Discovering specific common trends in a large set of disaggregates: statistical procedures, their properties and an empirical application*

Guillermo Carlomagno[a] and Antoni Espasa[b]

[a](*e-mail:gcarlomagno@bcentral.cl*) Central Bank of Chile and Center of Economic Research (CINVE)

[b](*e-mail:antoni.espasa@uc3m.es*) Department of Statistics and Instituto Flores de Lemus, University Carlos III of Madrid

Abstract

Macroeconomic variables are weighted averages of a large number of components. Our objective is to model and forecast all of the N components of a macro variable. The main feature of our proposal consists of discovering subsets of components sharing single common trends neither assuming pervasiveness of those trends, nor imposing special restrictions on the serial or cross-sectional idiosyncratic correlation. We adopt a pairwise approach and study its statistical properties. Our asymptotic theory works both, with fixed N and $T \to \infty$, and with $[T, N] \to \infty$. An extension of our strategy allows a wide type of conditional and unconditional heteroskedasticity. The paper includes an application to the breakdown of the US CPI in 159 components.

Total number of words (inclusive of footnotes, references and tables): 10980

Keywords: Cointegration, Factor Models, Disaggregation, Pairwise tests, Heteroskedasticity.

JEL: C01, C22, C32, C53.

*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.

I Introduction

Macroeconomic variables are weighted averages of a large number of components, thus, the usual focus on the aggregate alone implies neglecting a large amount of information. The objective of this paper is to develop a procedure to model and forecast all of the components of a macro or business variable at the maximum level of disaggregation. Our strategy consists of identifying and estimating relevant relationships between the components (disaggregates) and then exploiting those relationships in single-equation models for the disaggregates. This strategy can be useful in achieving two purposes: (i) providing relatively precise models and accurate forecasts of the components, our major interest; and (ii) generating an improved indirect forecast for the aggregate, in the sense that it is not significantly worse than direct forecasts. Attaining purpose ii, would be an indirect validation of the strategy for achieving purpose i.

When working with an aggregate composed of a large set of disaggregates, one of the main challenges in econometric modeling is how to deal with the trade-off between informational losses (when components are not considered) and estimation uncertainty (due to the increased number of parameters to be estimated when disaggregates are used). The informational losses would increase with the differences in the statistical distributions of the components, and the estimation uncertainty could be mitigated by using appropriate restrictions for the data.

One possibility for dealing with the informational losses vs. the estimation uncertainty trade-off is the consideration of common features, as proposed by Espasa and Mayo-Burgos (2013). These authors argue that when analyzing the components of a macro variable, it is usual to observe that while some components share features such as trends or cycles, others do not, probably because they incorporate changes in technology or in the preferences of economic agents in different ways. Thus, as Espasa and Mayo-Burgos argue, a valid hypothesis may be that specific subsets of components share common features, while others do not.

To exploit the restrictions derived from the existence of those subsets, Espasa and Mayo-Burgos suggest restrict the analysis to discover subsets of components such that all the

elements in a subset share a unique common feature (trends and cycles — see Engle and Kozicki, 1993 for a definition of common features), and then including the restrictions implied by those commonalities in single-equation models for the components. In this paper we focus just on common trends.

Castle and Hendry (2010) also highlight — as Mayo and Espasa (2009) propose[1] — the importance of including long and short-run common features restrictions in the individual models for the components.

The problem of how to discover unknown restrictions in multivariate models is also present in the Dynamic Factors Models (DFM) literature. Several authors have shown that if the data contain non-pervasive factors (factors that are common only to a reduced subset of series), results are more accurate when factors are extracted from data that are informative about them (see, e.g., Boivin and Ng, 2006 and Beck et al., 2015). Some proposals to deal with non-pervasive factors can be found in 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), BHP hereafter, and Ando and Bai (2015).

The closer approaches to ours are those of Ando and Bai (2015), BKP, and BHP. In all these cases, the authors restrict their attention to stationary series, assume that the cross-sectional dimension goes to infinity, and require the usual restrictions of DFM on the serial and cross-correlation of idiosyncratic components (see e.g., assumption C in Bai and Ng, 2002). All these assumptions do not fit our framework of interest.

We face the problem of identifying possibly small subsets of components that share just one common trend. Apart from dealing with $I(1)$ variables, our strategy has three additional advantages which derive from the fact that it does not rely on any type of cross-sectional averaging method. First, we do not need the trends to be pervasive. Second, the cross-sectional dimension of the subsets that have the non-pervasive trends does not need to go to infinity. Our theory works both, with fixed N and $T \to \infty$, and with $[T, N] \to \infty$. Third, as we do not need idiosyncrasies to average out, we do not need to impose special restrictions on idiosyncratic serial or cross-correlation.

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

An important contribution of this paper is to show that subsets of components sharing single common trends can be discovered by pairwise methods. For determining the cointegration rank in the pairwise models we consider both the Johansen's approach and procedures based on information criteria, which are robust to a wide type of heteroskedasticity. Additionally, in the accompanying online appendix we show that our strategy can be generalized to consider special situations in which some subsets may include two common trends, one '*general*' and one '*sectorial*'.

The rest of the paper is organized as follows. In §II, we study the statistical foundations of the pairwise procedure when using the Johansen's test, and in §III we extend the analysis for the information criteria-based approach. In §IV we present Monte Carlo evidence and in §V we apply the pairwise procedure to the US CPI broken down in 159 components. §VI is devoted to the conclusions.

II Statistical foundations of the pairwise procedure

General framework and assumptions

The general framework for the models we work with is given by a VAR model for all of the N components of an aggregate:

$$X_t = \mu_t + \Pi_1 X_{t-1} + ... + \Pi_k X_{t-k} + \epsilon_t \;\Rightarrow \Pi(L)X_t = \mu_t + \epsilon_t, \tag{1}$$

where X_t is a $N \times 1$ vector; Π_i are $(N \times N)$ coefficient matrices; ϵ_t is a vector of iid innovations with covariance matrix Σ_t; μ_t contains the deterministic components (constants, trends, seasonal dummies, and outliers and breaks indicators); $\Pi(z)$ is the characteristic polynomial; and L is the lag operator. If the system is cointegrated, it can be rewritten as a Vector Equilibrium Correction Model (VECM):

$$\Delta X_t = \mu_t + \alpha\beta' X_{t-1} + \Phi_1 \Delta X_{t-1} + ... + \Phi_{k-1}\Delta X_{t-k+1} + \epsilon_t, \tag{2}$$

where α and β are $N \times r$ matrices, with $0 < r < N$; r is the number of cointegration relationships; $\alpha\beta' = -I_n + \Pi_1 + ... + \Pi_k$; *and* $\Phi_i = -\sum_{j=i+1}^{k} \Pi_j$. The data structure for which our procedure is designed can be summarized in five assumptions:

Assumption 1 The N components are generated by the VECM in eq. (2).

Assumption 2 The N components are $I(1)$.

Assumption 3 There are J subsets (with $J > 0$), each one of size n_j (with $0 < j \leq J$), such that the components of each subset share a unique common stochastic trend. Thus, in each of those subsets there are $n_j - 1$ cointegration relationships and we denote them as *fully cointegrated subsets*.

Assumption 4 The innovations ϵ_t of eq. (2) are iid zero mean and multivariate normally distributed, which implies $\Sigma_t = \Sigma$. (this assumption is modified in §III).

Assumption 5 $\frac{T}{N^{1/\kappa}} \to \geq c$, when $[T, N] \to \infty$, for some $c > 0$ and $\kappa > 0$.

Assumption 6 $\frac{n_j}{N^{1-1/\kappa}} \to \geq c$, when $[T, N] \to \infty$, for $0 < j < J$, and some $c > 0$ and $\kappa > 0$.

Remark 1 Instead of Assumption 2, we could use the usual assumption that $\alpha_\perp \Phi(1) \beta_\perp$ is of full rank, which will ensure that the components are *at most* $I(1)$. This new assumption, say assumption 2 bis, would require proceeding as Johansen (1995) suggests: testing the significance of the cointegration relationships' coefficients (β) to detect $I(0)$ variables (that is how we proceed in the empirical application). Thus, Assumption 2 is only needed for the simulation exercises of §IV, for the rest of the paper we only need assumption 2 bis.

Remark 2 In principle, assuming that the innovations are iid would be enough for the Johansen's procedure to be asymptotically valid, so that the normality imposed in assumption 4 would be unnecessary. However, as it will become clear later in this section, normality of ϵ_t is necessary to guarantee the validity of the inference from the bi-variate models in which we test cointegration.

Remark 3 By allowing for outliers and location shifts in eq. (2), the identical distribution established in assumption 4 could be achieved after correcting for outliers and/or location shifts which, as Juselius (2015) argues, is a quite general assumption for macroeconomic VAR models.

Remark 4 Apart from having all its roots outside the unit circle, there is no restriction on the polynomial $(I - \Phi_1 L - ... - \Phi_k L)$, additionally the covariance matrix Σ has no particular restrictions. Thus, we are not imposing any additional restriction on the serial or cross-correlation of the components.

Remark 5 Even after considering remark 3, the present framework does not allow for heteroskedasticity in ϵ_t. As we argue in §V, although this is not an issue for our empirical application, it could be so in other cases. Therefore, in §III we generalize our proposal for the case of heteroskedastic innovations.

Remark 6 As we discuss latter in this section, we use assumptions 5 and 6 to control gauge and false discoveries when we let N going to infinity.

The pairwise strategy consists of three steps: (i) Test for conintegration in all of the $N(N-1)/2$ pairs of series with the Johansen's test using a nominal size φ (in §III we consider methods based on information criteria). (ii) Look for the largest subset of series in which all of them are pairwise cointegrated (fully cointegrated subset). (iii) Repeat previous step but excluding the series already included in the fully cointegrated subsets already discovered until no more fully cointegrated subsets are found.

We use the notation n_j both as the name of the fully cointegrated subsets that exist in the true DGP and as their cardinality. For the subsets constructed by the pairwise procedure we use the notation $\hat{n}_j$.

For each pair of components, a bivariate VAR model has to be estimated, and the lag length must be determined in each case using some information criteria. Another important decision when performing cointegration tests concerns the deterministic structure in the model, which is particularly relevant for trending variables. We adopt the two step procedure of Nielsen and Rahbek (2000); first, test the cointegration rank in a VECM that

includes all the deterministic components (constant, trends and interventions) in the cointegrating relationships and their differences in the VAR. Second, once the cointegration rank is determined, test the hypothesis that the coefficients of the deterministic components are zero. Given the asymptotic similarity of the cannonical correlations with respect to the coefficient of the trend (see Nielsen and Rahbek, 2000), it would be unnecessary to re-estimate cointegration relationship if that coefficient turns out to be not significant. Juselius (2006) adopts this strategy too (see also Doornik et al. (1998) for a discussion on this issue).

Using the outcomes of the cointegration tests and the resulting fully cointegrated subsets, the final phase of our proposal is to estimate a single-equation model for each component, including as potential regressors all of the possibly relevant cointegration relationships found in the previous step, if any, as well as each component's own lags and lags of other components. The selection of the relevant regressors can be carried out by the model selection algorithm *Autometrics* (see Doornik, 2009). This modeling strategy is something intermediate between the full vector model — which is unfeasible in our context of large N— and the univariate estimation of each component. Finally, the single-equation models can be used for forecasting all the components.

As argued by Zivot and Wang (2007), this single-equation strategy is justified by the fact that the $I(0)$ cointegration relationships, that are included as regressors, can be treated as known given $\beta's$ estimators super consistency.

Asymptotic properties of the pairwise procedure

The properties of the pairwise procedure for discovering fully-cointegrated subsets must be evaluated in three dimensions: i) *Potency:* The proportion of correct series that are included in $\hat{n}_j$. ii) *Gauge:* The proportion of wrong series that are included in $\hat{n}_j$[2]. iii) *False discovery:* The discovery of subsets in which none of the pairs are cointegrated.

Potency

In order to include all of the correct series in $\hat{n}_j$ we should find one cointegration relationship in all of the $n_j(n_j-1)/1$ pairs that exist in the true subset.

Since the pairwise procedure involves a large number of cointegration tests (e.g., 4950 for

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

$N = 100$), it may be thought to raise the probability of false rejection, which may affect the potency of the procedure. Specifically, if the two series of the pair belong to the same n_j, the problem could be wrongly rejecting the true hypothesis of $r = 1$ — one cointegration relationship. In what follows we study this issue in detail.

In the Johansen's procedure the null hypotheses $r = 0$ and $r = 1$ are usually tested sequentially. Since the asymptotic power of Johansen's test is 1, finding no cointegration between pairs in the same n_j is not an issue when n_j is fixed and $T \to \infty$. Therefore, the problem could be false rejecting $r = 1$ in favor of $r = 2$. If the tests were independent, the probability of finding one common trend between *all* of the series in the same n_j would be $(1 - \varphi)^{n_j(n_j-1)/2}$, which quickly decreases with n_j. But, as Theorem 1 indicates, these tests are asymptotically equivalent, in the sense that the probability of finding the same cointegration rank in all of them tends to one as T goes to infinity.

Theorem 1 (Asymptotic equivalence of pairwise cointegration tests in a fully cointegrated subset). Under assumptions 1 to 4, given a subset of n_j pairwise cointegrated series (i.e., there are $n_j - 1$ cointegration relationships among them and a single common trend), for a fixed n_j, the probability of finding the same cointegration rank by means of the Johansen's test in all of the $n_j(n_j - 1)/2$ pairs tends to 1 as T goes to infinity, when all the individual tests are performed with the same significance level.

Proof See Appendix A

Theorem 1 states that the joint probability of rejecting $r = 1$ for any number of pairs between series in n_j tends to φ. Therefore, even in the case in which the hypothesis of interest is the *universal* one (i.e., false rejecting at least one of the $n_j(n_j - 1)/2$ hypotheses) defined in the approaches that try to control multiple testing problems (see e.g., Romano and Wolf, 2005), p-values need not be corrected.

This result is in line with Johansen and Juselius (2014), who show that the random walk components of linear combinations of a vector of variables are cointegrated with those of the original (big) system.

The following proposition deals with the case of $n_j \to \infty$.

Proposition 1 Under assumptions 1 to 4, given a subset of n_j pairwise cointegrated series, the expected proportion of pairwise Johansen's tests that deliver the same cointegration rank tends to 1 as $T \to \infty$, both for fixed n_j and when it goes to infinity.

Proof See Appendix B.

Gauge

We now analyze the problem of including wrong series in the estimated fully cointegrated subsets ($\hat{n}_j$).

Proposition 2 Under assumptions 1 to 6, the expected proportion of wrong elements in the estimated fully cointegrated subsets can be controlled by setting by setting $\varphi = N^{-1/\kappa}$ (for some $\kappa > 0$), both when N is fixed and when it goes to infinity.

Proof See Appendix C.

Note that this proposition does not require pervasiveness in the sense of DFM (see, e.g., Assumption B in Bai, 2003), it just needs assumption 6. It is also to be noted that the use of $\varphi = N^{-1/\kappa}$ requires Assumption 5 to avoid power problems.

False discovery

A false discovery occurs when none of the pairs included in $\hat{n}_j$ are truly cointegrated. The distinction with previous argument is relevant since it allows to explicitly analyze how the pairwise procedure would work in a situation where there are no cointegrated pairs.

Proposition 3 Setting the nominal size of the Johansen's tests as $\varphi = N^{-1/\kappa}$ for $\kappa > 0$, the expected number of false fully cointegrated subsets of size λN (subsets with cardinality smaller than or equal to λN, with $0 < \lambda < 1$, in which there are none cointegrated pairs but pairwise cointegration tests indicate that all pairs are cointegrated) tends to zero at a rate larger than or equal to that of $N^{(1-\lambda N)/\kappa}$ as T and N go to infinity.

Proof See appendix D.

This result implies that choosing φ as an inverse function of N and disregarding small fully cointegrated subsets (those with cardinality smaller than or equal to λN) constitutes a strong protection against false discovery.

In obtaining previous results for gauge and false discovery, the fact that for including wrong series we need not to reject $r = 1$ with *all* of the series in $\hat{n}_j$ plays a crucial role.

The bi-variate systems

The strategy described above requires the estimation of systems which are linear transformations of a larger one because we assume the existence of a large VAR model for all of the components but estimate several partial bi-variate systems in which we test cointegration.

Linear transformations of a process X_t that follow a (possibly cointegrated) VAR model have an infinite VAR representation with exponentially decreasing coefficients' matrices. If the innovations of the infinite bi-variate VAR representation (say, e_t) are iid, and the fitted models have a lag length that increases with T, provided that the used lag length is $Op(T^{1/3})$, usual asymptotic inference is still valid (see, Saikkonen, 1992, Saikkonen and Lütkepohl, 1996 and Johansen and Juselius, 2014). However, as noted by Johansen and Juselius (2014), e_t is a white noise process which need not be iid, as we assumed for ϵ_t (the innovations of the large model). The case of ϵ_t being iid Gaussian ensures e_t also to be iid Gaussian.

Thus, in order to guarantee that our bi-variate models are valid for inference, we need Gaussian iid innovations in the model of eq. (2). This explains assumption 4 (see remark 3).

Apart from previous discussion, the strategy of testing for cointegration in bi-variate models could be thought to imply a loss of power compared to a 'complete' model approach (when feasible). Interestingly, Monte Carlo results (not reported) show that when the cointegration relationships are pairwise detectable, the power of the pairwise procedure for finding the true number of cointegration relationships is higher than that of the traditional Johansen's trace test. This result is in line with the fact that cointegration relationships are more easily detectable in systems with fewer stochastic trends (see, e.g., Lütkepohl et al. (2003) and Johansen (1995)).

III Extension to the case of unconditional heterskedastic innovations

There is a large body of applied literature suggesting that the unconditional variance of the shocks driving the behavior of macroeconomic time series is not stable over time. Most of the empirical analysis find a decline in the variance of US macro variables during the so called 'great moderation' period that started in the mid 80's (see, *inter alia*, Kim and Nelson, 1999, McConnell and Perez-Quiros, 2000, and Sensier and Van Dijk, 2004).

In this section we describe how our procedure has to be modified to deal with conditionally and/or un-conditionally heteroskedastic innovations and analyze the properties of the modified procedure.

Cavaliere et al. (2010) analyze the impact of non-stationary volatility in the innovations of cointegrated VAR models on the Johansen's cointegration tests, and show that inference based on standard homoskedastic critical values is, in general, invalid.

Cavaliere et al. (2016) study the properties of procedures based on information criteria for determining the lag length and the cointegration rank in VAR models with innovation processes that allow a wide class of conditional and un-conditional heteroskedasticity, including multiple covariance shifts, variances with broken trends, smooth variance shifts, and GARCH and stochastic volatility processes. Specifically, the assumption about the innovations is as follows:

Assumption 7 Denoting e_t the innovations of the bi-variate systems in which we test for cointegration, $e_t = z_t\sigma_t$, where σ_t is non-stochastic and defined by $\sigma(t/T)$ for $t = 1 \dots T$, where $\sigma(.)$ belongs to the space of $N \times N$ matrices of càdlàg functions on $[0, 1]$, and z_t is a martingale difference sequence with respect to the filtration F_t, with conditional variance matrix $h_t = E(z_t z_t'|F_t)$ satisfying $sup_t E||z_t||^{4r} < \infty$, for some $r > 1$ and $T^{-1}\sum_{t=1}^{T} h_t \xrightarrow{p} E(z_t z_t') = I_p$, where $\xrightarrow{p}$ denotes convergence in probability as $T \to \infty$.[3]

[3] As showed by Johansen and Juselius (2014) e_t is a linear process in ϵ_t, the innovations of the full system. Thus, this assumption relaxes assumption 4 in the sense that we do not need ϵ_t of eq. (2) to be normal iid anymore. For example if ϵ_t is a non-stationary independent sequence of $N(0, \Sigma_t)$ with $\Sigma_t = I_N$ up to $t = t^*$ and $\Sigma_t = \lambda I_N$, for some $\lambda > 0$, for $t > t^*$ (case C in the numerical results of Cavaliere et al. (2016)), e_t maintains this property.

The procedure consists of selecting the optimal lag length ($\hat{k}$) and cointegration rank ($\hat{r}$) either sequentially or jointly according to a generic information criterion IC. In the joint procedure, $\hat{k}$ and $\hat{r}$ are chosen as follows:

$$(\hat{k}, \hat{r}) = \underset{r=0,...N;k=0,...K}{arg\ min} IC(k, r), \tag{3}$$

where K denotes a given maximum number of lags.

Under assumption 7, provided that the variables in the system are at most $I(1)$ and that the penalty function (p_T) related to the information criterion satisfies $\frac{p_T}{T} + \frac{1}{p_T} \to 0\ as\ T \to \infty$, Cavaliere et al. (2016) show that $(\hat{k}, \hat{r}) \xrightarrow{p} (k_0, r_0)$, where k_0 and r_0 denote the true lag length and cointegration rank, respectively (see theorem 1 in Cavaliere et al., 2016). Among the three most widely used information criteria (AIC, BIC and HQ), the AIC is the only one which does not satisfy the required condition on its penalty function.[4]

In the following three propositions we show that our pairwise approach would remain valid when determining the cointegration rank by means of information criteria instead of Johansen's tests. But before that it is important to mention that we will assume that the results of Cavaliere et al. (2016), which are derived for finite order VAR models, are valid for the bi-variate models in which test for cointegration that, as discussed at the end of previous section, are approximations to infinite models with exponentially decreasing coefficient matrices. Though providing theoretical results on this issue is out of the scope of this paper, the fact that procedures based on information criteria remain robust when the lag length is under-specified (see Cavaliere et al. (2016)), and our simulation results in next section, suggest the procedure based on information criteria is still valid in our case of interest.

Proposition 4 Under assumptions 1 to 3 and 7 given a subset of n_j pairwise cointegrated series (i.e., there are $n_j - 1$ cointegration relationships among them and a single common trend), when determining the cointegration rank of the pairs by means of an information criterion whose penalty function (p_T) satisfies $\frac{p_T}{T} + \frac{1}{p_T} \to 0\ as\ T \to \infty$, the probability of

[4]Interestingly, Cavaliere et al. (2018) show that information criteria procedures remain valid for partial systems that condition on the excluded variables, independently of whether the conditioning variables are weakly exogenous or not.

finding a cointegration rank equal to 1 in all of the $n_j(n_j-1)/2$ pairs tends to 1 as T goes to infinity, for a fixed n_j. When $n_j \to \infty$ the expected proportion of pairs that deliver a cointegration rank of 1 tends to 1 as $T \to \infty$.

The proof follows directly from theorem 1 of Cavaliere et al. (2016) and the proof of proposition 1.

Proposition 5 Under assumptions 1 to 3, and 7, when $T \to \infty$ and N is fixed the expected proportion of wrong inclusions in the estimated fully cointegrated subsets ($\hat{n}_j$) constructed by means of tests based on information criteria, tends to zero. When $(T, N) \to \infty$, restricting the ratio N/n_j to be $Op(log(T))$ (when using the BIC) or $Op(log(log(T)))$ (when using the HQC) ensures that the expected proportion of wrong inclusions tends to zero as $[T, N] \to \infty$

Proof See appendix E.

Proposition 6 Under assumptions 1, 2 and 7, when estimating the cointegration rank by means of an information criterion, the expected number false fully cointegrated subsets of size λN(subsets of λN $I(1)$ series in which there are none cointegrated pairs but pairwise cointegration tests indicate that all pairs are cointegrated) tends to zero as T and N go to infinity (for any $0 < \lambda < 1$).

The proof follows directly from the proof of proposition 3 and theorem 1 of Cavaliere et al. (2016).

The three previous propositions indicate that the cointegration rank of the pairs could also be determined by means of procedures based on information criteria. The fact that these procedures are robust to a wide class of conditional and un-conditional heteroskedasticity implies that our procedure would remain valid in those cases. Additionally, Cavaliere et al. (2010) shows that the Gaussian pseudo maximum likelihood estimation of the parameters of the cointegrated VAR remains consistent under assumption 7. This implies that the single-equation forecasting strategy described §II would remain valid under assumption 7 when using the approach based on information criteria.

IV Simulations

The Monte Carlo experiments in this section are designed with the objective of analyzing the finite samples behavior of our procedure both when determining the cointegration rank by means of the Johansen's test and when using information criteria procedures.

General design of the experiments

We consider three alternative DGPs.

DGP 1

The general expression of the VECM for the N series is eq. (2) with only one lag, $\mu_t = \alpha c_0$ (i.e., the series do not have deterministic trends). We simulate a situation in which a subset n_1 of the N components share a unique common trend, and the rest of the components have their own trends. Thus, we will have $N - n_1 + 1$ 'common' trends in the system. Without loss of generality, we set matrix β such that:

$$\beta' = \begin{pmatrix} \beta_2 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \beta_3 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & & & & & \\ \beta_{n_1} & 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}_{r \times N},$$

where $r = n_1 - 1$. This normalization was suggested by Clements and Hendry (1995). Different normalizations change the exact shocks that drive the long-run behavior of the n_1 variables, but not the fact that they are determined by $N - r$ shocks and r adjusting mechanisms. The parameters β_j are all equal to -1, for $j = 2, ..., n_1$.

For the sake of simplicity, in this DGP matrix α is set to have the following structure:

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\alpha_2 & 0 & 0 & \cdots & 0 \\ 0 & -\alpha_3 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & -\alpha_{n_1} \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{N \times r} = \begin{pmatrix} \alpha^U \\ \alpha^D \end{pmatrix}, \tag{4}$$

where sub-matrix α^U is $n_1 \times r$; sub matrix α^D is a matrix of zeros with dimensions $(N-n_1)\times r$; and the values α_i are taken from the uniform distribution with parameters $[0.15, 0.3]$. These parameters are motivated by results in Espasa and Mayo-Burgos (2013) for CPI series.

In this case, the long run impact matrix ('C') related to the first n_1 variables is full of zeros, except for the first column. This simplicity of the long run structure is eliminated in DGP 3.

In this first DGP, Φ_1 is a diagonal matrix whose diagonal elements are drawn from the uniform distribution with parameters [0.5, 0.8]. Given that the residuals' covariance is also diagonal, there is no short run correlation between any of the N series. This simplicity is eliminated in DGPs 2 and 3.

For the innovations, we consider two possibilities, one with iid Gaussian innovations, and one with unconditionally heteroskedastic ones. In the first case (DGP1-iid), $\epsilon_t \sim N(0, I_N)$. The second case (DGP 1-hetero) mimics the case C of Cavaliere et al. (2016), where ϵ_t is a non-stationary independent sequence of $N(0, \Sigma_t)$ processes, where $\Sigma_t = I_N$ for $t < \lfloor 0.5T \rfloor$ and $\Sigma_t = 3I_N$ for $t \geq \lfloor 0.5T \rfloor$, with $\lfloor A \rfloor$ being the operator that takes the integer part of number A.[5]

[5]Since the innovations of the bi-variate systems, say e_t, are linear processes in ϵ_t, the covariance matrix of e_t maintains the property of being constant up to $t = \lfloor 0.5T \rfloor$ and increasing by a factor of 3 from then on.

DGP 2

DGP 2 is the same as DGP 1, except that we allow for some short-run interactions by including non-zero coefficients in the off-diagonal elements of matrix Φ_1. To do so, we first reorder the rows of matrices β and α to have the series in n_1 in positions $\lfloor iN/n_1 \rfloor$, for $i = 1, ..., n_1$. Then, denoting ϕ_{ij} the elements of Φ_1, we set:

$$\phi_{ij} = \begin{cases} p_i & \text{if } i = j, \\ u_i & \text{if } i \neq j, \text{ and } max(i-q,1) \leq j \leq min(i+q,N), \\ 0 & \text{otherwise,} \end{cases}$$

for $i, j = 1, ..., N$. Parameters p_i are taken from the uniform distribution $U_{[0.4,0.75]}$; $|u_i|$ is taken from the uniform distribution $U_{[0.05,0.1]}$; and

$$q = \begin{cases} 5 & \text{if } 5 < i < N-5, \\ 10-i & \text{if } i \leq 5, \\ 10-(N-i) & \text{if } i \geq N-5. \end{cases} \tag{5}$$

In this way, each series has non-zero short-run dependence with other nine (see Bai and Ng, 2002 for a similar strategy to generate short-run dependence). This DGP generates quite complex short-run dynamics, and there are no bivariate subsystems with purely finite VAR structures.

We consider the cases of iid and unconditionally heteroskedastic innovations in the same way as in DGP 1. Thus, we also have DGP 2-iid and DGP 2-hetero.

DGP 3

This DGP is the same as DGP 2 (non-diagonal Φ_1), except that we consider a more complex matrix α. We set:

$$\alpha = \begin{pmatrix} \alpha_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0\cdots & 0 \\ -\alpha_{21} & -\alpha_{22} & \cdots & -\alpha_{2,q_2} & 0 & 0 & 0 & 0\cdots & 0 \\ 0 & -\alpha_{32} & -\alpha_{33} & \cdots & -\alpha_{3,q_3+1} & 0 & 0 & 0\cdots & 0 \\ 0 & 0 & -\alpha_{43} & -\alpha_{44} & \cdots & -\alpha_{4,q_4+2} & 0 & 0\cdots & 0 \\ \vdots & & & & & & & & \\ 0 & \cdots & 0 & -\alpha_{i,i-1} & -\alpha_{i,i} & \cdots & -\alpha_{i,q_i+i-2} & 0\cdots & 0 \\ \vdots & & & & & & & & \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0\cdots & -\alpha_{n_1,n_1-1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \\ \vdots & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \tag{6}$$

where:

$$q_i = \begin{cases} \lfloor 0.3 \times n_1 \rfloor & \text{if } \lfloor 0.3 \times n_1 \rfloor + i - 2 < n_1 - 1, \\ i - (n_1 - 1) & \text{otherwise} \end{cases} \tag{7}$$

In this way, there are no weakly exogenous variables inside n_1, each variable reacts to $0.3 \times n_1$ cointegration relationships (except for the first one and those in the last positions). For instance, with $n_1 = 40$, each variable reacts to 12 cointegration relationships. Therefore, the long run impact matrix related to the variables in n_1 does not have zeros anymore.

As for previous DGPs, $\alpha_{i,i-1}$, for $i = 2, ..., n_1$, is taken from the uniform distribution with parameters $[0.15, 0.3]$. To avoid explosive patterns, α_{ij} for $j \neq i - 1$ is taken from the uniform distribution with parameters $\left[\frac{0.15}{q_i/2}, \frac{0.3}{q_i/2}\right]$.

Again, we allow the possibility of iid and unconditionally heteroskedastic innovations in the same way as in DGP 1 and DGP 2. Thus, we also have DGP 3-iid and DGP 3-hetero.

For the six DGPs described above, we consider three scenarios. In all of them, we set $N = 100$, and they differ in the choice of n_1 — recall that we are using the notation n_1 both as an indication of the size of the 'fully cointegrated subset' and as its label. The three choices are $n_1 = 10$, $n_1 = 25$ and $n_1 = 40$. Additionally, we consider three possible sample

sizes: $T = 100$, $T = 200$ and $T = 400$.

For each DGP, scenario and sample size, we perform 500 Monte Carlo replications. In each replication, the objective is to discover the series in n_1 by means of the procedure described in II (we disregard subsets with 4 or less series for avoiding false discoveries).

At this point, a word about the number of tests included in the Monte Carlo experiments and the computing time is in order. For a particular DGP, scenario and sample size we have 2.475 million sub-models (4950 pairs of series for each of the 500 replications). Since we have six DGPs, three scenarios and three sample sizes we have $(6 \times 3 \times 3) \times 2.475 = 133.650$ million sub-models to estimate. Additionally, the lag length for each of the 4950 sub-models of a particular replication is unknown. When using the Johansen test, we select it with the AIC in a model with one cointegration relationship and admitting between one and three lags in the VECM representation, which multiplies the number of sub-models to estimate by three. When using the information criteria procedures, we use the BIC to choose between the 9 possible models that emerge when considering cointegration ranks between 0 and 2 and lag lengths between 1 and 3, this multiplies the number of models by 9.

In a personal computer equipped with Windows 10 64 bits, with a processor i7-6700HQ of 2.60Hz, and 16GB of ram it takes 45 seconds to perform the 4950 pairwise Johansen's tests selecting models between 1 and 3 lags. When using the information criterion procedure the joint determination of the cointegration rank and the lag length of the 4950 models takes 95 seconds.

Monte Carlo results

iid innovations

Gauge and potency of the pairwise approach for the three DGPs with iid innovations are included in table 1. Since false discovery is very low for all scenarios and sample sizes, we omit these results in this part of the paper and include them in appendix G.

Focus first on the first panel of the table (DGP 1) and the case of Johansen's tests. The pairwise procedure applied with Johansen's test performs reasonably well for all scenarios and sample sizes. For $T = 400$, the probability of including all of the correct series is close to 99%. This outcome is in line with Theorem 1, which implies that the asymptotic probability

of finding a cointegration rank equal to 1 in all of the true cointegrated pairs tends to $(1-\varphi)$ as $T \to \infty$, with φ being the nominal size of the individual tests (we are using $\varphi = 1\%$).

On the other hand, the number of wrong series is quite low. For example, in scenario 3, for $T = 400$, the expected number of wrong series is $0.002 \times (100 - 40) = 0.12$. In appendix C we indicate that an upper bound for the expected number of wrong series in $\hat{n}_1$ is $(N - n_1)\varphi$; which means 0.6 series in scenario 3. This result shows that the actual average number of wrong series is far from this upper bound, meaning that the assumption that lead to that bound is quite extreme.

Finally, although gauge remains rather stable when the sample size changes, potency deteriorates as T decreases. For instance, with $n_1 = 25$ (scenario 2), we go from a potency of 99% with $T = 400$ to 67% with $T = 100$. Still, in this case, we get a low gauge and capture 67% of the correct series.

Focus now on the case of determining the cointegration rank by the BIC. Results are in line with the theory included in §III. In that section we showed that the probability finding $r = 1$ between all of the truly cointegrated series is asymptotically 1 and the probability of including wrong series is asymptotically 0, which is in line with the gauge and potency of table 1 for $T = 200$ and $T = 400$. For $T = 100$ we also observe a relevant deterioration, particularly in gauge.

The comparison of the information criteria procedure with the one based on Johansen's tests shows that although both performs very well in large samples ($T = 200$ and $T = 400$), the former is somewhat better. In short samples the information criteria approach dominates in terms of potency and the Johansen's alternative in terms of gauge.

Second panel of table 1 includes the results for DGP 2. The performance of the Johansen's-based procedure is somewhat worse than for DGP 1, except for scenario 3 with $T = 100$, where the deterioration in terms of potency is substantial. In this situation (scenario 3 and $T = 100$), we include, on average, $0.26 \times 40 = 10.4$ correct series and almost no wrong ones $(0.004 \times (100 - 40) = 0.24)$. As the sample size increases, results improve in terms of gauge and potency. For $T = 400$, the procedure almost recovers its performance of DGP 1 in all scenarios. These comments are also valid for the information criteria-based procedure.

The comparison of the information criteria procedure with the one based on Johansen's tests shows, again, that although both performs very well in large samples ($T = 200$ and $T = 400$), the former is somewhat better, except for scenario 3. In short samples the information criteria procedure dominates.

The third panel of table 1 includes the results for DGP 3, which are similar to those of DGP 2 for both types of procedures. We get a deterioration in small samples, but as the sample size increases we recover the results of DGP 1. Although in small samples the deterioration in terms of gauge of the Johansen-based procedure seems important, the figures in the table imply an average of just 0.7, 1.2 and 0.5 wrong series in scenarios 1, 2 and 3, respectively.

As discussed in the introduction, an alternative strategy to ours may be the estimation of Dynamic Factor Models. As we argued, since we are dealing with non-pervasive factors, small sizes of the groups, and serially and cross-correlated idiosyncrasies, DFM cannot be expected to show a good performance. Still, as the DFM assumptions about pervasiveness, and serial 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 (available upon request) show that our procedures (both the Johansen and the information criteria based) dominate in almost all situations.

Unconditionally heteroskedastic innovations

As argued in §III the Johansen's test is not valid under unconditionally heteroskedastic innovations, even asymptotically. Thus, in this section we focus just in the approach based on information criteria. Gauge and potency for the three DGPs, the three scenarios and the three sample sizes described above are included in table 2, we use the BIC for determining the cointegration rank. Again, results about false discovery are included in appendix G. As expected from §III, results do not differ significantly from the case of iid innovations. In moderately large samples ($T = 200$ and $T = 400$) we observe a very good performance both in terms of gauge and potency in all DGPs and scenarios. With $T = 100$ we observe an

TABLE 1:
Gauge and potency of the pairwise strategy in DGPs with iid innovations (nominal size of Johansen's tests $\varphi = 0.01$)

	$n_1 = 10$		$n_1 = 25$		$n_1 = 40$	
	Gauge	***Potency***	***Gauge***	***Potency***	***Gauge***	***Potency***
			DGP 1 - Johansen			
T=100	0.4	77.0	0.2	67.0	0.2	63.2
T=200	0.3	96.4	0.2	95.3	0.1	94.5
T=400	0.3	99.2	0.2	99.0	0.2	98.6
			DGP 1 - BIC			
T=100	2.2	95.0	0.7	84.0	1.7	81.3
T=200	0	100	0	98	0	100
T=400	0	100	0.67	100	0	100
			DGP 2 (non-diag. Φ) - Johansen			
T=100	2.2	67.5	1.1	60.2	0.4	26.0
T=200	0.2	90.6	0.2	84.2	0.1	80.5
T=400	0.2	98.7	0.2	98.2	0.1	96.9
			DGP 2 (non-diag. Φ) - BIC			
T=100	1.1	90.0	0.7	84.0	0.8	43.8
T=200	0.6	100.0	0.0	92.0	0.8	91.3
T=400	0.0	100.0	0.0	100.0	0.8	98.8
			DGP 3 (non-diag. Φ and complex long-run structure) - Johansen			
T=100	7.0	70.3	4.9	69.3	1.3	38.1
T=200	0.5	95.1	0.4	95.5	0.3	92.9
T=400	0.6	98.1	0.5	99.0	0.4	98.7
			DGP 3 (non-diag. Φ and complex long-run structure) - BIC			
T=100	2.8	90.0	1.3	92.0	0.8	50.0
T=200	0.0	100.0	0.0	94.0	0.0	92.5
T=400	0.0	100.0	0.0	100.0	0.0	98.8

- $Gauge = \frac{100}{(N-n_1)Nexp}\sum_{i=1}^{Nexp} Z_{2,i}$. - $Pot = \frac{100}{n_1 Nexp}\sum_{i=1}^{Nexp} Z_{1,i}$. - Z_2 = number of wrong series included in $\hat{n}_1$. - Z_1 = number of correct series included in $\hat{n}_1$. - $Nexp$ = number of experiments (500). - Scenario 1: $n_1 = 10$. - Scenario 2: $n_1 = 25$. - Scenario 3: $n_1 = 40$.

important deterioration of potency in scenario 3 in all DGPs, and also a relevant increase of gauge in scenarios 1 and 2 of DGPs 2 and 3.

Before concluding this section we recall that, as mentioned in the introduction, the pairwise approach can be extended for cases in which the set series under consideration has one general and several sectorial common trends. Interested readers can find the analysis of this case in appendix F.

TABLE 2:
Gauge and potency of the BIC-based pairwise strategy in DGPs with unconditionally heteroskedastic innovations

	$n_1 = 10$		$n_1 = 25$		$n_1 = 40$	
	Gauge	***Pot***	***Gauge***	***Pot***	***Gauge***	***Pot***
			DGP 1			
T=100	0.0	85.0	0.7	78.0	0.0	73.8
T=200	0.0	95.0	0.0	98.0	0.8	96.3
T=400	0.0	100.0	0.0	100.0	0.8	100.0
			DGP 2 (non-diagonal Φ *)*			
T=100	3.3	90.0	2.0	94.0	1.7	43.8
T=200	0.6	100.0	0.0	92.0	0.0	93.8
T=400	0.0	100.0	0.7	100.0	0.0	100.0
			DGP 3 (non-diagonal Φ*, complex long-run structure)*			
T=100	2.2	90.0	3.3	94.0	0.0	41.3
T=200	0.0	95.0	0.0	94.0	0.0	90.0
T=400	0.0	95.0	0.0	100.0	0.0	100.0

See notes to table 1.

V Empirical application: US CPI

In this section we apply the pairwise procedure to the US CPI. We focus just on the forecasting exercise, detailed results about the cointegration tests and the conformation of the fully cointegrated 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 usually will also end up with a better understanding of the aggregate. As a by-product of a disaggregated analysis we have that aggregating the forecasts of the components we get an indirect forecast of the aggregate. In arriving to the specification of fully-cointegrated 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 cointegrated 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 the direct forecast with this indirect one we could end up with an extra support for our disaggregated strategy.

Data

The CPI break down 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). The total number of components is 174. Not all the series have data for the whole sample period, after dropping those with less than 150 observations we keep 169 components. From these series we exclude nine that evolve by steps (regulated prices) so that we end up with 160 series which, considering 2016 weights, represent 92% of the CPI[6]. Among the remaining series, *Owners' equivalent rent of primary residence* weights approximately 24% of the CPI, which, comparing with other developed economies CPI or other aggregated variables in the US, is quite a rare situation.

In order to avoid the global results to be driven by the ability to forecast 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 fully cointegrated subsets, nor for the forecasting exercises. We denote the aggregate corresponding to the 159 series as CPI^*. An indirect forecast of the overall aggregate (the CPI) could be done from a regression of the CPI on the CPI^*, or by forecasting the excluded components individually and aggregating those forecasts.

Unconditional heteroskedasticity tests

We performed two types of unconditional heteroskedasticity tests; the White (1980) tests using only squares and using squares and cross products, and tests for a single shift in unconditional variance as those considered by Sensier and Van Dijk (2004). In both cases, we found almost 90% of the series to be homoskedastic (detailed results are available upon request).

This finding, led us to focus the application using Johansen's test. Anyway, as we comment below, the procedure based on information criteria led to the same results.

[6]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

Design of the forecasting exercise

In order to obtain economically and statistically sensible cointegration relationships between the components of the CPI^* we consider only those which satisfy the following four conditions: *(i)* the cointegration relationship does not require a deterministic trend, *(ii)* coefficients of both prices are statistically significant, *(iii)* the bivariate VAR characteristic polynomial's second largest root is not close to one, and *(iv)* the cointegration relationship is stable over time.

Cointegration tests are performed at the 1% of significance and the number of lags for each pair is determined with the AIC in a model with one cointegration restriction and without trend in the cointegration relationship. Centered seasonal dummies are included in all models. Subsets with less than five series were disregarded [7].

The presence of outliers may seriously distort cointegration tests results. For controlling this issue, we adopt a strategy that combines the GLS procedure of Saikkonen and Lütkepohl (2000) and Lütkepohl et al. (2004) for testing cointegration with the Impulse Indicator Saturation (IIS) technique for detecting outliers (see Santos et al., 2008, Johansen and Nielsen, 2009, and Castle et al., 2012)). Outliers detected with IIS are used in the pairwise GLS cointegration tests.

We found 7 subsets that jointly include 41 series, which represent 25.8% of the components and 23.5% of the total weight we are considering. When applying the procedure based on information criteria (BIC) to determine the cointegration rank, the same 7 groups are obtained.

For building the single-equation models we use the automatic model selection algorithm *Autometrics*, in which the initial big model to be reduced is called General Unrestricted Model (GUM), and includes IIS (see Doornik, 2009 and Castle et al., 2011).

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 from a scalar model of the aggregate, and with an indirect procedure in which all

[7] We performed previous seasonal unit root test 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 sensible.

components are forecast with univariate models. We denote our indirect approach by I-PW (the 'I' stands for indirect and 'PW' for pairwise), the direct one by D, and the univariate indirect by 'I-B' (indirect basic).

For these approaches (D, I-PW and I-B) we consider three broad possibilities depending on the regressors to be included in the GUM. Apart from own lags, seasonal dummies, outlier indicators, and cointegration relationships (when it is the case), we may include: 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 last option as *Dissaggregated information* (DI)

For each of the three possibilities (a to c), in the I-PW procedure series which do not belong to any fully cointegrated subset can be modeled individually or all together in a scalar model for the sub-aggregate that adds up all those series, we denote the latter option as I-PW-Rest. If forecasting the components which do not belong to any fully cointegrated subset is also 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.

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[8].

Forecasting comparison

Table 3 includes the results of a pseudo 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 computation of the fully cointegrated subsets, and the corresponding cointegration relationships, is carried out only each December. Hence, in PW approaches we are using less information than the truly

[8]In models D-DFM and I-B-DFM the q-dimensional factors (F) are computed from the first difference of all the components. 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 correction of large residuals.

available, except in January.

First row of table 3 includes the root mean squared forecast error (RMSFE) of $\Delta_{12} log(CPI^*)$ 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 4 includes p-values for the Diebold-Mariano tests for comparing the RMSFE of the I-PW and I-PW-Rest with selected competitors.

The main conclusions of the forecasting exercise are the following:

i. Table 3 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.

ii. 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.

iii. When modeling the disaggregates without including additional information beyond the own lags (I-B approach in row 4 of table 3) 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 (but it is to be noted that for large horizons the number of independent forecasts is small, 6 for $H = 12$).

iv. The I-PW approach (row 8) delivers better results than the baseline in all horizons. Reductions in RMSFE range from 10 to 12 percentage points and are statistically significant.

v. The comparison between I-PW and I-B indicates that the former beats the latter in horizons 1 to 10 and the gains range between 1 and 12 percentage points. As table 4 shows, these differences are statistically significant for horizons 1 to 7.

vi. The I-PW-Rest approach produces a small deterioration in the RMSFE with respect to I-PW for horizons 1 to 3, but differences are not statistically significant (see last line

of table 4). For horizons 4 to 12 improvements in RMSFE of the I-PW-Rest approach with respect to I-PW are remarkable. Starting with a RMSFE gain of 1 percentage point, the differences in favor of I-PW-Rest increase with the forecasting horizon and reach 22 percentage pints in horizon 12. As table 4 shows the differences are statistically significant for horizons 6 to 12.

vii. The forecasting gains of I-PW-Rest with respect to the baseline are even more remarkable. Staring from 7 percentage points in horizon 1, improvements in RMSFE systematically increase with the horizon to reach 33 points in H=12.

In summary, the pairwise approach performs better than all other alternatives considered in this paper, and it is the only approach that beats the baseline in all horizons. It gets bigger improvements in forecasting accuracy when using the restricted alternative (I-PW-Rest).

TABLE 3: 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**	0.89**	0.89**	0.90**	0.88**	0.88**	0.88**	0.89**	0.89**	0.88**	0.88**	0.89**	0.89**
(9)	**I-PW-CPI**	0.93	0.97	1.01	1.02	1.04	1.04	1.03	1.02	1.01	1.01	1.01	1.01
(10)	**I-PW-DI**	0.96	0.97	0.96	0.95	0.94	0.91	0.90**	0.87**	0.86**	0.87**	0.86**	0.86**
(11)	**I-PW-Rest**	0.93	0.91	0.91	0.87	0.83*	0.78**	0.76**	0.75**	0.72**	0.69**	0.67**	0.67**
(12)	**I-PW-Rest-CPI**	0.94	0.98	1.01	0.99	0.96	0.90	0.87*	0.84**	0.81**	0.78**	0.76**	0.75**
(13)	**I-PW-Rest-DI**	0.99	0.99	1.00	0.96	0.91	0.83*	0.79**	0.76**	0.71**	0.70**	0.66**	0.65**

* Significantly different from the baseline at the 5% level using the Diebold-Mariano test.
** Significantly different from the baseline at the 1% level using the Diebold-Mariano test.

TABLE 4: 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	0.00	0.01	0.02	0.03	0.04	0.04	0.07	0.19	0.26	0.40	0.49	0.38
2 vs 8	0.05	0.14	0.30	0.21	0.18	0.28	0.30	0.27	0.16	0.11	0.07	0.07
6 vs 8	0.03	0.03	0.06	0.06	0.06	0.20	0.47	0.20	0.16	0.14	0.05	0.01
2 vs 11	0.18	0.27	0.40	0.25	0.14	0.09	0.05	0.03	0.01	0.00	0.00	0.00
6 vs 11	0.17	0.13	0.23	0.17	0.09	0.04	0.04	0.06	0.04	0.02	0.02	0.03
8 vs 11	0.18	0.40	0.45	0.44	0.27	0.08	0.04	0.02	0.01	0.00	0.00	0.00

- Row names refer to the lines of table 3
- The null hypothesis is that the RMSFE are equal.

VI Concluding Remarks

In this paper we studied the properties of a pairwise procedure for testing cointegration in all possible pairs between the components of an aggregate at the maximum level of disaggregation. This procedure allows to discover subsets of series that share a unique common trend (*fully cointegrated subsets*).

As our methodology does not rely on any cross-sectional averaging procedure, we need neither to assume pervasiveness of the common trends, nor to impose special restrictions on serial or cross-correlation of idiosyncratic components. Furthermore, we do not need the cross-sectional dimension to go to infinity.

We showed that the main theoretical results are valid when determining the cointegration rank of the pairs by means of the Johansen's approach and when using procedures based on information criteria, which are robust to a wide type of conditional and unconditional heterosckedasicity in the innovations of the models. We also argued that the pairwise approach can be extended for sets of macro variables (not necessarily components of a single one) with *general* and *sectorial* trends.

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 fully cointegrated 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. Our results suggest that the pairwise approach shows a promising way for choosing useful disaggregations. Therefore, if disaggregated information exists, it is not efficient to ignore it, and looking for single-trend subsets of components is a useful and feasible strategy to exploit that information.

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.

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 (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*.

Beck, G., K. Hubrich, and M. Marcellino (2015). On the importance of sectoral and regional shocks for price-setting. *Journal of Applied Econometrics , forthcomming*.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169–194.

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

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.

Cavaliere, G., L. De Angelis, and L. Fanelli (2018). Co-integration rank determination in partial systems using information criteria. *Oxford Bulletin of Economics and Statistics 80*(1), 65–89.

Cavaliere, G., L. De Angelis, A. Rahbek, and A. R. Taylor (2016). Determining the cointegration rank in heteroskedastic var models of unknown order. *Econometric theory*, 1–34.

Cavaliere, G., A. Rahbek, and A. R. Taylor (2010). Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics 158*(1), 7–24.

Clements, M. P. and D. F. Hendry (1995). Forecasting in cointegrated systems. *Journal of Applied Econometrics 10*(2), 127–146.

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.

Doornik, J., D. Hendry, and B. Nielsen (1998). Inference in cointegrating models: Uk m1 revisited. *Journal of economic surveys 12*(5), 533–572.

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.

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.

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. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*, Volume 9. Cambridge University Press.

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. (2006). *The cointegrated VAR model: methodology and applications*. Oxford University Press, USA.

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.

Kim, C.-J. and C. R. Nelson (1999). Has the us economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. *The Review of Economics and Statistics 81*(4), 608–616.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2003). Comparison of tests for the cointegrating rank of a var process with a structural shift. *Journal of Econometrics 113*(2), 201–229.

Lütkepohl, H., P. Saikkonen, and C. Trenkler (2004). Testing for the cointegrating rank of a var process with level shift at unknown time. *Econometrica 72*(2), 647–662.

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)).

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review 90*(5), 1464–1476.

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

Nielsen, B. and A. Rahbek (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics 62*(1), 5–22.

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.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics 22*(2), 129–162.

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

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.

Saikkonen, P. and H. Lütkepohl (2000). Testing for the cointegrating rank of a var process with structural shifts. *Journal of Business & Economic Statistics 18*(4), 451–464.

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

Sensier, M. and D. Van Dijk (2004). Testing for volatility changes in us macroeconomic time series. *Review of Economics and Statistics 86*(3), 833–839.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

Zivot, E. and J. Wang (2007). *Modeling financial time series with S-Plus*, Volume 191. Springer Science & Business Media.

Appendix A FOR ONLINE PUBLICATION ONLY: Proof of theorem 1

For simplicity, and without loss of generality, no deterministic terms are considered in this proof. From the Granger representation theorem we can write any individual series, say X_a, in n_1 as:

$$X_{at} = \delta_a CT_t + w_{at}; \quad a = 1, ..., n_1, \tag{A.1}$$

where CT_t is a random walk (the common stochastic trend) and w_{at} is a stationary process.

Defining the vector X_t as $[X_{at}, X_{bt}]'$, for $[a, b] = 1, .., n_1$ and $a \neq b$, the Johansen's procedure consists of, firstly, concentrating out the model with respect to $\alpha\beta'$, which is done by regressing ΔX_t and X_{t-1} on $(\Delta X_{t-1}, ..., \Delta X_{t-k+1})$, to obtain the residuals R_{0t} and R_{1t}, respectively. Define R_i as the $2 \times T$ matrix $[R_{i1}, ..., R_{iT}]'$, for $i = 0, 1$. Then, define matrices S_{ij} as $T^{-1}R_iR_j'$. The likelihood ratio test for the null $r = 1$ vs. $r = 2$ is: $-Tln(1 - \hat{\lambda}_2)$, where $\hat{\lambda}_2$ is the smallest eigenvalue of the generalized eigenvalue problem:

$$(S_{10}S_{00}^{-1}S_{01})v = \lambda S_{11}v, \tag{A.2}$$

whose eigenvalues are the solution of, $|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$.

Assume that matrices S_{ij} refer to the vector , $X_t = [X_{1t}, X_{2t}]'$. We now derive the test statistic for any other pair in n_1 given that for X_t. From eq. (A.1), any series in n_1 can be expressed as linear combination of any other series in n_1 and a stationary component. Thus, we can write:

$$\begin{aligned} X_{at} &= \gamma_{a1}X_{1t} + \eta_{a1,t}; \\ X_{bt} &= \gamma_{b2}X_{2t} + \eta_{b2,t}; \end{aligned} \tag{A.3}$$

call $X_t^* = [X_{at}, X_{bt}]'$ and assume, without loss of generality, that $p^* \geq p$ (where p^* is the lag length of the model for X_t^*, and p is the lag length of the model for X_t). Writing the auxiliary regressions for ΔX_t^* and X_{t-1}^* to obtain R_{0t}^* and R_{1t}^*, and using eq. (A.3), it can be shown that the new (2×2) matrices S_{ij}^* are:

$$S_{ij}^* = T^{-1}(\Gamma R_i + e_i)(\Gamma R_j + e_j)', \ for \ [i, j] \in [1, 2], \tag{A.4}$$

where $\Gamma = \begin{bmatrix} \gamma_{a1} & 0 \\ 0 & \gamma_{b1} \end{bmatrix}$, and e_i and e_j are stationary processes. Then,

$$S_{11}^* = T^{-1}[\Gamma R_1 R_1'\Gamma + \Gamma R_1 e_1' + e_1 R_1'\Gamma + e_1 e_1'], \tag{A.5}$$

where we used $\Gamma = \Gamma'$.

In eq. (A.5), all terms inside the brackets are $Op(T)$, except for $\Gamma R_1 R_1'\Gamma$, which is $Op(T^2)$[9].

[9]Since ΓR_1 is $I(1)$ with zero mean, its variance is $Op(T)$ and $\Gamma R_1 R_1'\Gamma$ is $Op(T^2)$. Since e_1 is $I(0)$ with

Thus, S_{11}^* is $Op(T)$ and its long-run behavior is dominated by $\Gamma S_{11}\Gamma$. That is, $S_{11}^* \simeq \Gamma S_{11}\Gamma$ for large T.

The the generic expression for S_{ij}^* (which is $Op(1)$ except for S_{11}^*) is:

$$S_{ij}^* = \Gamma S_{ij}\Gamma + \Omega_{ij}, \quad i,j \in [1,2], \tag{A.6}$$

where $\Omega_{ij} = T^{-1}[\Gamma R_i e_j' + e_i R_j'\Gamma + e_i e_j']$ is $Op(1)$, for $i,j \in [1,2]$.

The new eigenvalue problem is: $(S_{10}^* S_{00}^{*-1} S_{01}^*)v^* = \lambda^* S_{11}^* v^*$. Using eq. (A.6), we get:

$$[(\Gamma S_{10}\Gamma + \Omega_{10})(\Gamma S_{00}\Gamma + \Omega_{00})^{-1}(\Gamma S_{01}\Gamma + \Omega_{01})]v^* = \lambda^*(\Gamma S_{11}\Gamma + \Omega_{11})v^*. \tag{A.7}$$

Note that $(\Gamma S_{00}\Gamma + \Omega_{00})^{-1}$ can be written as:

$$(\Gamma S_{00}\Gamma + \Omega_{00})^{-1} = c(\Gamma S_{00}\Gamma)^{-1} + \tilde{\Omega}_{00} = c\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \tilde{\Omega}_{00}, \tag{A.8}$$

where the equality $(\Gamma S_{00}\Gamma)^{-1} = \Gamma^{-1}S_{00}^{-1}\Gamma^{-1}$ follows from the fact that Γ is diagonal, $c = \dfrac{|\Gamma S_{00}\Gamma|}{|\Gamma S_{00}\Gamma + \Omega_{00}|}$, and $\tilde{\Omega}_{00} = \dfrac{Adj[\Gamma S_{00}\Gamma + \Omega_{00}] - Adj[\Gamma S_{00}\Gamma]}{|\Gamma S_{00}\Gamma + \Omega_{00}|}$ (note that $0 < c \leq 1$).

Hence, plugging eq. (A.8) into eq. (A.7) and doing some algebra, we get:

$$[c\Gamma S_{10}S_{00}^{-1}S_{01}\Gamma + \Psi]v^* = \lambda^*(\Gamma S_{11}\Gamma + \Omega_{11})v^*, \tag{A.9}$$

where, $\Psi = (\Gamma S_{10}\Gamma\tilde{\Omega}_{00} + c\Omega_{10}\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \Omega_{10}\tilde{\Omega}_{00})(\Gamma S_{01}\Gamma + \Omega_{01})$, is $Op(1)$.

Left multiplying eq. (A.9) by Γ^{-1} we obtain: $[cS_{10}S_{00}^{-1}S_{01}\Gamma + \Gamma^{-1}\Psi]v^* = \lambda^*(S_{11}\Gamma + \Gamma^{-1}\Omega_{11})v^*$. Now, let $\underset{\sim}{\Psi} = \Gamma^{-1}\Psi\Gamma^{-1}$, and $\underset{\sim}{\Omega}_{11} = \Gamma^{-1}\Omega_{11}\Gamma^{-1}$, to get:

$$[cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi}]\Gamma v^* = \lambda^*(S_{11} + \underset{\sim}{\Omega}_{11})\Gamma v^*. \tag{A.10}$$

Comparing eq. (A.10) with eq. (A.2), we can make three considerations:

(i) If $X_a \equiv X_1$ and $X_b \equiv X_2$, we get $\underset{\sim}{\Psi} = 0$, $\underset{\sim}{\Omega}_{11} = 0$, $c = 1$, and $\Gamma = I$, so we recover

zero mean, its covariance with ΓR_1 is $Op(1)$ and $\Gamma R_1 e_1'$ is $Op(T)$. Similarly, the variance of e_1 is $Op(1)$ and $e_1 e_1'$ is $Op(T)$.

the original problem. (ii) In the unlikely case that $\Omega_{ij} = 0$, we also get $\underset{\sim}{\Psi} = 0$ and $c = 1$, so that the eigenvalue problem would be: $[S_{10}S_{00}^{-1}S_{01}\Gamma]v^* = \lambda^*(S_{11}\Gamma)v^*$, the solution of which is $\lambda^* = \lambda$ and $v^* = \Gamma v$. Hence, even in small samples, the cointegration test statistic would be exactly the same as that for the pair (X_1, X_2). (iii) In the general case that $\Omega_{ij} \neq 0$, we will have $\underset{\sim}{\Psi} \neq 0$, $\underset{\sim}{\Omega}_{11} \neq 0$, and $c \neq 1$. Note that the eigenvalues of the problem eq. (A.10) are the solutions of the second-order polynomial in λ^* $|\lambda^*(S_{11} + \underset{\sim}{\Omega}_{11}) - (cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi})| = 0$.

Focus on the general case that $\Omega_{ij} \neq 0$. As Johansen (1995) shows, the test statistic $-T\sum_{r+1}^{p} ln(1-\lambda_i)$ converges to a non-standard distribution that does not depend on S_{00}. Given that S_{11} is $Op(T)$ and the other matrices are $Op(1)$, the asymptotic behavior of λ and λ^* is dominated by the same terms. To see this, let $\Theta = S_{10}S_{00}^{-1}S_{01}$, and $\Theta^* = cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi}$. The original eigenvalues λ_1 and λ_2 ($\lambda_1 > \lambda_2$) are the roots of the polynomial:

$$\lambda^2|S_{11}| + \lambda \underbrace{(s_{12}\theta_{21} + s_{21}\theta_{12} - s_{11}\theta_{22} - s_{22}\theta_{11})}_{B} + \underbrace{(\theta_{11}\theta_{22} - \theta_{21}\theta_{12})}_{C} = 0, \tag{A.11}$$

where s_{ij} and θ_{ij} are the elements of the matrices S_{11} and Θ, respectively.

Since $B < 0$, $\lambda_2 = \dfrac{-B - \sqrt{B^2 - 4|S_{11}|C}}{2|S_{11}|} = \dfrac{G}{2|S_{11}|}$.

If the series are cointegrated $|S_{11}| \sim Op(T)$, and since $B \sim Op(T)$, the expression under the square root is dominated by B^2, meaning that $(B^2 - 4|S_{11}|C) \simeq B^2$. Thus $G \to 0$.

Now, replace θ_{ij} by θ_{ij}^*, and s_{ij} by s_{ij}^* in eq. (A.11) to get B^*, C^* and G^*. Since $S_{11}^* = S_{11} + Op(1)$, under cointegration, $|S_{11}^*|$ is still $Op(T)$. Additionally, since θ_{ij} and θ_{ij}^* are $Op(1)$, the asymptotic behavior of G^* is the same as that of G, as the expression under the square root is also dominated by B^{*2}, which is determined by the same s_{ij}'s as B. ■

Appendix B FOR ONLINE PUBLICATION ONLY: Proof of proposition 1

When n_j, the size of the fully cointegrated subset, is fixed, the proof follows directly from the proof of Theorem 1, so focus on the case of $n_j \to \infty$.

Choose a pair from the n_j series, say *pair* 0, and define Z_i as the random variable that

takes the value 1 if the test for the i^{th} pair delivers the same cointegration rank as *pair* 0, and takes the value 0 otherwise. The expected proportion of tests that deliver the same cointegration rank as pair 0 is $n_j^* E(Z_i)/n_j^*$, where n_j^* is the number of pairs between the n_j series. From Theorem 1 $E(Z_i) = 1$, thus, the mentioned expected proportion goes to 1 as T goes to infinity. ■

Appendix C FOR ONLINE PUBLICATION ONLY: Proof and discussion of proposition 2

Under Assumption 2, the true number of cointegration relationships between one series inside and one outside a fully cointegrated subset is $r = 0$. Thus, for wrongly including a series in $\hat{n}_j$ we require that the cointegration tests for all of the pairs between that series and every series in $\hat{n}_j$ reject $r = 0$.

Let X_{out} be a series outside n_j and let WR_i be the event of wrongly rejecting $r = 0$ between X_{out} and the i^{th} series in $\hat{n}_j$. Since for wrongly including X_{out} in $\hat{n}_j$, we need to wrongly reject $\hat{n}_j$ hypotheses, the probability of inclusion is $P(WR_1 \cap ...WR_{\hat{n}_j})$. This probability can be factorized as:

$$P(WR_1 \cap ...WR_{\hat{n}_j}) = P(WR_1|WR_2, ..., WR_{\hat{n}_j}) \times ... \times P(WR_{\hat{n}_j - 1}|WR_{\hat{n}_j}) \times P(WR_{\hat{n}_j}), \quad \text{(C.1)}$$

where $P(WR_i)$ is the nominal size of the pairwise tests (φ). Using the extreme assumption that all of the $\hat{n}_1 - 1$ conditional probabilities in eq. (C.1) are equal to 1, the probability of wrongly including X_{out} in $\hat{n}_1$ would be φ and, the expected number of wrong series in $\hat{n}_j$ would be $(N - n_j)\varphi$. Thus, under this extreme assumption, the expected ratio of wrong elements in the estimated n_j over its true size would be:

$$E[W] = \frac{(N - n_j)\varphi}{n_j}. \quad \text{(C.2)}$$

Therefore, with fixed N, we may have $N >> n_j$ and the procedure will still work properly.

For instance, with $N - n_j = 100$ and $\varphi = 0.01$, we will have just one wrong series in $\hat{n}_j$.

If we consider $N \to \infty$, and φ is fixed, we would need n_j to grow at the same rate as N to avoid the proportion of wrong elements going to infinity. This implies a *pervasiveness* requirement similar to that of DFM (see, e.g., Assumption B in Bai, 2003). There are two possible ways to avoid imposing the pervasiveness assumption.

One way is to set φ as an inverse function of N. We could set $\varphi = N^{-1/\kappa}$, with κ being a fixed constant strictly larger than zero. In this way, n_j must grow at a rate larger than or equal to that of $N^{1-1/\kappa}$, which is smaller than that of N. Note however that the use of such a significance level may generate power issues for the pairs in which $r = 1$. For those pairs, the test statistic for the null of $r = 0$ diverges at the rate of T, we need $N^{1/\kappa}$ to be $O_p(T)$. This is ensured by assumption 5. Since κ can be larger than one, this assumption does not impose any restriction on the asymptotic behavior of the ratio T/N.

Note that in with $\kappa < 1$, we may have $n_j \to 0$ and our procedure will still work properly in terms of gauge. This is possible because, by assumption 5, $\kappa < 1$ implies $T/N \to \infty$.

Another way to avoid the need of pervasiveness is to relax he extreme assumption that all conditional probabilities in eq. (C.1) are equal to one. Once this is done we can use a fixed φ and still have non-pervasive common trends. A detailed discussion about this issue is available upon request.

Appendix D FOR ONLINE PUBLICATION ONLY: Proof of proposition 3

In this appendix we show that the expected number false fully cointegrated subsets of size λN(subsets of λN $I(1)$ series in which there are none cointegrated pairs but pairwise cointegration tests indicate that all pairs are cointegrated) tends to zero as T and N go to ∞.

The degree of dependence between the distributions of the cointegration tests statistics for pairs of series in a set in which none of them cointegrates, is unknown. In what follows

we consider the two extreme cases.

Independent distributions

In this case, the expected number of (wrongly) estimated fully cointegrated subsets composed by Q series of the $N - \sum_j n_j \equiv n$ is

$$E[Q_{worng}] = C_Q^n \varphi^{Q(Q-1)/2}, \tag{D.1}$$

where φ is the nominal size of the individual tests and C_K^n is the binomial coefficient. This expectation is almost zero for, say, $\varphi = 0.01$, $Q > 4$ and moderately large n.

When n goes to infinity, letting $Q = \lambda n$ (with $0 < \lambda < 1$), we show that $E[Q_{worng}]$ goes to zero at a rate that depends on whether φ is fixed or not.

Using, $log(C_Q^n) \approx nlog(n) - (n-Q)log(n-Q) - Qlog(Q)$, and taking logs in eq. (D.1) we get:

$$log(E[Q_{worng}]) \approx nlog(n) - (n-Q)log(n-Q) - Qlog(Q) + \frac{Q(Q-1)}{2}log(\varphi) \tag{D.2}$$

Using $Q = \lambda n$ and plugging into eq. (D.2):

$$log(E[Q_{worng}]) \approx nlog(n) - n(1-\lambda)log(n(1-\lambda)) - \lambda nlog(n\lambda) - \frac{\lambda n(\lambda n - 1)}{2}log(\varphi)$$

Reordering:

$$log(E[K_{worng}]) \approx \frac{\lambda n(\lambda n - 1)}{2}log(\varphi) - n(1-\lambda)log(1-\lambda) - \lambda nlog(\lambda) \tag{D.3}$$

Thus $E[Q_{worng}]$ converges to zero. If φ is fixed, the rate of convergence is that of $\varphi^{\lambda n(\lambda n-1)/2}$. If we use $\varphi = n^{-1/\kappa}$, the rate of convergence is that of $n^{\lambda n(1-\lambda n)/2\kappa}$.

Maximal dependence

In a set of Q series there are $Q-1$ independent pairs, meaning that the minimum number of independent tests is $Q-1$. Thus, in the worst possible situation, the expected number of (wrongly) estimated fully cointegrated subsets composed by Q series of the $N - \sum_j n_j \equiv n$

is $E[Q_{worng}] = C_Q^n \varphi^{(Q-1)}$, where φ is the nominal size of the individual tests and C_Q^n is the binomial coefficient.

Note that this is an extreme and very unlikely situation because it means that in a triplet of not cointegrated $I(1)$ series, finding cointegration in two of the three pairs implies that cointegration will be found in the third pair with probability one. Simulation results in §IV show that this is far from reality.

Still, if this were the case, $E[Q_{worng}]$ is acceptably small for, say, $\varphi = 0.01$, $Q > 4$ and moderate n (for the these values and $n < 106$ the expectation is smaller than 1).

When n goes to infinity, letting $\varphi = n^{-1/\kappa}$ and $Q = \lambda n$, and using the same reasoning as in the case of independence, we get that $E[Q_{worng}]$ goes to zero at the rate of $n^{(1-\lambda n)/\kappa}$.

Implications

Previous results imply that choosing φ as an inverse function of n and disregarding small fully cointegrated subsets (those with cardinality smaller than or equal to λn) constitutes a strong protection against false discovery. In practice n is unknown but we can, safely, use $\varphi = N^{-1/\kappa}$ and disregard subsets smaller than or equal to λN because $N > n$.

Appendix E FOR ONLINE PUBLICATION ONLY: Proof of proposition 5

In this proof we use a slight modification of the arguments made in §II to study the gauge of the pairwise procedure applied with Johansen's tests.

For wrongly including a series in $\hat{n}_j$ we require that the estimation of the cointegration rank for all of the pairs between that series and every series in $\hat{n}_j$ deliver $r = 1$.

Let X_{out} be a series outside n_j and let WR_i be the event of wrongly concluding $r = 1$ between X_{out} and the i^{th} series in $\hat{n}_j$. Since for wrongly including X_{out} in $\hat{n}_j$, we need to $\hat{n}_j$ wrong rank conclusions, the probability of inclusion is $P(WR_1 \cap ...WR_{\hat{n}_j})$. This probability can be factorized as:

$$P(WR_1 \cap ...WR_{\hat{n}_j}) = P(WR_1|WR_2, ..., WR_{\hat{n}_j}) \times ... \times \ P(WR_{\hat{n}_j-1}|WR_{\hat{n}_j}) \times \rho, \qquad \text{(E.1)}$$

where ρ is the marginal probability of wrongly concluding $r = 1$. Using the extreme assumption that all of the $\hat{n}_1 - 1$ conditional probabilities in eq. (E.1) are equal to 1, the probability of wrongly including X_{out} in $\hat{n}_1$ would be ρ and, the expected number of wrong series in $\hat{n}_j$ would be $(N - n_j)\rho$. Thus, under this extreme assumption, the expected ratio of wrong elements in the estimated n_j over its true size would be:

$$E[W] = \frac{(N - n_j)\rho}{n_j}. \tag{E.2}$$

By theorem 1 of Cavaliere et al. (2016) $\rho \to 0$ as $T \to \infty$, therefore, with fixed N, we may have $N >> n_j$ and, not only the expected proportion, but also the expected number of wrong inclusions will tend to zero.

Consider now the case of $N \to \infty$. If the common trends are pervasive (n_j grows at the same rate as N), then, we still have the $E[W]$ going to zero.

In the case of non-pervasive trends ($N/n_j \to \infty$), we need to impose a restriction on the asymptotic behavior of N/n_j as $[T, N] \to \infty$.

From lemma 1 and theorem 1 of Cavaliere et al. (2016) $\rho = P(A_T > 0)$, where $A_T \to \infty$ and $A_T/T \to 0$ (see Cavaliere et al. (2016) for the exact expression of A_T). In the case of the Bayesian Information Criterion (BIC), A_T diverges at the rate of $log(T)$, and in the case of the Hannan and Quinn Information Criterion (HQC) it diverges at the rate of $log(log(T))$. Thus, in the former case we need N/n_j to be $Op(log(T))$ and in the latter $Op(log(log(T)))$. These restrictions ensure $E[W]$ going to zero as $[T, N] \to \infty$.

Appendix F FOR ONLINE PUBLICATION ONLY: Extension to sets of series with one 'general' and several 'sectorial' common trends

Throughout the paper we focused on the specific case that the data set at hand contains several trends, among which some are common to reduced groups of series such that each of

those groups have only one common trend. As Espasa and Mayo-Burgos (2013) argue, this is a sensible assumption 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 DFM.

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

If this is the situation, the pairwise procedure proposed in this paper will not be useful. Since the only cointegrated pairs are those formed by series with a single common trend (e.g., series that have only the general factor and no sectorial one), the procedure will be unable to discover the 'true' data structure. Thus, we need to modify our approach for this situation.

Provided that in the set of N series there is a subset of series that have just the *general* trend, we can apply the following algorithm:

(i) Apply the pairwise procedure proposed in this paper. This will lead us to discover the subset of series that have only the *general* trend — call it n_1. (ii) Test for cointegration in all of the possible triplets formed by one series inside $\hat{n}_1$ and a pair of outsiders. For the triplets in which the outsiders have the same *sectorial* trend, we will find one cointegration relationship (two common trends). (iii) Construct a $(N-\hat{n}_1)\times(N-\hat{n}_1)$ symmetric *adjacency matrix* for the series outside $\hat{n}_1$ such that each cell of this matrix represents a pair of the components outside $\hat{n}_1$. Each of those pairs belongs to $\hat{n}_1$ different triplets: one for each element of $\hat{n}_1$. Then, in each cell of the adjacency matrix, put a 1 if all of the corresponding $\hat{n}_1$ triplets have just one cointegration relationship; otherwise, put a 0. (iv) Look for maximal fully connected sub-graphs in the previous *adjacency matrix*. This will lead us to discover the series in each sector.

Remark 7 By Theorem 1, in point iii above, it would be asymptotically irrelevant if in testing cointegration in a given triplet formed by a pair outside $\hat{n}_1$ and an element inside $\hat{n}_1$

we do it (a) with all of the series in $\hat{n}_1$, (b) with some of them, or (c) with the estimated common trend of $\hat{n}_1$. When dealing with small samples, requiring to find one cointegration relationship in all of the $\hat{n}_1$ triplets that contain the same pair of series outside $\hat{n}_1$ and one series of that subset (case a) may be too stringent. Instead, we could relax this requirement by allowing *a few* of those triplets to fail in showing the existence of one cointegration relationship. In a simulation experiment (available online), we consider three possibilities: requiring a cointegration relationship in all of the $\hat{n}_1$ mentioned triplets; in all but one of those triplets; and in all but two of them.

This procedure contributes to the literature in one relevant aspect: while the usual practice is to assume the sectorial structure as given, we can estimate it. Ando and Bai (2015) estimate the sectorial structure but for stationary variables, with a size of sectors that goes to infinity (in their simulation experiments the smallest sector has 100 units) and restricted serial and cross-correlation of the error terms. The Global VAR models proposed by Pesaran et al. (2004) are also related to our proposal. Among other relevant differences, we determine the 'regions' (sectors) statistically and do not have restrictions on the number of variables per region, which can be large.

Simulation design

As argued above, the generalization for the case of *general* and *sectorial* trends requires testing cointegration not only in pairs, but also in some triplets of series. Thus, the computational cost somehow rises with respect to the *pure* pairwise approach. Assume a case with $N = 100$ and $n_1 = 10$ (now, n_1 is the subset of series that have only the *general* trend). Assume also that $\hat{n}_1 = 10$. After testing cointegration in all of the 4950 pairs, the procedure requires making other $10 \times 85(85-1)/2 = 35700$ cointegration tests. However, as highlighted in Remark 7, this issue could be strongly mitigated by testing cointegration only with the estimated common trend of $\hat{n}_1$, so that the additional tests in the previous example would be only 3570. We do not explore this possibility.

In this section, we consider two DGPs. The DGPs are DGP 1 and DGP 3 described in §IV, modified to have *general* and *sectorial* trends. Call these modified DGPs as DGP 4 and DGP 5, respectively. Thus, DGP 4 represents a process in which each variable in n_1 or in

some sector reacts only to one cointegration relationship, and idiosyncratic components are independent. On the other hand, in DGP 5 variables in n_1 or in some sector react to more than one cointegration relationship, and there is idiosyncratic cross-correlation between all of the N variables. The rest of this sub-section is devoted to describe the DGPs in more detail.

Let s_i be the number of variables that, in addition to the general trend, also have the trend of sector i. Using the same normalization for matrix β as in DGPs of previous section, without loss of generality, we normalize all cointegration relationships with respect to one of the variables in n_1. To have a simple example of $\beta's$ structure, assume that $N = 10$, $n_1 = 3$, $s_1 = 3$, $s_2 = 3$, and that the remaining series has its own trend. In this case, we can set β such that:

$$\beta' = \begin{pmatrix} \beta_{11} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{21} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{31} & 0 & 0 & \beta_{34} & 1 & 0 & 0 & 0 & 0 & 0 \\ \beta_{41} & 0 & 0 & \beta_{44} & 0 & 1 & 0 & 0 & 0 & 0 \\ \beta_{51} & 0 & 0 & 0 & 0 & 0 & \beta_{57} & 1 & 0 & 0 \\ \beta_{61} & 0 & 0 & 0 & 0 & 0 & \beta_{67} & 0 & 1 & 0 \end{pmatrix} \tag{F.1}$$

An important difference with respect to §IV is that we cannot set the coefficients β_{ij} equal to -1 because the series in n_1 would be cointegrated with all of the other series in the system. To avoid this, we need some variation in the coefficients β_{ij}. Thus, we take those coefficients from the uniform distribution with parameters $[-5, -0.1]$. For DGP 4 (DGP 5), matrix Φ_1 is generated in the same way as in DGP 1 (DGP 3).

For DGP 4, matrix α has exactly the same structure as in DGP 1 (see eq. (4)), except that the number of columns (r) is now $n_1 + s_1 + s_2 - 3$. With this structure, the series in first position of n_1, s_1 and s_2 are weakly exogenous. The other series, react to a single cointegration relationship that affects itself, the series in the first position of n_1, and, for the series that belong to some sector, the first series of the sector (see matrix β in eq. (F.1)).

In DGP 5, we set matrix α such that each variable j belonging to n_1 or to some sector

reacts to $q_j + 1$ cointegration relationships and there are no weakly exogenous variables. To get a visual example, assume that apart from the general common trend, there are two sectorial ones. In this case, matrix α can be partitioned as follows:

$$\alpha = \begin{bmatrix} A1_{n_1 \times (n_1 - 1)} & 0 & 0 \\ 0 & A2_{s_1 \times (s_1 - 1)} & 0 \\ 0 & 0 & A3_{s_2 \times (s_2 - 1)} \\ 0 & 0 & 0_{(N-(n_1+s_1+s_2)) \times (s_2 - 1)} \end{bmatrix}, \tag{F.2}$$

where $A1$, $A2$ and $A3$ have the same structure as matrix α of DGP 3 (see eq. (6)).

We consider four scenarios and one sample size, $T = 400$. In the four scenarios, there is a single general trend, two sectors, and some series with their own trends. In *scenario 1* we set $N = 35$, $n_1 = 10$, $s_1 = 10$, $s_2 = 10$, and the remaining five series have their own trends. In *scenario 2* we add more noise; instead of only five series with their own trends, we have 30, thus, in this second scenario $N = 60$. In *scenario 3* $N = 80$, $n_1 = 25$, $s_1 = 25$, $s_2 = 25$, and the remaining five series have their own trends. In *scenario 4* $N = 105$, $n_1 = 25$, $s_1 = 25$, $s_2 = 25$, and the remaining 30 series have their own trends.

Note that in scenarios 2 and 4, not even the 'general' trend is clearly pervasive.

We use scenarios 1 and 2 both for DGP 4 and DGP 5. For saving computing time, we simulate scenarios 3 and 4 only for DGP 5.

Results

Table F.1 includes the *gauge* and *potency* of the algorithm described above for discovering the *general* and the *sectorial* trends. Figures under 'Sectors' columns are averages for the two sectors. As the table shows, in general, the procedure has high potency for discovering the true series in each sector with little cost in terms of gauge.

In DGP 4 ('simple' matrix α and diagonal Φ_1), potency for n_1 is close to 99% in both scenarios, and gauge is 1%. For the sectors, when we require a cointegration relationship in *all* of the triplets formed by a pair of series outside $\hat{n}_1$ and each of the insiders, potency is somewhat lower, but still high (92%). When we allow some of those triplets to fail in showing a cointegration relationship (see remark 7), potency figures of the sectors get close

to those of n_1. This improvement in potency is costless in terms of gauge.

In DGP 5 ('complex' matrix α and non-diagonal Φ_1), potency for n_1 is almost the same as in DGP 4. Gauge is somewhat larger, but we still have acceptable low figures. Potency results for the sectors show a relevant deterioration that is mitigated by allowing some failures in the cointegration tests of the triplets. This improvement in potency is costless in terms of gauge, which is somewhat larger than in DGP 4 but is still acceptably low. Note that a gauge of 0.05 scenarios 1 and 2 implies an average of 0.5 wrong series in the estimated subsets. For scenarios 3 and 4 the same gauge implies an average of 1.25 wrong series.

TABLE F.1:
Gauge and potency of the 'pairwise' procedure for the case of *general* and *sectorial* trends (nominal size $\varphi = 0.01$, $T = 400$)

		Scenario 1		**Scenario 2**		**Scenario 3**		**Scenario 4**	
		n_1	*Sectors*	n_1	*Sectors*	n_1	*Sectors*	n_1	*Sectors*
DGP 4: All	*Potency*	0.98	0.92	0.99	0.92	—	—	—	—
	Gauge	0.01	0.02	0.01	0.01	—	—	—	—
DGP 4: All but one	*Potency*	0.98	0.95	0.99	0.96	—	—	—	—
	Gauge	0.01	0.01	0.01	0.01	—	—	—	—
DGP 4: All but 2	*Potency*	0.98	0.95	0.99	0.96	—	—	—	—
	Gauge	0.01	0.01	0.01	0.01	—	—	—	—
DGP 5: All	*Potency*	0.97	0.82	0.96	0.73	0.96	0.68	0.96	0.60
	Gauge	0.05	0.03	0.03	0.03	0.04	0.01	0.04	0.01
DGP 5: All but one	*Potency*	0.97	0.88	0.96	0.84	0.96	0.82	0.96	0.76
	Gauge	0.05	0.03	0.03	0.02	0.04	0.01	0.04	0.01
DGP 5: All but two	*Potency*	0.97	0.89	0.96	0.86	0.96	0.87	0.96	0.83
	Gauge	0.05	0.03	0.03	0.02	0.04	0.01	0.04	0.01

- $Gauge = \frac{100}{(N-n_1)Nexp}\sum_{i=1}^{Nexp} Z_{2,i}$. - $Pot = \frac{100}{n_1 Nexp}\sum_{i=1}^{Nexp} Z_{1,i}$. - Z_2 = number of wrong series included in $\hat{n}_1$. - Z_1 = number of correct series included in $\hat{n}_1$. - $Nexp$ = number of experiments (500). - n_1 is the group of series that have the general trend only. - *Scenario 1:* $N = 35$, $n_1 = 10$, $s_1 = s_2 = 10$. - *Scenario 2:* $N = 60$, $n_1 = 10$, $s_1 = s_2 = 10$. - *Scenario 3:* $N = 80$, $n_1 = 25$, $s_1 = s_2 = 25$. - *Scenario 4:* $N = 105$, $n_1 = 25$, $s_1 = s_2 = 25$. - Figures in 'Sectors' columns are averages for the two sectors. - *All but one* and *All but two* rows indicate that we are allowing one or two triplets to fail in showing a cointegration relationship (see remark 7).

Appendix G FOR ONLINE PUBLICATION ONLY: False discovery in Monte Carlo experiments

In this appendix we present the figures of false discovery corresponding to the Monte Carlo experiments of §IV. For each experiment, DGP, scenario and sample size we count the number fully cointegrated subsets in excess of the true number, which is always one. We denote the number of additional subsets as *Num ex.Subsets*. For each of these additional subsets we count the number of series which are not fully cointegrated (*Num Sers*). Note that when the first fully cointegrated subset does not include all the correct series, the additional subsets may include some fully cointegrated series.

Table G.1 include the average across experiments of *Num ex.Subsets* and *Num Sers*. As the table shows, false discovery is not a relevant issue for $T = 200$ and $T = 400$ in any case. With $T = 100$ it could be problematic, particularly in DGPs 2 and 3. Given that the number of wrong series is always less than 6, disregarding subsets with less than 6 series will solve the problem for this sample size.

TABLE G.1:
False Discovery in DGPs with Gaussian innovations

	$n_1 = 10$	$n_1 = 25$	$n_1 = 40$
	Num. ex.Subsets × *Num. Sers*	*Num. ex.Subsets* × *Num Sers*	*Num. ex.Subsets* × *Num. Sers*
		DGP 1 - Johansen	
T=100	0.12 x 4.3	0.56 x 1.3	1.15 x 0.5
T=200	0.05 x 4.9	0.05 x 3.4	0.06 x 1.5
T=400	0.03 x 4.9	0.03 x 3.6	0.02 x 1.9
		DGP 1 - BIC	
T=100	0.13 x 5.1	0.27 x 4	0.71 x 3.1
T=200	0.002 x 5	—	—
T=400	—	—	—
		DGP 2 (non-diagonal Φ*) - Johansen*	
T=100	1.64 x 5.2	1.95 x 3.6	2.07 x 2.1
T=200	0.04 x 5	0.12 x 2	0.51 x 0.4
T=400	0.03 x 5.1	0.04 x 3.3	0.06 x 1.9
		DGP 2 (non-diagonal Φ*) - BIC*	
T=100	1.08 x 5.3	0.85 x 4.4	2.17 x 1.3
T=200	0.01 x 5.3	0.01 x 4.8	0.05 x 0.6
T=400	—	—	—
		DGP 3 (non-diagonal Φ *and complex long-run structure) - Johansen*	
T=100	2.45 x 5.1	2.74 x 3.8	2.97 x 2.9
T=200	0.07 x 4.4	0.14 x 1.5	0.53 x 0.3
T=400	—	—	—
		DGP 3 (non-diagonal Φ *and complex long-run structure) - BIC*	
T=100	1.46 x 5.4	1.22 x 4.9	2.81 x 2.1
T=200	0.01 x 5	0.02 x 3.1	0.08 x 0.8
T=400	—	0.002 x 5	—