to zero as T goes to infinity.

Define now Z_h as the random variable that takes the value one if the variable h is wrongly included in $\hat{SC}_j$ and zero otherwise. Then, the number of wrong inclusions is:

$$\sum_{h=1}^{N-SC_j} Z_h. \quad (6)$$

Therefore, the expected number of wrong inclusions is:

$$E[\textstyle\sum_{h=1}^{N-SC_j} Z_h] = \sum_{h=1}^{N-SC_j} E[Z_h] = (N - SC_j)E[Z_h], \quad (7)$$

and the expected proportion of wrong elements in the estimated single-cycle subset (ω) is:

$$E[\omega] = \frac{(N - SC_j)E[Z_h]}{S\hat{C}_j} \tag{8}$$

Under the extreme assumption that all conditional probabilities in eq. (5) are equal to one, $E[Z_h] = P(W_h)$. Since $P(W_h)$ goes to zero as $T \to \infty$, using assumption D, $plim_{T\to\infty} E[Z_h] = 0$. When N is fixed and T goes to infinity, eq. (8) goes to zero at the same rate as $P(W_h)$ does. When both N and T go to infinity, for avoiding eq. (8) to grow without limit, we need N/SC_j to grow at a rate smaller than to equal to that of $1/P(W_h)$. Since this condition includes cases in which the common cycles are not pervasive, i.e., $N/SC_j \to \infty$, we will have good gauge properties even in that case.

This argument can also be made without relying on the asymptotic power of the common cycles tests. For doing that we need to change the extreme assumption that all conditional probabilities in eq. (5) are equal to one, for the following assumption:

Assumption F *There is a proportion γ of the $\rho\hat{SC}_j - 1$ conditional probabilities in eq. (5) that do not exceed a fixed threshold p_{max}, with $0 \leq p_{max} < 1$. The proportion γ is assumed to satisfy the condition $\gamma_{min} \leq \gamma \leq 1$, with γ_{min} being some fixed value larger than zero.*

With this new assumption, an upper bound for $E[Z_h]$ will be $\overline{E[Z_h]} = p_{max}^{\gamma(\rho\hat{SC}_j - 1)} P(W_h)$. Thus, an upper bound for the expected proportion of wrong elements in the estimated single-cycle subset would be:

$$\overline{E[\omega]} = \frac{(N - SC_j)p_{max}^{\gamma(\rho\hat{SC}_j - 1)} P(W_h)}{S\hat{C}_j} \tag{9}$$

Proposition 1 *Under assumption F, a sufficient condition for eq. (9) not to grow, as $N \to \infty$, is $SC_j/log(N) \to \geq \kappa$, where κ is some positive constant.*

Proof *See appendix B*

Again, we do not need pervasiveness in the sense of DFM.

Consider now the problem of *false discovery*. Matula (1976) shows that the size of the maximal fully connected sub-graph (*maximal clique*) in a random graph with M elements and edge probability p has a strong peak around $2log(M)/log(1/p)$. In our case, $M = N - \sum_j SC_j$, and $p \to 0$ as $T \to \infty$. Thus, for fixed N the size of the largest *false fully connected* subset goes to

zero. When $N \to \infty$ we need $N - \sum_j SC_j$ to grow at a smaller rate than that of $1/p$. Given $1/p$ is $O_p(T)$, this implies $N - \sum_j SC_j$ can grow, at most, at the same rate as T does.

Since arguments based on random graph theory require independent edge probabilities they may not be adequate in our case. Another argument, that do not require independence is as follows. Define $M = N - \sum_j SC_j$ as the number of series which do not belong to any single cycle subset; $M^* = \frac{M(M-1)}{2}$ as the number of pairs between those series; and Y_m as a random variable that takes the value 1 if a common cycle is wrongly found for a pair m. Using the same reasoning as that used for expressions 7 and 8, the expected number of false discoveries (number pairs for which a common cycle is wrongly found) is $M^* \times E[Y_m]$.

Since $E[Y_m] = p$ (with p being the probability of wrongly finding a common cycle in one of the M^* pairs), and $p \to 0$ as $T \to \infty$, for finite N, the expected number of false discoveries tends to zero. When $N \to \infty$, a sufficient condition having a fixed expected number of false discoveries is $M^* \times p \to 0$, what requires $N - \sum_j SC_j$ to grow a rate smaller than or equal to that of $p^{-1/2}$. Which implies that $N - \sum_j SC_j$ can grow, at most, as the same rate as $T^{1/2}$ does (see assumption E).

3.3 Partial models

The pairwise strategy consists of testing for a common cycle in all possible pairs of series and look for the largest subset in which *almost all* the pairs have the cycle. This strategy requires estimating partial models and it could be thought to imply a power loss with respect to a 'complete' model approach (when feasible). To analyze this issue we performed a small simulation study to compare the ability of the pairwise vs. the full model approach when the latter is feasible (small N). Since our focus in this paper is relatively large N, we do not report the results (available upon request). The main conclusion of these experiments is that, when common cycles are pairwise detectable, nothing is lost by proceeding in a pairwise fashion. On the contrary, important power gains for finding the true number of common cycles in short samples can result from this procedure with respect to the full model approach.

4 Algorithm

In order to discover subsets (SC_j) in which all the series share a single cycle we proceed in six steps:

i. Perform common cycles tests between all possible pairs of components, store the resulting p-values, and construct a $N \times N$ boolean adjacency matrix, A, that contains a 1 in the cell

(i,j) if the corresponding pair has a common cycle (the null of $s = 0$ has not been rejected) and zero otherwise.

ii. Find the maximal clique on A using, for example, the Bron-Kerbosch algorithm (see Bron and Kerbosch (1973)). The maximal clique is defined as the largest subgraph in which all nodes are pairwise connected (see also, Bollobás and Erdös (1976)). We rename the maximal clique as *single cycle subset*, $\hat{SC}_1$.

iii. Define the relaxation parameter ($1 \leqslant \lambda < \hat{SC}_1$, with $\hat{SC}_1$ being the number of series in the estimated single cycle subset) to identify the candidates to enter in the *almost fully connected* subset. A series outside the original subset is a candidate if it satisfies two conditions:

   (a) a single cycle — at the original φ of confidence — is found with at least $\hat{SC}_1 - \lambda$ of the series already in the subset $\hat{SC}_1$.

   (b) when the nominal size of the test is relaxed to φ^* the candidate has a common cycle with all the series already in the subset $\hat{SC}_1$.

iv. Construct the set of candidates C_0. If all the candidates have a common cycle between each other (at the original φ), let all of them in $\hat{SC}_1$ and go to point vi (because there are not more potential candidates). If not, find the maximal clique (see point ii) inside C_0 and let in $\hat{SC}_1$ all the series in the maximal clique. Note that after including these series there could still remain some potential candidates, so check for this possibility; construct a new set of candidates C_1 using conditions a and b above, and repeat the procedure in this point.

v. If, according with conditions a and b, there are no candidates that share the cycle, try to include them sequentially starting with the one which is has a common cycle with more of the series already in the subset. In case of conflict (there are candidates that share the cycle with the same number of variables already in the subset), use the p-values stored in step i to decide. An adhoc criteria could be, for example, to include the series whose sum of p-values for the null $s > 0$ is the minimum.

vi. Repeat steps ii to v but excluding the series already included in some almost fully-connected subset.

5 Simulations

In this section we perform some Monte Carlo experiments to fulfill two objectives: confirm the asymptotic properties studied in §3.2, and analyze the small sample properties of the procedure.

5.1 Design of the experiments

We consider two alternative DGPs. Both of them have the same general structure:

$$X_t = c + \Pi X_{t-1} + \epsilon_t, \tag{10}$$

where $\epsilon_t \sim N(0, \Sigma)$; and the roots of $det(I - \Pi L)$ are all outside the unit circle. We want to simulate situations in which only a subset SC of series share a single cycle and there are no more common cycles restrictions in the system. In order to simplify the dynamics of the systems, matrix Π has the following structure:

$$\begin{bmatrix} A_{SC\times SC} & 0 \\ 0 & D_{N-SC\times N-SC} \end{bmatrix},$$

where $A = \delta_\perp^* \Psi^{*'}$, with $\delta_\perp^*$ and Ψ^* are $SC \times 1$ vectors, and D is a diagonal matrix. This does not imply that series outside SC are independent between each other, or with respect to series inside SC, as Σ is not necessarily diagonal. Partition the vector X_t in its first SC elements and the remaining $N - SC$, and call X_t^{sc} to the first sub-vector. Then, the common cycle is $\Psi^{*'} X_{t-1}^{sc}$, and $\delta_\perp^*$ contains the coefficients of the common cycle in each of the first SC series.

Since we want X_t in eq. (10) to be stationary, we need the roots of the characteristic polynomial $det(I - \Pi L)$ to be outside the unit circle. Calling π_{ij} the elements of matrix Π, it can be shown that, after setting $|\pi_{ii}| < 1$, $\forall\, i \geq sc + 1$, the stationarity condition of eq. (10) is:

$$|\sum_{i=1}^{sc} d_i \psi_i| < 1.$$

where d_i and ψ_i are the elements of $\delta_\perp^*$ and $\Psi^{*'}$, respectively.

There are infinitely many different possibilities of $\delta_\perp^*$ and Ψ^* that would satisfy the stationarity condition. Three of them that may be of interest are:

DGP 0 : $\delta_\perp^*$ is filled with uniform random values between 0.7, and 1 and Ψ^* is filled with uniform random values between $\frac{1}{1.2SC}$ and $\frac{1}{1.1SC}$.

DGP 1 : Same as DGP 0 but imposing z zeroes in Ψ^* so that we can change SC in the denominator for $SC - z$ and the non zero entries will be larger. We set $SC - z = 2$, so that the common cycles will be generated by two of the series in SC (those whose coefficients in $\Psi^{*'}$ are different from zero).

DGP 2 : Same as DGP 1 but allowing some negative values in $\delta_{\perp}^{*}$ (there will be some counter-cyclical variables). This allows to increase the non-zero values in Ψ^* with respect to option DGP 1. We impose negative coefficients to 20% of the variables in $\delta_{\perp}^{*}$, hence, the number of non zero coefficients in Ψ^* is $(2 + 0.2 \times SC)$.

Note that in DGP 0 all values of vector ψ will be rather small, even for relatively small SC. Thus, in order to statistically distinguish those values from zero, we would need quite large samples. To avoid this issue, we focus only on DGPs 1 and 2.

Finally, the residuals ϵ_t are generated as:

$$\epsilon_{it} = \eta_{i,t} + \sum_{j \neq 0,\, j=-Q}^{Q} \beta \eta_{i-j,t}, \tag{11}$$

where $\eta_t \sim N(0, I_N)$.

In this way, when $\beta \neq 0$ and $Q \neq 0$, the residuals of each equation are cross-correlated with other $2Q$ residuals. For avoiding higher cross-correlation inside SC than outside it, the position of insiders is randomly set, so they are not in positions 1 to SC.

For the two DGPs we consider three scenarios and three sample sizes, in all cases $N = 100$. In scenario 1 we set $SC = 10$; in scenario 2, $SC = 25$; and in scenario 3, $SC = 40$. The sample sizes are $T = 100$, $T = 200$, and $T = 400$.

For each DGP, scenario and sample size, we perform 500 Monte Carlo replications. In each replication we simulate a 100-dimensional model in which a subset of SC series share a single common cycle. Our objective is to discover the series in SC. To do that, we perform SCCF tests on all the 4950 bi-variate VAR sub-models that exist among the 100 series. Thus, for a particular DGP, scenario and sample size, we have 2.475 million sub-models (4950 for each replication). Since we have two DGPs, three scenarios, and three sample sizes, we have $(2 \times 3 \times 3) \times 2.475 = 44.55$ million sub-models to estimate. Additionally, since the lag length for each of the 4950 sub-models of a particular replication is unknown, we select it with the AIC, admitting between one and five lags, this multiplies the number of models by five. Furthermore, we consider four

alternative combinations for β and Q of eq. (11), which multiplies the number of models by 4.

5.2 Monte Carlo results

Table 2 includes the *gauge* and *potency* of the pairwise strategy for DGP 1. As Conclusions for DGP 2 are the same, details are omitted.

As expected from §3.2, the gauge of the pairwise procedure is close to zero for all scenarios and sample sizes, independently of whether residuals are cross-correlated or not. Results in terms of potency are also very good, as we get values above 0.9 in all cases (except for the case of $T = 100$, $SC = 40$, and independent residuals).

Table 2: Gauge and Potency of the Pairwise procedure. DGP 1 ($\lambda = 2$, $\varphi = 5\%$, $\varphi^* = 0.5\%$)

	$SC = 10$		$SC = 25$		$SC = 40$	
	Gauge	***Potency***	***Gauge***	***Potency***	***Gauge***	***Potency***
	$\beta = 0$, $Q = 0$ *(independent residuals)*					
$T = 400$	0.0	97.9	0.0	95.4	0.0	93.9
$T = 200$	0.0	97.2	0.0	94.0	0.0	92.2
$T = 100$	0.2	88.2	0.2	83.1	0.1	79.1
	$\beta = -0.3$, $Q = 10$ *(non zero corr. with 20 other residuals)*					
$T = 400$	0.0	97.1	0.0	96.1	0.0	94.8
$T = 200$	0.0	96.9	0.0	95.6	0.0	93.7
$T = 100$	0.1	93.3	0.1	91.6	0.0	89.3
	$\beta = -0.3$, $Q = 20$ *(non zero corr. with 40 other residuals)*					
$T = 400$	0.0	97.4	0.0	96.5	0.0	94.6
$T = 200$	0.0	96.3	0.0	95.6	0.0	94.2
$T = 100$	0.1	94.6	0.1	92.2	0.1	90.5
	$\beta = -0.3$, $Q = 30$ *(non zero corr. with 60 other residuals)*					
$T = 400$	0.0	97.2	0.0	95.2	0.0	94.6
$T = 200$	0.0	96.7	0.0	95.6	0.0	94.3
$T = 100$	0.1	95.2	0.1	93.4	0.1	91.0

- Number of experiments: 1000. y.
- $Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
- $Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
- Z_2 = number of wrong series included in $\hat{SC}$
- Z_1 = number of correct series included in $\hat{SC}$
- $Nexp$ = number of experiments

As discussed in the introduction, an alternative strategy to ours may the estimation of Dynamic Factor Models. As we argued, since we are dealing with non-pervasive factors, small sizes of the groups, and cross-correlated idiosyncrasies, DFM cannot be expected to show a good

performance. Still, as the DFM assumptions about pervasiveness, and cross-correlation of the residuals are asymptotic, it could be of interest to compare our approach with DFM alternatives.

We applied the usual Principal Components strategy and the QML approach of Doz et al. (2012), and grouped the series with statistically significant factor loadings. Results (not reported) are comparable to those of table 2 only for cases with $SC \geq 25$, $T \geq 200$, and $Q = \beta = 0$. This could be expected as those cases may satisfy the assumptions of pervasiveness and limited idiosyncratic cross-correlation. Small alterations of these conditions make the DFM alternatives to fail.

6 Possible extensions

In this section we describe some possible extensions of our procedure which are not implemented in this paper.

As mentioned in §2 the pairwise approach can be generalized both to consider $I(1)$ variables that may have cointegration relationships, and to allow non-contemporaneous short run commonalities. These extensions require considering WF and $PSCCF$ (see §2) instead of $SCCF$ structures. Though when considered in a pairwise fashion, those structures are not transitive, simple modifications of the testing strategy make them so.

The lack of transitivity of WF tests derives from the fact that cointegration relationships that are relevant for a series, say, X_i, and need to be considered in the WF tests that include that series, are not detectable in all pairs that contain X_i. For example, the cointegration relationship between X_i and X_j will not be detectable in the pair (X_i, X_h) — for $h \neq j$. Hence, when testing WF in the pair (X_i, X_h), this issue can be solved by including *all* of the cointegration relationships that are relevant for the two series, not only that one between X_i and X_h (a proof of this statement is available upon request).

This strategy is also valid for $PSSCF$, with the only modification that we should include the lags of all relevant variables instead of the cointegration relationships.

A third extension of interest is allowing subsets with more common cycles. In this paper we focused on the case that the data can be grouped in subsets in which series inside them share just one common cycle (as we argued we do not need these subsets to be large and there can be series which do not belong to any subset). As Espasa and Mayo-Burgos (2013) argue, this situation is a good approximation to the reality when dealing with the components of a macro variable. In fact, they show that the pairwise procedure leads to more accurate forecasts of

different CPIs than do other alternative methodologies, including Dynamic Factor Models.

Nonetheless, when dealing with a large data set of macro variables (not necessarily the components of a single one), the situation could be different. It is usual in the litterature to see a *general* factor that affects more or less all variables and *sectorial* factors that affect specific groups (see, e.g., Karadimitropoulou and León-Ledesma (2013), Moench et al. (2013), and Breitung and Eickmeier (2015)).

In this case, the pairwise procedure proposed in this paper will not be applicable. Since the only pairs that have common cycles are those formed by series with a single common cycle (e.g., series that have only the general factor and no sectorial one), the procedure will be unable to discover the 'true' data structure.

Our approach could be adapted to this situation, for what we need a new assumption:

Assumption G *in the set of N series there is a subset of series that just have the general cycle.*

Under this assumption we can proceed with the following algorithm:

(i) Apply the pairwise procedure proposed in this paper. Under assumption G, this will lead us to discover the subset of series that only have the *general* cycle — call it SC_1. (ii) Test for a common cycle in all of the triplets formed by one series inside $\hat{SC}_1$ and a pair of outsiders. For the triplets in which the outsiders have the same *sectorial* cycle, we will find two common cycles ($s = 1$). (iii) Construct a $(N - \hat{SC}_1) \times (N - \hat{SC}_1)$ symmetric *adjacency matrix* for the series outside $\hat{SC}_1$ such that each cell of this matrix represents a pair of the components outside $\hat{SC}_1$. Each of those pairs belongs to $\hat{SC}_1$ different triplets: one for each element of $\hat{SC}_1$. Then, in each cell of the adjacency matrix, put a 1 if *almost all* of the corresponding $\hat{SC}_1$ triplets have two common cycles; otherwise, put a 0. (iv) Look for maximal fully connected sub-graphs in the previous *adjacency matrix*. This would lead us to discover the general and the sectorial cycles.

7 Concluding remarks

This paper deals with the issue of discovering common cycles, which can be pervasive or non-pervasive, in a large set of disaggregates. We showed that, when focusing on groups in which the series have single common cycles, the discovery can be carried out in a pairwise fashion.

The strategy consists of testing for common cycles between all possible pairs of series and constructing groups in which *almost* all pairs showed a common cycle. Statistical properties of this procedure were studied both when N and $T \to \infty$ and when N is fixed and $T \to \infty$.

Theoretical results indicate that the pairwise strategy has good properties in both cases.

An interesting characteristic of our proposal is that it does not rely on any type of cross-sectional averaging method. This explains why we can deal with pervasive and non-pervasive common cycles, both when N is fixed and when it goes to infinity. Additionally, as we do not need idiosyncrasies to average out as N increases, we do not need to restrict idiosyncratic cross-correlation.

Monte Carlo results confirmed the theoretical analysis and showed a good performance in small samples.

Extensions of this paper include generalizations for $I(1)$ variables which may be cointegrated, non-contemporaneous short run commonalities, and the consideration of general and sectorial common cycles.

Acknowledgments

Both authors gratefully acknowledge financial support from the Spanish Government project ECO2012-32401.

References

Ando, T. and J. Bai (2015). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics*.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bailey, N., S. Holly, and M. H. Pesaran (2015). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2015). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*.

Beveridge, S. and C. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics* 7(2), 151–174.

Bollobás, B. and P. Erdös (1976). Cliques in random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 80, pp. 419–427. Cambridge Univ Press.

Breitung, J. and S. Eickmeier (2015). Analyzing business cycle asymmetries in a multi-level factor model. *Economics Letters 127*, 31–34.

Bron, C. and J. Kerbosch (1973). Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM 16*(9), 575–577.

Candelon, B., A. Hecq, and W. Verschoor (2005). Measuring common cyclical features during financial turmoil: Evidence of interdependence not contagion. *Journal of International Money and Finance 24*(8), 1317–1334.

Castle, J., J. Doornik, and D. Hendry (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics 3*(169), 239–246.

Castle, J. L. and D. F. Hendry (2010). Nowcasting from disaggregates in the face of location shifts. *Journal of Forecasting 29*(1-2), 200–214.

Cubadda, G. (2007). A unifying framework for analysing common cyclical features in cointegrated time series. *Computational Statistics & Data Analysis 52*(2), 896–906.

Cubadda, G. and A. Hecq (2001). On non-contemporaneous short-run co-movements. *Economics Letters 73*(3), 389–397.

Derényi, I., G. Palla, and T. Vicsek (2005). Clique percolation in random networks. *Physical review letters 94*(16), 160202.

Doz, C., D. Giannone, and L. Reichlin (2012). A quasi–maximum likelihood approach for large, approximate dynamic factor models. *Review of Economics and Statistics 94*(4), 1014–1024.

Engle, R. and S. Kozicki (1993). Testing for common features. *Journal of Business & Economic Statistics 11*(4), 369–380.

Engle, R. F. and J. V. Issler (1995). Estimating common sectoral cycles. *Journal of Monetary Economics 35*(1), 83–113.

Ericsson, N. (1993). Comment (on engle and kozicki,1993). *Journal of Business & Economic Statistics 11*(4), 380–383.

Espasa, A. and I. Mayo-Burgos (2013). Forecasting aggregates and disaggregates with common features. *International Journal of Forecasting 29*(4), 718–732.

Hecq, A., F. Palm, and J. Urbain (2006). Common cyclical features analysis in var models with cointegration. *Journal of Econometrics 132*(1), 117–141.

Juselius, K. (2015). Haavelmo's probability approach and the cointegrated var. *Econometric theory 31*(02), 213–232.

Karadimitropoulou, A. and M. León-Ledesma (2013). World, country, and sector factors in international business cycles. *Journal of economic dynamics and control 37*(12), 2913–2927.

Long Jr, J. B. and C. I. Plosser (1987). Sectoral vs. aggregate shocks in the business cycle. *The American Economic Review 77*(2), 333–336.

Lucas, R. E. (1977). Understanding business cycles. In *Carnegie-Rochester conference series on public policy*, Volume 5, pp. 7–29. Elsevier.

Matula, D. W. (1976). The largest clique size in a random graph. *Department of Computer Science, Southern Methodist University, Technical Report CS 7608*.

Mayo, I. and A. Espasa (2009). Forecasting aggregates and disaggregates with common features. Technical report, Universidad Carlos III de Madrid. Departamento de Estadística, Working Paper (later published as Espasa and Mayo-Burgos (2013)).

Moench, E., S. Ng, and S. Potter (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics 95*(5), 1811–1817.

Newman, M. (2009). *Networks: an introduction*. OUP Oxford.

Vahid, F. and R. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics 8*(4), 341–60.

Vahid, F. and J. V. Issler (2002). The importance of common cyclical features in var analysis: a monte-carlo study. *Journal of Econometrics 109*(2), 341–363.

Appendix A Transitivity of *common cycles*

In this appendix we show that $SCCF$ are transitive.

To see that $SCCF$ is transitive let X_1 and X_2 be two $I(0)$ series that share a $SCCF$, and write:

$$\begin{aligned} X_{1t} &= c_1^1 + \varphi_1^1 CC_{t-1}^1 + \epsilon_{1,t}^1, \\ X_{2t} &= c_2^1 + \varphi_2^1 CC_{t-1}^1 + \epsilon_{2,t}^1, \end{aligned} \tag{A.1}$$

where $CC_{t-1}^1 = \psi'[X'_{t-1}, ..., X'_{t-k+1}]'$; $[\varphi_1^1, \varphi_2^1]' = \delta_\perp$; $\epsilon_{i,t}^1$ is white noise; and $X_t = [X'_{1t}, X'_{2t}]'$.

If X_1 and X_3 also share a $SCCF$, then:

$$\begin{aligned} X_{1t} &= c_1^2 + \varphi_1^2 CC_{t-1}^2 + \epsilon_{1,t}^2, \\ X_{3t} &= c_2^2 + \varphi_2^2 CC_{t-1}^2 + \epsilon_{2,t}^2. \end{aligned} \tag{A.2}$$

Equalizing the first line of eq. (A.1) and eq. (A.2), solving for CC_{t-1}^2, and plugging the result in the second line of eq. (A.2) we get:

$$\begin{aligned} X_{3t} = c_2^2 + \frac{\varphi_2^2}{\varphi_1^2}[(c_1^1 - c_1^2) + \varphi_1^1 CC_{t-1}^1 + (\epsilon_{1,t}^1 - \epsilon_{1,t}^2)] + \epsilon_{2,t}^2 \;\Rightarrow \\ X_{3t} = c_3 + \varphi_3 CC_{t-1}^1 + \epsilon_{3,t}, \end{aligned} \tag{A.3}$$

where $c_3 = c_2^2 + \frac{\varphi_2^2}{\varphi_1^2}(c_1^1 - c_1^2)$; $\varphi_3 = \frac{\varphi_2^2 \varphi_1^1}{\varphi_1^2}$; and $\epsilon_{3t} = \frac{\varphi_2^2}{\varphi_1^2}(\epsilon_{1,t}^1 - \epsilon_{1,t}^2) + \epsilon_{2t}^2$. Since ϵ_{3t} is white noise, X_3 has the same $SCCF$ as X_1 and X_2.

Another way to see the transitivity of the $SCCF$ is to notice CC_{t-1}^2 can be written as a linear function of CC_{t-1}^1 plus a constant and a white noise.

Appendix B Proof of proposition 1

First, take logs in eq. (9):

$$log(\overline{E[\omega]}) = log(N - SC_i) + \gamma(\rho \hat{SC}_i - 1) log(p_{max}) + log(P(W_h)) - log(\hat{SC}_i)$$

Now let $SC_i = cN^{1/\lambda}$, with $\lambda > 1$ and $c > 0$. Using the results from §3.2.1 we can assume that $\hat{SC}_i$ grows at the same rate as SC_i. That is, we can assume that $\hat{SC}_i = \tilde{c}N^{1/\lambda}$, with $\tilde{c}$ not

necessarily equal to c. Then:

$$log(\overline{E[\omega]}) = log(N - cN^{1/\lambda}) + \gamma(\rho\tilde{c}N^{1/\lambda} - 1)log(p_{max}) + log(P(W_h)) - log(\tilde{c}N^{1/\lambda})$$

Let $log(\overline{E^*[\omega]}) = log(N) - c^*N^{1/\lambda}$, with:

$$c^* = -\tilde{c}\gamma\rho log(p_{max}).$$

Since p_{max} is fixed, for sufficiently large N, $log(\overline{E[\omega]}) \leq log(\overline{E^*[\omega]})$. Then, having a constant $\overline{E^*[\omega]}$ is a sufficient condition for keeping $\overline{E[\omega]}$ lower than a certain threshold.

To find the condition for constant $log(\overline{E^*[W]})$ write;

$$log(N) - c^*N^{1/\lambda} = C,$$

from where;

$$\lambda = \frac{log(N)}{log[log(N) - C] - log(c^*)} \longrightarrow \frac{log(N)}{log[log(N)]} > 1.$$

Noting that $N^{log(log(N))/log(N)} = log(N)$, completes the proof. ■