Discovering pervasive and non-pervasive common cycles with an application to the US CPI

Guillermo Carlomagno*.
Center of Economic Research (CINVE), University ORT Uruguay
and
Antoni Espasa
Department of Statistics, Instituto Flores de Lemus,
University Carlos III of Madrid

Abstract

The objective of this paper is to propose a strategy for exploiting short-run commonalities in the sectoral components of macroeconomic variables in order to obtain better models and more accurate forecasts of the components and, hopefully, of the aggregate. 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 methods 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 of 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. Furthermore, the procedure works both with fixed N and $T \to \infty$, and with $[T, N] \to \infty$. We perform an application to the US CPI and find good results of our procedure.

Keywords: Common features, 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. One reason for this is that very often decision makers need to analyse the disaggregates to get specific knowledge of them, to grasp a better understanding of the aggregate and eventually to make better decisions. However, the usual macroeconometric analyses that focus on modeling and forecasting economic aggregates (e.g. GDP, CPI, industrial production, employment, imports and exports, etc) do not make full use of the large amount of information contained in the disaggregates.

*Correspondence to Guillermo Carlomagno. E-mail: gcarlomagno@cinve.org.uy.

The objective of this paper is developing a strategy to discover and exploit short-run commonalities in the sectoral components of macroeconomic variables in order to model and forecast those components. Our main contribution concerns cases in which the number of components is large and traditional multivariate approaches are not feasible. As an important byproduct, this approach allows constructing an indirect forecast of the aggregate that hopefully could 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 (see, e.g., Lucas, 1977, Long Jr and Plosser, 1987, Engle and Issler, 1995, Engle and Kozicki, 1993, Engle and Kozicki, 1993, Vahid and Engle, 1993, Candelon et al., 2005, Hecq et al., 2006, Cubadda, 2007, Li et al., 2015). In this paper the study of common cyclical features in done on the components of macroeconomic series. This is relevant for understanding their interrelations, and consequently for constructing better empirical models and obtaining more accurate forecasts. As Espasa and Mayo-Burgos (2013) argue, one should work with the components at the maximum disaggregation level (basic components) because intermediate sub-aggregates may add up components that do not share common cycles, what could introduce invalid restrictions in a DGP refereed to the basic components.

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 improvements in forecasting accuracy over unrestricted models.

As Espasa and Mayo-Burgos (2013) show, an important characteristic of the sectoral components of macro variables is that they can be grouped into relatively small subsets in which all of the series show the same short-run dynamic 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. 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 the present paper we show that subsets of components that share single common cycles

[1]A revised version of this working paper was later published as Espasa and Mayo-Burgos (2013).

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, a generalization to make it useful when the size of the subsets may be large (what we latter call 'relaxation procedure'), a Monte Carlo study in which the finite samples behavior of the procedure are studied, and an application to the US CPI.

The pairwise strategy consists of testing for common cycles in all of the $N(N-1)/2$ pairs that exist among the N components of an aggregate (N is usually greater than a hundred), and then, constructing subsets such that in each subset all pairs of elements have 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 can be consistently estimated by OLS. The components not belonging to any subset can be modeled independently by univarariate models, or by restricting their forecasts to add up to the forecast of the intermediate sub-aggregate formed by them. This last possibility can be carried out using the *general combining rule* proposed by Guerrero and Peña (2003).

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 only to a reduced group of components), 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 non-pervasive factors have been proposed in the DFM literature, most of them assume 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 (2016) (BKP, hereafter), Bailey, Holly, and Pesaran (2016) (BHP, hereafter), and Ando and Bai (2016), work with unknown non-pervasive structures. In these cases 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 C in Bai, 2003 or Assumption A2 in Doz et al., 2012) are required. These assumptions do not fit our framework of interest.

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. Lastly, though in this paper we focus on $I(0)$ series, we will argue in §6 that a generalized version of our procedure is applicable both when the series are $I(0)$ and when they are $I(1)$. In the latter case we do not need to differentiate.

There is, however, one aspect in which our procedure is less general than the DFM alterna-

tives described above. While in those 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, which is our main objective, 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 the other 'sectorial', in §6 we describe this generalization.

The rest of this paper is organized as follows. In §2 we give the precise definitions of *common cycles* that are used throughout 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, and in §6 we discuss some possible extensions. In §7 we include an empirical application to the US CPI broken down in 159 components and §8 concludes the paper.

2 Concepts of common cycles

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

Engle and Kozicki (1993) define a *serial correlation common feature* (SCCF, hereafter) to be present 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 stationary VAR:

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

where X_t is an $N \times 1$ vector, ϵ_t an iid N-dimensional process and c an $N-$dimensional vector of constants. 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, which implies that $\delta' \Pi_i = 0$ 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

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

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 the 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 that the smallest s eigenvalues are zero, the statistic has a chi-squared distribution with $s^2 + sNk + sr - sN$ degrees of freedom.

As noted above, Engle and Kozicki (1993) develop their method for stationary variables, thus, in most cases, the analysis must be carried out for the differenced variables. 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.

A natural extension of the notion of $SCCF$ in cointegrated series is to allow the possibility that the SCCF vectors cancel the short-run dynamics, but are not related in any 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. As we briefly mention in §6, in that case, we can deal both 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 the definition of SCCF requires. To deal with this possibility, Cubadda and Hecq (2001) introduce the concept of *polynomial serial correlation common feature* (PSCCF). Although all the results of the present paper can be generalized to the case of PSCCF (see §6), in order to keep things simple, we leave that 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 common cycles being 'transitive'. That is, it must be the case that if series A_t and B_t share the cycle, and series A_t and C_t also share the cycle, one can conclude that B_t and C_t also have the same cycle. A simple argument, available upon request, shows that SCCF structures are in fact transitive.

3.1 Assumptions

Our general framework includes the following four assumptions:

Assumption A *The N components are generated by the VAR in eq. (1), which may be gen-*

eralized to include outliers and/or location shifts.

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

Assumption C *There is at least one subset containing S_1 components (with $2 \leq S_1 \leq N$) that share a single common cycle (we will use notation S_1 as the name of the subset and as its cardinality).*

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

Assumption E $\frac{N}{\sqrt{T}} \to \leq c$ *as* $[T, N] \to \infty$, *for some fixed constant* c.

Remark 1 *In principle assuming that the innovations are iid would be enough for the common cycles tests to be asymptotically valid, so that the normality imposed in assumption B would be unnecessary. However, as will become clear later, normality of ϵ_t is necessary to guarantee the validity of the asymptotic inference from the bi-variate models in which we test for common cycles.*

Remark 2 *Assumption B is related to the innovations, 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 the 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.*

Assumption C makes our objective of discovering single-cycle subsets to be relevant, and assumption D rules out the trivial common cycles that will appear if some components are white noise. As we argue below, assumption E is required to control false discoveries when we let N going to infinity.

Remark 3 *Instead of assumption D we could require that at least S components satisfy it. This flexibilization would require testing the significance of the coefficients of the estimated common cycles ($\delta_\perp$). This is how we proceed in the empirical application of § 7.*

3.2 Statistical properties

Define S_j to be a subset in which all of the series share a SCCF (recall that we are using S_j both as the name of the subset, and to indicate the number of series inside it). Abusing notation, we will denote the subset constructed by the pairwise procedure as $\hat{S}_j$.

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 $\hat{S}_j$. ii) *Gauge:* The proportion of wrong series that are included in $\hat{S}_j$. iii) *False discovery:* The discovery of nonexistent single-cycle subsets[2].

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

3.2.1 Potency

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

If we were testing a single hypothesis, the probability of not falsely 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, $m = S_j(S_j-1)/2$ may be quite large. Thus, if tests are independent, the probability of including all of the correct series in $\hat{S}_j$ will be close to zero. Simulation results (available upon request) show that, under some circumstances, common cycle tests between the series in S_j may be independent. This means that the probability of including all of the correct series in $\hat{S}_j$ may be a rapidly decreasing function of the number of series in the true subset. This would be an undesirable property for our procedure.

To mitigate this problem we exploit two facts. First, since the tests are transitive, and each series is included in several pairs, we could infer the correct result for one pair using the results of other ones. 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 $\hat{S}_j$ passes the pairwise test for a common cycle with *all* of the other series in the subset, letting a series entering into $\hat{S}_j$ when it passes the test with *almost all* of the other series in the subset. We call this strategy a *relaxation strategy* (step iii of the algorithm in §4).

Second, since the asymptotic power of common cycle 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. In what follows we develop an informal argument suggesting that the procedure is to be expected to have high potency asymptotically (large T).

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 S_j is $1-\varphi$. For any other pair this probability is zero. Hence, a natural way to see the problem of finding a single-cycle subset is in terms of the theory of random graphs.

To put it simple, a random graph can be seen as a square symmetric matrix of zeros and ones in which each entry has probability p of having a one and $1-p$ of having a zero, independently of the other entries. When the (i,j)th entry is a one, we say that there is an *edge* between the nodes 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 the (i,j)th entry corresponds to the pair formed by series i and j. Thus, ones would appear in those pairs of series for which a common

cycle was found. Although when testing for common cycles the edge probabilities might not be independent, the case of independence is the worst possible 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 S_j are zero, we can focus on the sub-graphs formed by the series in a particular S_j.

As we describe in 4, our procedure starts looking for the largest single-cycle subset. 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 *maximal clique* problem described in the random graph literature. The maximal clique is defined as the largest subgraph in which all nodes are pairwise connected (see, *inter alia*, 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 S_j will be almost surely connected with, at least, $k-1$ other elements. The authors denote this probability as *percolation probability* and it is given by

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

Using eq. (4), we can fix a proportion ρ and choose $k = \rho(S_j - 1) + 1$, such that eq. (4) will give the edge probability required to find a subset of size S_j in which, for each series, a common cycle will be found with at least $\rho(S_j - 1)$ of the remaining series.

Figure 1 shows this threshold probability for different alternatives of ρ and S_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 common cycles tests). For instance, for $S_j = 40$ and $\rho = 0.8$, we would need $(1-\varphi) = 0.8$, meaning that with a φ smaller than or equal to 0.2 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 of the correct series in the estimated S_j would be $(1-\varphi)^{\frac{S_j(S_j-1)}{2}}$, which is virtually zero even for small values of φ and moderate S_j. This implies that relaxing the requirement from full connection, to almost full connection may lead to a an important increment in the probability of including all of the correct series in the estimated S_j.

In deriving eq. (4) Derényi et al. (2005) assume $S_j \to \infty$. Thus, eq. (4) could be a rough measure of the percolation probability when S_j is fixed.

To better understand the properties of the relaxation procedure when S_j is fixed, we perform a small simulation study that replicates the case of $T \to \infty$ and S_j fixed. Simulations for finite T are delayed to §5.

In each of the 5000 experiments we simulate a random graph of size S_j with independent

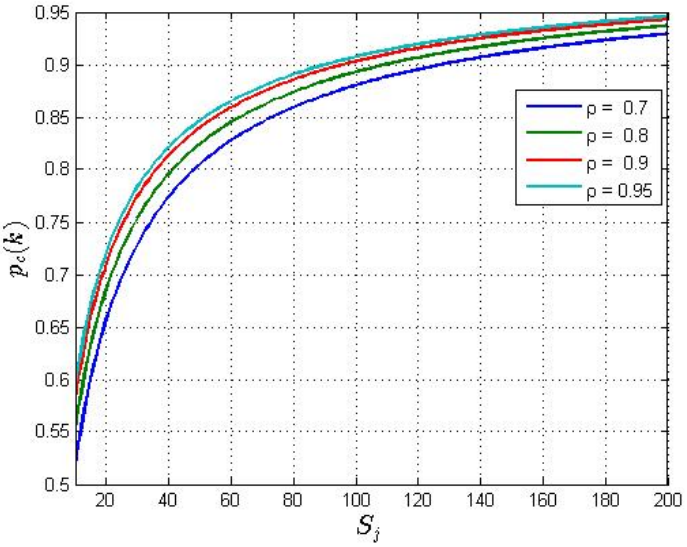

Figure 1: Percolation probabilities $p_c(k)$, for $k = \rho(S_j - 1) + 1$

edge probability p. This is done by generating a square symmetric matrix with zeros and ones in which each entry has probability p of having a one and $1-p$ of having a zero, independently of the other entries.

The asymptotic probability of finding a common cycle between two series that truly have it is $1-\varphi$. Hence, by fixing p equal to $1-\varphi$ this simulation replicates the random graph that would be obtained when testing for common cycles between all of the pairs among S_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 S_j (mean($\hat{S}_j/S_j$)) and the probability of including *all* of the elements in $\hat{S}_j$ (mean($I_{\hat{S}_j=S_j}$)), both when applying the strict full connection criteria and when relaxing it. For relaxing the strict full connection criteria, 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 into $\hat{S}_j$).

The results are included in Table 1. As a general conclusion, the relaxation procedure allows reaching 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 of the size of the estimated subset to the true size of the subset is 0.97.

In summary, when T goes to infinity, our procedure is expected to have high potency, regardless of the size of S_j; when $S_j \to \infty$ eq. (4) applies, when S_j is fixed results in table 1 suggest that potency will also be high.

3.2.2 Gauge

To include wrong series in some $\hat{S}_j$ common cycle tests should lead to concluding $s = 1$ when the true hypothesis is $s = 0$, i.e., not to reject the false null hypothesis of $s > 0$.

Let $\mathrm{X}_{\mathrm{out}}$ be a series outside S_j and let W_i be the event of wrongly not rejecting $s > 0$ with the i^{th} series in the estimated S_j ($\hat{S}_j$). As for wrongly including $\mathrm{X}_{\mathrm{out}}$ in $\hat{S}_j$ we need to find

Table 1: Fully connected *vs* almost fully connected subsets

		S=5	S=10	S=25	S=40	S=50
				$\boldsymbol{p = 0.95}$		
Full contection	mean($\hat{S}/S$)	0.91	0.83	0.69	0.61	0.55
	mean($I_{\hat{S}=S}$)	0.59	0.10	0.00	0.00	0.00
Almost Full conection	mean($\hat{S}/S$)	0.99	0.95	0.98	0.97	0.97
	mean($I_{\hat{S}=S}$)	0.94	0.58	0.56	0.27	0.25
				$\boldsymbol{p = 0.99}$		
Full contection	mean($\hat{S}/S$)	0.98	0.96	0.90	0.85	0.83
	mean($I_{\hat{S}=S}$)	0.90	0.64	0.05	0.00	0.00
Almost Full conection	mean($\hat{S}/S$)	1.00	1.00	1.00	1.00	1.00
	mean($I_{\hat{S}=S}$)	1.00	0.97	1.00	1.00	1.00

- The relaxation parameter (λ) is $max(1, 0.2S)$.
- 'mean' denotes the mean across experiments.
- The number of experiments is 5000.
- $\hat{S}$ is the number of series included in the largest (almost) fully connected subset.
- $I_{\hat{S}=S}$ denotes the indicator function that takes the value 1 if $\hat{S} = S$ and 0 otherwise.

a common cycle with $\rho\hat{S}_j$ other series, we need to wrongly not reject $\rho\hat{S}_j$ hypotheses. The probability of this event is $P(W_1 \cap ... \cap W_{\rho\hat{S}_j})$, and can be factorized as

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

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

Note that when $s = 0$ the eigenvalue λ_1 of eq. (3) is be strictly larger than zero and the test statistic for the hypothesis $s > 0$ vs $s = 0$ diverges at the rate of T. Thus, $P(W_i)$ is $Op(T^{-1})$.

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

$$\sum_{h=1}^{N-S_j} Z_h, \tag{6}$$

and the expected number of wrong inclusions is

$$E[\textstyle\sum_{h=1}^{N-S_j} Z_h] = \sum_{h=1}^{N-S_j} E[Z_h] = (N - S_j)E[Z_h]. \tag{7}$$

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

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

Under the extreme assumption that all conditional probabilities in eq. (5) are equal to one, $E[Z_h] = P(W_h)$. Since $P(W_h)$ is $Op(T^{-1})$, using assumption D, $lim_{T\to\infty} E[Z_h] = 0$. When N is fixed and T goes to infinity, eq. (8) is $Op(T^{-1})$, meaning that the gauge tends to zero as $T \to \infty$. When both N and T go to infinity, in order to avoid having eq. (8) growing without limit, we need N/S_j to be, at most, $Op(T)$. Clearly, this condition includes cases in which the common cycles are not pervasive, i.e., $N/S_j \to \infty$, so that we will have good gauge properties even in that case.

Note that, by imposing N to be at most $Op(T^{1/2})$, assumption E ensures N/S_j being at most $Op(T)$, but in a very restrictive way. As will became clear below, that assumption is necessary for controlling false discoveries.

Previous argument can also be made without relying on the asymptotic power of the common cycle tests. A proof of this statement is available upon request.

3.2.3 False discovery

Define $M = N - \sum_j S_j$ 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 the pair m. Using the same reasoning as that used for expressions 7 and 8, the expected number of false discoveries (the number of 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 for having a fixed expected number of false discoveries is $M^* \times p \to 0$, which requires $N - \sum_j S_j$ to grow a rate less than or equal to that of $p^{-1/2}$. This implies that $N - \sum_j S_j$ can grow, at most, at the same rate as $T^{1/2}$ (see assumption E).

3.3 Partial models

The pairwise strategy described in previous sections requires partial systems' estimation in the sense that we assume the existence of a full VAR model for all of the components but estimate several partial bi-variate systems in which we test common cycles.

Linear transformations of processes that follow VAR models have an infinite VAR representation with exponentially decreasing coefficients' matrices (see Johansen and Juselius, 2014), so that the bi-variate models in which we test common cycles will be, in general, approximations to the true data generating process.

If the innovations of the infinite VAR representation are iid, the lag length is, at most, $Op(T^{1/3})$, and finite order VAR models are fitted with a lag length that increases with T, usual asymptotic inference is still valid (see, Saikkonen, 1992, Saikkonen and Lütkepohl, 1996 and Johansen and Juselius, 2014). To the best of our knowledge, this result has not been proven for non-iid innovations *and* non-Gaussian innovations. As shown by Johansen and Juselius (2014) the innovations of the linear transformations will be white noise but not necessarily iid, except if the original innovations are Gaussian iid. In the latter case the innovations of

the linear transformations will also be Gaussian iid.

Thus, in order to guarantee that our bi-variate models are valid for inference, we need Gaussian iid innovations in the model of eq. (1). This explains assumption B.

Apart from previous discussion, the strategy of testing for common cycles in partial models could be thought to imply a loss of power compared to a 'complete' model approach (when feasible). To analyze this issue, we performed a small simulation study to compare the ability of the pairwise approach with that of the full model approach, when the latter is feasible (small N). Since our focus in this paper is on relatively large N, we do not report the results (available upon request). The main conclusion of those 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, compared to the full model approach.

4 The algorithm

In order to discover subsets S_j — within a set of N series — in which all of the series share a single cycle we apply a six-step algorithm:

i. Perform common cycle tests between all possible pairs of series, store the resulting p-values, and construct an $N \times N$ Boolean adjacency matrix, A, that contains a 1 in the (i, j)th entry 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 in A, for example, using the Bron-Kerbosch algorithm (see Bron and Kerbosch, 1973). Recall that the maximal clique is defined as the largest subgraph in which all nodes are pairwise connected (see also Bollobás and Erdös, 1976). We will refer to the maximal clique as the largest *single-cycle subset*, $\hat{S}_1$.

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

   (a) A single cycle — at the original nominal size, φ — was found in step i with at least $\hat{S}_1 - \lambda$ of the series already in the subset $\hat{S}_1$.

   (b) When the nominal size of the test is relaxed to φ^*, the candidate is found to have a common cycle with all the series already in the subset $\hat{S}_1$.

iv. Using previous results, construct the set of candidates C_0.

   - If all the candidates have a common cycle between each other (with the original nominal size, φ), let all of them enter into $\hat{S}_1$ and go to Step v (because there are no more potential candidates).

- If not, find the maximal clique (see Step ii) inside C_0 and enlarge $\hat{S}_1$ with all the series in this maximal clique. Note that after including these series, there could still remain some potential candidates. To check for this possibility: construct a new set of candidates C_1 (these candidates must satisfy conditions iiia and iiib with the enlarged $\hat{S}_1$), and repeat the procedure in the present Step.
- If, according to conditions iiia and iiib, there are no candidates that share the cycle, try to include them sequentially, starting with the one which has a common cycle with more of the series already in the subset. In case of a conflict (i.e., 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 ad hoc criterion could be, for example, to include the series whose sum of p-values for the null $s > 0$ is the minimum.

v. In order to find other single-cycle subsets repeat steps ii to iv but excluding the series already included in some almost fully-connected subset. This will lead to discover other almost fully-connected subsets with descending sizes.

5 Simulations

In this section we perform some Monte Carlo experiments to analyze the finite sample properties of the procedure.

5.1 Design of the experiments

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

$$X_t = c + \Pi X_{t-1} + \epsilon_t, \tag{9}$$

where $\epsilon_t \sim N(0, \Sigma)$, and the roots of $det(I - \Pi L)$ are all outside the unit circle. We simulate situations in which only a subset of S_1 series share a single cycle and there are no more common cycle restrictions in the system. In order to simplify the dynamics of the systems, Π has the following structure:

$$\Pi = \begin{bmatrix} A_{S_1 \times S_1} & 0 \\ 0 & D_{N-S_1 \times N-S_1} \end{bmatrix}, \tag{10}$$

where $A = \delta_\perp^* \Psi^{*'}$, with $\delta_\perp^*$ and Ψ^* being $S_1 \times 1$ vectors (see §2 for a justification of this structure), and D a diagonal matrix. This does not imply that series outside S_1 are independent of each other, or with respect to series inside S_1, as Σ is not necessarily diagonal (the correlation structure of the innovations is detailed below). Partition the vector X_t into its first S_1 elements and the remaining $N - S_1$, and denote the first sub-vector as $X_t^{S_1}$. Then, the common cycle is $\Psi^{*'} X_{t-1}^{S_1}$, and $\delta_\perp^*$ contains the coefficients of the common cycle in each of the first S_1 series.

Calling π_{ij} the elements of Π, it can be shown that, after imposing the condition that

$|\pi_{ii}| < 1$, $\forall\, i \geq S_1 + 1$, the stationarity condition of eq. (9) is:

$$|\sum_{i=1}^{S_1} d_i \psi_i| < 1.$$

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

There are infinitely many different possibilities for $\delta_\perp^*$ 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 $\frac{1}{1.2S_1}$ and $\frac{1}{1.1S_1}$.

DGP 1 : The same as DGP 0 but imposing z zeroes in Ψ^* so that we can change the S_1 in the denominator to $S_1 - z$ and the non-zero entries will be larger. We set $S_1 - z = 2$, so that the common cycles will be generated by two of the series in S_1 (those whose coefficients in $\Psi^{*'}$ are different from zero).

DGP 2 : The same as DGP 1 but allowing some negative values in $\delta_\perp^*$ (there will be some counter-cyclical variables). This allows increasing the non-zero values in Ψ^* with respect to option DGP 1. We limit the number of negative coefficients to 20% of the variables in $\delta_\perp^*$. Hence, the number of non-zero coefficients in Ψ^* is $(2 + 0.2 \times S_1)$.

Note that in DGP 0 all the entries of ψ will be rather small, even for relatively small S_1. 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 innovations ϵ_t are generated by

$$\epsilon_{it} = \eta_{i,t} + \sum_{j\neq 0,\, j=-Q}^{min(Q,i-1)} \beta \eta_{i-j,t}, \tag{11}$$

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 another $2Q$ residuals. To avoid having higher cross-correlation inside S_1 than outside it, the rows of matrix Π in eq. (10) are randomly disordered so that series inside the subset S_1 are not in the first S_1 positions of vector X_t. We consider four alternative combinations for β and Q: $[\beta = 0, Q = 0]$, $[\beta = -0.3, Q = 10]$, $[\beta = -0.3, Q = 20]$, and $[\beta = -0.3, Q = 30]$.

For the two DGPs we consider three scenarios and three sample sizes. In all cases $N = 100$. In scenario 1 we set $S_1 = 10$, in scenario 2, $S_1 = 25$; and in scenario 3, $S_1 = 40$. The sample sizes are $T = 100$, $T = 200$, and $T = 400$.

For each DGP, scenario, and sample size, we perform 500 Monte Carlo replications. Our objective is to discover the series that are in S_1. To do that, we carry out 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 5. Furthermore, as we consider four alternative combinations for β and Q of eq. (11), this further multiplies the number of models by 4.

Finally, a relaxation parameter λ has to be defined, i.e., we need to define the maximum number of tests that a series can fail and still enter in $\hat{S}_1$ (see step iii of the algorithm). Our baseline choice is to set $\lambda = 0.4 \times \hat{S}_1^{\lambda=0}$, where $\hat{S}_1^{\lambda=0}$ is the number of series in the subset obtained using $\lambda = 0$.

A word about the computing time is now in order. In a personal computer equipped with Windows 10 64 bits, with a processor i7-6700HQ of 2.60Hz, and 16GB of ram it takes 7.8 seconds to perform the 4950 pairwise tests selecting models between 1 and 5 lags.

5.2 Monte Carlo results

Table 2 includes the gauge and potency of the pairwise strategy for DGP 1 and $\lambda = 0.4 \times \hat{S}_1^{\lambda=0}$, results for different choices of λ are included in appendix A. As analyzed above, false discovery is not an issue asymptotically and it turns out not be relevant in finite samples either. Thus, for clarity, we omit the results about false discovery in this part of the paper and include them in appendix B. As the conclusions for DGP 2 are the same, the details are omitted, but are available upon request.

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. The results in terms of potency are also very good, as we get values above 0.9 in all cases (except for the case of $T = 100$ with independent innovations).

Table A.1 in appendix A replicates the results for values of the relaxation parameter, λ that are smaller than or equal to the one used in Table 2. The choices are $\lambda = 0$ (no relaxation), $\lambda = min[2, 0.4 \times \hat{S}_1^{\lambda=0}]$, $\lambda = min[5, 0.4 \times \hat{S}_1^{\lambda=0}]$ and $\lambda = min[10, 0.4 \times \hat{S}_1^{\lambda=0}]$.

Two main conclusions emerge from table A.1. First, the relaxation leads to great improvements in potency. For example, in scenario 3 ($S_1 = 40$) with independent innovations and no relaxation we get potencies of 70.2, 65.8 and 52.0 for $T = 400$, $T = 200$, and $T = 100$, respectively. With $\lambda = min[2, 0.4 \times \hat{S}_1^{\lambda=0}]$ we already get a great improvement — of, approximately, 15 percentage points — for all sample sizes. This improvement in potencies continues up to the figures in table 2 which are around 25 percentage points higher than those with $\lambda = 0$. These observations are also valid for the other scenarios and correlation of innovations.

The second conclusion from table A.1 is that the improvements in potency derived from the relaxation procedure are almost costless in terms of gauge. Gauges in table 2 are almost the same as those in the first block of table A.1 ($\lambda = 0$), and very close to zero as well.

These two conclusions were expected from the analysis in §3.2.

Table 2: Gauge and potency of the pairwise procedure. DGP 1 ($\lambda = 0.4 \times S^{\lambda=0}$, $\varphi = 5\%$, $\varphi^* = 0.5\%$)

	$S_1 = 10$		$S_1 = 25$		$S_1 = 40$	
	Gauge	***Potency***	***Gauge***	***Potency***	***Gauge***	***Potency***
	$\beta = 0$, $Q = 0$ *(independent innovations)*					
$T = 400$	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 innovations)*					
$T = 400$	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 innovations)*					
$T = 400$	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 innovations)*					
$T = 400$	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: 500.
- $Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$
- $Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$
- Z_2 = number of wrong series included in $\hat{S}$
- Z_1 = number of correct series included in $\hat{S}$
- $Nexp$ = number of experiments

As discussed in the Introduction, an alternative strategy to ours could be the estimation of Dynamic Factor Models. As we argued above, 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 the 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 $S_1 \geq 25$, $T \geq 200$, and $Q = \beta = 0$. This was to be expected as those cases might 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 that are not implemented in this paper.

As mentioned in §2 the pairwise approach can be generalized 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. Although when considered in a pairwise fashion these structures are not transitive, simple modifications of the testing strategy make them so, and the extensions can be implemented.

A third extension of interest is allowing subsets with more than one common cycle. In this paper we focused on the case that the data can be grouped into subsets in which the series inside them share just one common cycle. 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.

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. However, assuming that in the set of N series there is a subset of series that just have the *general* cycle, our approach could be adapted to this situation. Under this assumption we could proceed with the following algorithm:

(i) Apply the pairwise procedure proposed in this paper. This will lead us to discover the subset of series that only have the general cycle — call it S_0. (ii) Test for a common cycle in all of the triplets formed by each of the series inside $\hat{S}_0$ and every pair of outsiders. Since there are $\frac{(N-\hat{S}_0)(N-\hat{S}_0-1)}{2}$ pairs, we have $\hat{S}_0 \times \frac{(N-\hat{S}_0)(N-\hat{S}_0-1)}{2}$ triplets. For the triplets in which the outsiders have the same sectorial cycle, we will find two common cycles ($s = 1$) because all the series share the same general cycle. (iii) Construct an $(N - \hat{S}_0) \times (N - \hat{S}_0)$ symmetric *adjacency matrix* for the series outside $\hat{S}_0$ such that each entry of this matrix represents a pair of the components outside $\hat{S}_0$. Each of those pairs belongs to $\hat{S}_0$ different triplets: one for each element of $\hat{S}_0$. Then, in each entry of the adjacency matrix, put a 1 if almost all of the corresponding $\hat{S}_0$ 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 Empirical application to the US CPI

In this section we apply the pairwise procedure analyzed in previous sections to the US CPI. For space reasons we focus just on the forecasting exercise, detailed results about the

common cycles tests and the resulting single-cycle subsets are available upon request.

The main aim of the disaggregated analysis proposed in this paper is to arrive to a better understanding, modeling and forecasting of the components of a macro variable, which at the same time would lead to a better understanding of the aggregate. As a by-product of the disaggregated analysis we obtain an indirect forecast of the aggregate that consists of aggregating the forecasts of the components.

In arriving to the specification of the single-cycle subsets and to the forecasting models of the components we apply a battery of tests, which give guaranties to the validity of the results. But if the single-cycle subsets are going to be really relevant we expect that the mentioned indirect forecast for the aggregate is not significantly worse — hopefully significantly better — than the direct forecast. Therefore, comparing a direct forecast coming from a scalar model for the aggregate with this indirect one is a good way to evaluate the overall forecasting performance of our procedure.

In applying our procedure to a macro variable we could find that at its maximum level of disaggregation there are components which move mainly by steps. In a CPI index this happens because those components incorporate administrative prices or prices which usually are only revised once a year. In developed economies the total weight of those components in a macro variable is, in general, small. In order to identify single-cycle subsets of an aggregate, A, the moving-by-steps components should be excluded. We denote as A^* the aggregate that exclude the mentioned components. Then, the comparison between the indirect forecasts with direct ones should be referred to A^*.

In the presence of moving-by-steps components, an indirect forecast of the overall aggregate, A, could be obtained from a regression of A on its corresponding A^*. A more accurate procedure could be to forecast the moving-by-steps components individually and aggregate those forecasts to the forecast of A^*.

7.1 Data

The CPI breakdown used in this analysis corresponds to the maximum disaggregation level available to the public in the *Bureau of Labor Statistics* (seasonally un-adjusted CPI-U for all urban consumers) for the period $1999.1 - 2016.12$ (216 observations)[3]. The total number of components is 174. Not all the series have data for the whole sample. After dropping those with less than 150 observations we keep 169 components. From these series we exclude nine that evolve by steps so that we end up with 160 series that, considering the 2016 weights, represent 92% of the CPI[4]. Among the remaining series, the item *Owners' equivalent rent of primary residence* weights approximately 24% of the CPI (a single component with such a large weight is a rare case).

[3]Series are available at https://download.bls.gov/pub/time.series/cu/ or upon request.

[4]The nine excluded series are: Tuition other school fees and childcare, College tuition and fees, Elementary and high school tuition and fees, Child care and nursery school, Technical and business school tuition and fees, Postage, Delivery services, Limited service meals and snacks, Other lodging away from home including hotels and motels

In order to avoid that the global results could be driven by the forecast accuracy of a single series, we also exclude Owners' equivalent rent of primary residence from the analysis. Thus, all in all, we will work with 159 series, the remaining ones are neither considered for the construction of the single-cycle subsets, nor for the forecasting exercises. We denote the aggregate corresponding to the 159 series as CPI^*.

7.2 Design of the forecasting exercises

In order to obtain economically and statistically sensible common cycles restrictions between the components of the CPI^* we consider only those in which the coefficients $\delta_{\perp,i}$ of both prices are statistically significant, and the ψ_i are stable over time.

Common cycles tests are performed at the 5% of significance in a VAR in differences and the number of lags for each pair is determined with the AIC. Centered seasonal dummies and outliers' indicators detected with Impulse Indicator Saturation (IIS) are included in all models[5]. IIS is a procedure that saturates the model with one indicator variable for each observation and selects the relevant ones (see Santos et al. (2008), Johansen and Nielsen (2009), and Castle et al. (2012)).

Using a relaxation parameter $\lambda = min[10, 0.2 \times \hat{S}_j^{\lambda=0}]$, we found 8 subsets that jointly include 46 series, which represent 28.9% of the components and 26.1% of the total weight of the CPI^*. Using $\lambda = min[10, 0.4 \times \hat{S}_j^{\lambda=0}]$ produce similar results, both in terms of the specific subsets, and the total series and weights.

For building the single-equation models that will be employed to forecast the 159 components, we use the automatic model selection algorithm *Autometrics*. Starting from a General Unrestricted Model (GUM) that includes all potentially relevant regressors, and using a multiple path search, Autometrics reduces the GUM to a simpler model that encompasses it and passes a battery of diagnostic tests (see Doornik (2009) and Castle et al. (2011)). The specific GUMs considered in this application are detailed latter on table 3. Additionally, we control for data irregularities in the models by including impulse indicators in periods with large residuals[6].

To assess the forecasting accuracy of our procedure we compare the forecast of the aggregate obtained indirectly by aggregating the forecasts of the components with a direct forecast coming from a scalar model for the aggregate, and with an indirect procedure in which all components are forecast with single-equation models that do not incorporate common cycle restrictions. We denote our indirect approach by I-PW (the 'I' stands for indirect and 'PW' for pairwise), the direct one by D, and the unrestricted indirect by 'I-B' (indirect basic).

For these approaches (D, I-PW and I-B) we consider three broad possibilities depending on

[5]Previously, we performed seasonal unit root tests as proposed by Osborn et al. (1988) to all of the components. The results indicate that they do not show seasonal unit roots in general and that the assumption of only one unit root, linear growth and deterministic seasonality seems reasonable (details are available upon request).

[6]In a real application correcting outliers with IIS would be a better option than jut correcting large outliers. However, as our objective in this application is comparing methods we opt for correcting large residuals, in all the forecasting strategies, because it is a simpler and less computationally costly alternative.

the regressors to be included in the General Unrestricted Model (GUM). For the models of the components that belong to some single-cycle subset we only include the corresponding cycle, seasonal dummies and outlier indicators. For the other models, apart from own lags, seasonal dummies and outlier indicators, we consider three possibilities: a) No other regressor, b) Lags of the aggregated CPI^* (only for the indirect procedures), c) Lags of eight broad categories which add up to the CPI, we denote this option as *Dissaggregated information* (DI).

For the I-PW procedure, the components which do not belong to any single-cycle subset can be modeled individually or all together with a scalar model for the sub-aggregate *Rest* that adds up all those series. This last possibility can be considered to be a restricted version of I-PW in which the forecast of the components in *Rest* is restricted to add up to the forecast of its sub-aggregate. We denote this approach as I-PW-Rest.

If when using the approach I-PW-Rest forecasting the components in *Rest* is of interest, one could proceed as in Guerrero and Peña (2003), whose *general combining rule* allows to produce individual forecasts restricted to add up to the forecast of the aggregate[7]

Thus, we have six different I-PW possibilities, three I-B, and two direct (D). For the D and I-B alternatives, we add an additional possibility consisting of including dynamic factors estimated from all the disaggregates (D-DFM and I-B-DFM). Therefore, we end up with 13 alternatives.

Table 3 includes a summary of the equations for the different forecasting procedures. From options a to c above, only option a is included in the table, the other two options are quite clear extensions.

Equations in table 3 represent the initial GUMs from where final forecasting models are selected using *Autometrics* with correction for large residuals. The selection of the models is carried out in two steps. First we use a nominal size of 0.25% to select variables, lags and impulse indicators. Retained indicators are stored. In a second step we consider the same GUM augmented with the retained impulse indicators and a target size of 5% with no outlier correction.

7.3 Forecasting comparison

Table 4 includes the results of an out of sample forecasting exercise for the evaluation period $2011.1 - 2016.12$. At each month of this period the 13 forecasting models described above are estimated using information up to the previous month, and multi-step ahead forecasts are produced for horizons $H = 1$ to $H = 12$. The search of the single-cycle subsets, and the corresponding estimation of the common cycles, is carried out only each December. Hence, for the PW approaches we are using less information than the available in real-time forecasting, except for January.

First row of table 4 includes the root mean squared forecast error (RMSFE) of $\Delta_{12} log(CPI^*)$

[7]Note that given the potentially large number of series in *Rest* a VARMA model for those components (as proposed by Guerrero and Peña (2003)) would be, in general, unfeasible, so it would have to be substituted by ARIMA models for the individual components. Thus, it would be necessary to redefine the corresponding variance-covariance matrices of the procedure.

Table 3: Summary of the forecasting exercises

Model	Description
D	$\Delta cpi_t = c + \sum_{k=1}^{5} \phi_k \Delta cpi_{t-k} + \phi_{12}\Delta cpi_{t-12} + \phi_{24}\Delta cpi_{t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$
D-DFM	$\Delta cpi_t = c + \sum_{k=1}^{5} \phi_k \Delta cpi_{t-k} + \phi_{12}\Delta cpi_{t-12} + \phi_{24}\Delta cpi_{t-24} + \sum_{k=1}^{5} \delta_k F_{t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$
I-PW	*case i)* Series inside some subset $\Delta x_{i,t} = c + \delta \hat{CC}_{it} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$ *case ii)* Series not in any subset $\Delta x_{i,t} = c + \sum_{k=1}^{5} \phi_k \Delta x_{i,t-k} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$
I-PW-Rest	For series inside some subset, same as I-PW case i. For the others, only its corresponding sub-aggregate is forecast in a model with the same structure as I-PW case ii
I-B	$\Delta x_{i,t} = c + \sum_{k=1}^{5} \phi_k \Delta x_{i,t-k} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$
I -B-DFM	$\Delta x_{i,t} = c + \sum_{k=1}^{5} \phi_k \Delta x_{i,t-k_1} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} + \sum_{k=1}^{5} \delta_k F_{t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$

- Lower case letters denote logarithms, and *cpi* denotes $log(CPI^*)$.
- $\hat{CC}_{it}$ stands for the estimated common cycle of the subset to which series i belongs.
- All the equations represent the initial GUMs form where models are selected using *Autometrics* with control for large residuals. The selection is carried out in two steps. First we use a target size of 0.25% to select variables, lags and impulses. Retained impulses are stored. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% control for outliers. For the I-PW approach selection is not performed over $\hat{CC}_{it}$.
- $S_{i,t}$ are centered seasonal dummies.
- In models D-DFM and I-B-DFM the q-dimensional factors (F) are computed from the difference of all the components of the CPI^*. The optimal number of factors is chosen with the information criteria of Bai and Ng (2002). The factors are forecast in a VAR model, where lags are selected with *Autometrics* with control for large residuals. The same two step procedure for selecting impulses and regressors explained above applies in this case.

for horizons $H = 1$ to $H = 12$ of the direct procedure. All the other values in the table are ratios with respect to the first row. Table 5 includes pvalues of the Diebold-Mariano tests for comparing forecasting accuracy of selected methods.

In what follows we list the main conclusions from the forecasting comparison:

i. Table 4 shows that the use of disaggregated information in a scalar model for the aggregate (approach I-DI), as proposed by Hendry and Hubrich (2011), improve the RMSFE between 2 and 8 percentage points, and these improvements are statistically significant for horizons 7 to 12. The inclusion of dynamic factors extracted form the disaggregates improves

the forecasting accuracy only in short horizons, but the improvements are statistically significant only for one step ahead forecasts.

ii. When modeling the disaggregates without including additional information beyond the own lags (I-B approach in row 4 of table 4) the forecasting accuracy for short and medium horizons (1 to 6) is similar (not statistically distinguishable) to that of the baseline. For long horizons improvements in RMSFE range from 5 to 12 percentage points and are statistically significant.

iii. These improvements are even more important when including disaggregated information in the individual models of the components (approach I-B-DI in row 6) as they range from 8 to 17 percentage points. Again, the disaggregated information in terms of lags of official sub-aggregates of the CPI is much more relevant that the information from dynamic factors.

iv. The inclusion of the lagged CPI^* (row 5) or dynamic factors (row 7) in the models for the components (approaches I-B-CPI and I-DFM) deliver similar results as those of the baseline, and are dominated by simple disaggregated approach I-B in long horizons (row 4).

v. The I-PW approach (row 8) delivers somewhat better results than the baseline in short horizons (1 to 6), but differences are not statistically significant. In horizons 7 to 12 reductions in the RMSFE range from 9 to 18 percentage points and are statistically significant.

vi. The comparison between I-PW and I-B indicates that the former beats the latter in all horizons with improvements that range between 1 and 7 percentage points. As table 5 shows, these differences are statistically significant at the 10% for horizons 9 to 12.

vii. When adding lags of the broad categories of CPI in the pairwise approach I-PW-DI (row 10) and in the simple disaggregated approach I-B-DI(row 6), the former dominates in horizons 1, 3 and 9 to 12, in the other horizons RMSFE are the same. For the horizons that I-PW-DI dominates differences are statistically significant at the 10% only for horizons 11 and 12, as table 5 shows.

viii. The I-PW-Rest approach improves the RMSFE with respect to I-PW in all horizons. These differences are not statistically significant for horizons 1 to 4 (see table 5). For horizons 5 to 12 the improvements are remarkable. Starting with a RMSFE gain of 13 percentage points, the differences in favor of I-PW-Rest increase with the forecasting horizon and reach 23 percentage points in horizon 12. As table 5 shows the differences are statistically significant for those horizons.

ix. The forecasting gains of I-PW-Rest with respect to the baseline are even more remarkable. Staring from 6 percentage points in horizon 1, improvements in RMSFE systematically increase with the horizon to reach 42 points in H=12.

In summary, the pairwise approach performs better than all other alternatives considered in this paper. It gets bigger improvements in forecasting accuracy when using the restricted alternative (I-PW-Rest). This implies forecasting the series outside the single-cycle subsets with the restriction that their forecasts aggregate to the direct forecast of their corresponding sub-aggregate. It must be noted that this strategy can only be implemented in the pairwise approach that singles out the series which can be forecast with such restriction. Clearly, the option of using the aggregate of all series, baseline model, is not good at all.

It is relevant to note that the relative forecasting performance of I-PW-Rest with respect the baseline model is exclusively due to the specification and individual treatment of the components with common cycles restrictions. This points out that extending the pairwise approach by including not only common cycles, but common trends, and eventually common breaks and common seasonality which could affect the other components, seems a very promising approach.

Table 4: Relative RMSFE $\Delta_{12} log(CPI^*)$. (First row: RMSFE for the baseline. All the others are ratios with respect to the first)

		H=1	**H=2**	**H=3**	**H=4**	**H=5**	**H=6**	**H=7**	**H=8**	**H=9**	**H=10**	**H=11**	**H=12**
(1)	**D (baseline)**	0.26	0.46	0.61	0.71	0.80	0.90	1.02	1.14	1.25	1.36	1.46	1.57
(2)	**D-DI**	0.98	0.95	0.93	0.95	0.95	0.93	0.92*	0.92**	0.92**	0.93**	0.94**	0.95**
(3)	**D-DFM**	0.93**	0.93	0.96	0.97	0.99	1.01	1.02	1.03	1.02	1.02	1.02	1.02
(4)	**I-B**	0.97	1.01	1.01	0.99	0.98	0.97	0.95*	0.92**	0.90**	0.89**	0.89**	0.88**
(5)	**I-B-CPI**	0.97	0.99	1.01	1.03	1.05	1.07*	1.06*	1.04	1.03	1.04	1.04	1.04
(6)	**I-B-DI**	0.98	0.99	0.97	0.95	0.94	0.92*	0.89**	0.87**	0.85**	0.85**	0.84**	0.83**
(7)	**I-DFM**	0.94	0.99	1.04	1.07	1.09*	1.08*	1.05	1.02	1.00	0.98	0.97	0.96
(8)	**I-PW**	0.95	0.99	0.99	0.99	0.98	0.95	0.91*	0.88**	0.85**	0.84**	0.82**	0.82**
(9)	**I-PW-CPI**	0.93	0.95	0.95	0.96	0.98	0.98	0.97	0.95	0.93*	0.92*	0.92*	0.92*
(10)	**I-PW-DI**	0.96	0.99	0.96	0.95	0.95	0.93	0.89*	0.87**	0.84**	0.83**	0.82**	0.81**
(11)	**I-PW-Rest**	0.94	0.95	0.92	0.90	0.85*	0.79**	0.75**	0.71**	0.67**	0.63**	0.60**	0.58**
(12)	**I-PW-Rest-CPI**	0.96	0.98	0.95	0.92	0.87*	0.82**	0.79**	0.76**	0.73**	0.68**	0.64**	0.63**
(13)	**I-PW-Rest-DI**	0.95	0.97	0.97	0.96	0.94	0.88	0.82**	0.77**	0.71**	0.65**	0.61**	0.59**

* Significantly different from the baseline at the 10% level using the Diebold-Mariano test.
** Significantly different from the baseline at the 5% level using the Diebold-Mariano test.

Table 5: P-values of Deibold-Mariano tests for selected comparisons

	H=1	H=2	H=3	H=4	H=5	H=6	H=7	H=8	H=9	H=10	H=11	H=12
4 vs 8	0.16	0.22	0.21	0.43	0.43	0.30	0.21	0.15	0.09	0.05	0.04	0.04
2 vs 8	0.28	0.29	0.25	0.33	0.40	0.37	0.44	0.09	0.00	0.00	0.00	0.00
6 vs 8	0.27	0.48	0.34	0.28	0.29	0.28	0.31	0.39	0.49	0.35	0.32	0.39
2 vs 11	0.28	0.50	0.45	0.25	0.08	0.02	0.00	0.00	0.00	0.00	0.00	0.00
6 vs 11	0.29	0.33	0.28	0.22	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.00
8 vs 11	0.45	0.31	0.21	0.14	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00
6 vs 10	0.28	0.44	0.37	0.49	0.42	0.45	0.48	0.48	0.36	0.17	0.10	0.08

- Row names refer to the lines of table 4
- The null hypothesis is that the RMSFE are equal.

8 Concluding remarks

This paper deals with the issue of proposing a strategy to discover specific common cycles 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 using bivariate VAR models.

The strategy consists of testing for common cycles between all possible pairs of series and constructing groups in which almost all pairs show a common cycle. The 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 showed a good performance of our procedure in finite samples.

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

We applied the procedure to the US CPI broken down in 159 components. Our proposal for forecasting the aggregate is to do it indirectly, by constructing single-equation models for each component including the restrictions derived from the single-cycle subsets and, then, adding up the components' forecast. In a forecasting competition exercise we compared the ability of our procedure for forecasting the aggregate with other direct and indirect alternatives. The results show that disaggregation could be greatly relevant for forecasting, and it can be even better when it is done selectively. This means forecasting by single-equation models the components with single-cycle restrictions and the rest by a scalar model for its sub-aggregate. This points out that considering other common features that could involve more components, seems promising. In this way, we could discover by pairwise methods more complex relationships between the components and improve further the indirect forecast of the aggregate. In conclusion, if disaggregated information exists, it is not efficient to ignore it, and looking for single-feature subsets of components is a very useful and feasible strategy to exploit that information.

Acknowledgments

Both authors gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness research projects ECO2012-32401 and ECO2015-70331-C2-2. The first author also acknowledges the support form the Uruguayan Agencia Nacional de Investigación e Innovación research project FMV 3 2016 1 126200.

Appendix A Monte Carlo results for different relaxation parameters

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

	$S_1 = 10$		$S_1 = 25$		$S_1 = 40$	
	Gauge	*Potency*	*Gauge*	*Potency*	*Gauge*	*Potency*
$\boldsymbol{\lambda = 0}$ (no relaxation)						
$\beta = 0$, $Q = 0$ (independent innovations)						
$T = 400$	0.0	86.3	0.0	75.7	0.0	70.2
$T = 200$	0.0	84.4	0.0	72.0	0.0	65.8
$T = 100$	0.1	73.1	0.0	59.0	0.0	52.0
$\beta = -0.3$, $Q = 10$ (non zero correlation with 20 other innovations)						
$T = 400$	0.0	86.7	0.0	78.6	0.0	73.4
$T = 200$	0.0	86.6	0.0	76.5	0.0	70.9
$T = 100$	0.0	81.6	0.0	70.1	0.0	64.6
$\beta = -0.3$, $Q = 20$ (non zero correlation with 40 other innovations)						
$T = 400$	0.0	87.8	0.0	80.1	0.0	73.7
$T = 200$	0.0	86.2	0.0	77.8	0.0	73.0
$T = 100$	0.0	83.5	0.0	71.5	0.0	66.6
$\beta = -0.3$, $Q = 30$ (non zero correlation with 60 other innovations)						
$T = 400$	0.0	88.8	0.0	79.2	0.0	74.7
$T = 200$	0.0	87.6	0.0	78.2	0.0	73.1
$T = 100$	0.0	84.4	0.0	74.0	0.1	67.9
$\boldsymbol{\lambda = min[2, 0.4 \times \hat{S}_1^{\lambda=0}]}$						
	$S_1 = 10$		$S_1 = 25$		$S_1 = 40$	
	Gauge	*Potency*	*Gauge*	*Potency*	*Gauge*	*Potency*
$\beta = 0$, $Q = 0$ (independent innovations)						
$T = 400$	0.0	97.0	0.0	90.2	0.0	85.6
$T = 200$	0.0	95.7	0.0	86.8	0.0	81.0
$T = 100$	0.2	85.3	0.1	72.6	0.0	65.2
$\beta = -0.3$, $Q = 10$ (non zero correlation with 20 other innovations)						
$T = 400$	0.0	96.6	0.0	91.2	0.0	86.7
$T = 200$	0.0	96.2	0.0	89.8	0.0	84.7
$T = 100$	0.1	92.5	0.0	83.9	0.0	78.0
$\beta = -0.3$, $Q = 20$ (non zero correlation with 40 other innovations)						
$T = 400$	0.0	96.8	0.0	92.2	0.0	87.2
$T = 200$	0.0	95.8	0.0	90.6	0.0	86.0
$T = 100$	0.1	94.0	0.1	84.7	0.0	80.2
$\beta = -0.3$, $Q = 30$ (non zero correlation with 60 other innovations)						
$T = 400$	0.0	96.6	0.0	91.0	0.0	87.3
$T = 200$	0.0	96.1	0.0	90.4	0.0	86.5
$T = 100$	0.1	94.5	0.0	86.9	0.1	81.4

Table A.1 Continued:

	$S_1 = 10$		$S_1 = 25$		$S_1 = 40$	
	Gauge	*Potency*	*Gauge*	*Potency*	*Gauge*	*Potency*
$\boldsymbol{\lambda = min[5, 0.4 \times \hat{S}_1^{\lambda=0}]}$						
$\beta = 0$, $Q = 0$ (independent residuals)						
$T = 400$	0.0	97.9	0.0	94.9	0.0	92.3
$T = 200$	0.0	97.2	0.0	92.8	0.0	88.9
$T = 100$	0.2	88.2	0.1	80.9	0.1	73.8
$\beta = -0.3$, $Q = 10$ (non zero correlation with 20 other innovations)						
$T = 400$	0.0	97.1	0.0	95.4	0.0	92.7
$T = 200$	0.0	96.9	0.0	95.0	0.0	91.1
$T = 100$	0.1	93.3	0.1	90.5	0.0	85.4
$\beta = -0.3$, $Q = 20$ (non zero correlation with 40 other innovations)						
$T = 400$	0.0	97.4	0.0	95.9	0.0	92.6
$T = 200$	0.0	96.3	0.0	94.8	0.0	91.9
$T = 100$	0.1	94.6	0.1	91.2	0.0	86.9
$\beta = -0.3$, $Q = 30$ (non zero correlation with 60 other innovations)						
$T = 400$	0.0	97.2	0.0	94.6	0.0	92.6
$T = 200$	0.0	96.7	0.0	94.9	0.0	92.1
$T = 100$	0.1	95.2	0.1	92.5	0.1	87.7
$\boldsymbol{\lambda = min[10, 0.4 \times \hat{S}_1^{\lambda=0}]}$						
	$S_1 = 10$		$S_1 = 25$		$S_1 = 40$	
	Gauge	*Potency*	*Gauge*	*Potency*	*Gauge*	*Potency*
$\beta = 0$, $Q = 0$ (independent innovations)						
$T = 400$	0.0	97.9	0.0	95.4	0.0	93.8
$T = 200$	0.0	97.2	0.0	94.0	0.0	92.0
$T = 100$	0.2	88.2	0.2	83.1	0.1	78.7
$\beta = -0.3$, $Q = 10$ (non zero correlation with 20 other innovations)						
$T = 400$	0.0	97.1	0.0	96.1	0.0	94.8
$T = 200$	0.0	96.9	0.0	95.6	0.0	93.6
$T = 100$	0.1	93.3	0.1	91.6	0.0	89.2
$\beta = -0.3$, $Q = 20$ (non zero correlation with 40 other innovations)						
$T = 400$	0.0	97.4	0.0	96.5	0.0	94.5
$T = 200$	0.0	96.3	0.0	95.6	0.0	94.1
$T = 100$	0.1	94.6	0.1	92.2	0.1	90.4
$\beta = -0.3$, $Q = 30$ (non zero correlation with 60 other innovations)						
$T = 400$	0.0	97.2	0.0	95.2	0.0	94.6
$T = 200$	0.0	96.7	0.0	95.6	0.0	94.2
$T = 100$	0.1	95.2	0.1	93.4	0.1	91.0

See notes to table 2

Appendix B Monte Carlo results for false discovery

In this appendix we present the figures of false discovery corresponding to the Monte Carlo experiments of §5. For each experiment, scenario and sample size we count the number single cycle subsets in excess of the true number (which is always one) that is composed mainly by wrong series (50% or more wrong series). We denote the number of additional subsets as *Num ex.Subsets*. For each of these additional subsets we count the number of series which do not share the cycle (*Num Sers*). Note that when the first single cycle subset does not include all the correct series, the additional subsets may include some that share the cycle.

Table B.1 include the average across experiments of *Num ex.Subsets* and *Num Sers*. In this table we include the worst possible situation, as it refers to the maximum relaxation parameter (λ) considered in the paper. As the table shows, false discovery is not a relevant issue for any sample size or data structure.

Table B.1: False Discovery

	$\lambda = min[10, 0.4 \times \hat{S}_j^{\lambda=0}]$		
	Num ex.Subsets × Num. Sers	Num ex.Subsets × Num. Sers	Num ex.Subsets × Num. Sers
	$\beta = 0$, $Q = 0$ (independent innovations)		
$T = 400$	—	—	—
$T = 200$	—	—	—
$T = 100$	0.064 x 0.121	0.202 x 0.346	0.21 x 0.303
	$\beta = -0.3$, $Q = 10$ (non zero correlation with 20 other innovations)		
$T = 400$	—	—	—
$T = 200$	—	—	—
$T = 100$	0.02 x 0.046	0.068 x 0.107	0.048 x 0.088
	$\beta = -0.3$, $Q = 20$ (non zero correlation with 40 other innovations)		
$T = 400$	—	—	—
$T = 200$	—	—	—
$T = 100$	0.02 x 0.046	0.048 x 0.096	0.054 x 0.076
	$\beta = -0.3$, $Q = 30$ (non zero correlation with 60 other innovations)		
$T = 400$	—	—	—
$T = 200$	—	—	—
$T = 100$	0.01 x 0.026	0.034 x 0.059	0.066 x 0.11

References

Ando, T. and J. Bai (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics 31*(1), 163–191.

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

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bailey, N., S. Holly, and M. H. Pesaran (2016). A two stage approach to spatio temporal

analysis with strong and weak cross sectional dependence. *Journal of Applied Econometrics 31*(1), 249–280.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics 31*(6), 929–960.

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., J. A. Doornik, and D. F. Hendry (2012). Model selection when there are multiple breaks. *Journal of Econometrics 169*(2), 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.

Doornik, J. (2009). Autometrics. In *Castle and Shephard, The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 88–121. Oxford University Press.

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.

Guerrero, V. M. and D. Peña (2003). Combining multiple time series predictors: a useful inferential procedure. *Journal of Statistical Planning and Inference 116*(1), 249–276.

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

Hendry, D. and K. Hubrich (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. *Journal of business & economic statistics 29*(2), 216–227.

Johansen, S. and K. Juselius (2014). An asymptotic invariance property of the common trends under linear transformations of the data. *Journal of Econometrics 178*, 310–315.

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In *Castle, and Shephard (2009), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 1–36. Oxford University Press.

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.

Li, X.-L., T. Chang, S. M. Miller, M. Balcilar, and R. Gupta (2015). The co-movement and causality between the us housing and stock markets in the time and frequency domains. *International Review of Economics & Finance 38*, 220–233.

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.

Osborn, D. R., A. P. Chui, J. P. Smith, and C. R. Birchenhall (1988). Seasonality and the order of integration for consumption*. *Oxford Bulletin of Economics and Statistics 50*(4), 361–377.

Saikkonen, P. (1992). Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric theory 8*(1), 1–27.

Saikkonen, P. and H. Lütkepohl (1996). Infinite-order cointegrated vector autoregressive processes. *Econometric theory 12*(5), 814–844.

Santos, C., D. Hendry, and S. Johansen (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics 23*(2), 317–335.

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.