Discovering common trends in a large set of disaggregates: statistical procedures and their properties

Guillermo Carlomagno*
Department of Statistics and Econometrics, University Carlos III of Madrid
and
Antoni Espasa
Department of Statistics and Econometrics, Instituto Flores de Lemus,
University Carlos III of Madrid

Abstract

Macroeconomic variables are, in general, weighted averages of a large number of components. The objective of this paper is to design a procedure to model all the N components. Our contribution concerns cases with a large number of components, for which traditional multivariate approaches are not feasible. We extend the pairwise approach initially proposed by Espasa and Mayo-Burgos (2013) and study its statistical properties analytically. The main feature of our procedure consists of discovering groups of components that share single common trends. We also provide Monte Carlo evidence on small samples and carry out a comparison with a DFM alternative.

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

JEL: C01, C22, C32, C53.

1 Introduction

Macroeconomic variables are weighted averages of a large number of components. Therefore, 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 the components of a macro or business variable at the maximum level of disaggregation. Our strategy consist of identifying and estimating relevant relationships between the components and then exploiting those relationships in single-equation models for all the disaggregates. This strategy can produce relatively precise forecasts of the components and may lead to an accurate indirect forecast for the aggregate.

There are several relevant reasons for disaggregating a macro-variable. Probably the most important one is that the analysis of disaggregated data in themselves may be of interest for

*The first author greatly acknowledges the travel grant from the International Association of Applied Econometrics. Correspondence to Guillermo Carlomagno. E-mail: gcarloma@est-econ.uc3m.es. Phone: (+34)916249674. Fax: (+34)916249847.

decision makers. When dealing with disaggregates of macro variables, a great heterogeneity in their dynamic patterns is observed. This heterogeneity implies that the analysis of the aggregate is irrelevant when the interest is in the components. Furthermore, even when the interest is in the 'big-picture', disregarding the components implies a great deal of informational losses.

The discussion about the 'optimal' level of disaggregation is still open, and exceeds the scope of this paper. Still, two remarks about this issue are in order. First, the fact that all statistical offices around the world are reporting more and more disaggregated data suggests that this data is relevant for decision makers, and should be analyzed. Second, since the 'optimal' level of disaggregation depends on the objectives of the analysis and on the dynamic properties of the data, it is mainly an empirical matter. Accepting that for deciding on the level of disaggregation one should use a 'general to specific approach', better than a 'specific to general' one, the analysis at the maximum disaggregation level becomes necessary, even if we end up using a lower one.

When working with an aggregate composed by 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).

Giacomini and Granger (2004) and Hendry and Hubrich (2011) develop alternatives to tackle this trade-off. Giacomini and Granger work with spatially correlated series and propose to use a restricted multivariate model in which restrictions are assumed known. Hendry and Hubrich, whose objective is to model and forecast only the aggregate, propose to use disaggregate information in the model for the aggregate.

Another alternative for dealing with the informational loses vs. the estimation uncertainty trade-off is the consideration of common features proposed by Espasa and Mayo-Burgos (2013). The 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 technological or changes in preferences in different ways. Thus, as Espasa and Mayo-Burgos argue, a valid hypothesis may be that specific subsets of components share common features, but others do not.

To exploit the restrictions derived form the existence of those subsets, Espasa and Mayo-Burgos suggest trying to discover blocks of components that share unique common features (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. The search for those blocks is carried out by performing common features tests between all the $N(N-1)/2$ pairs that exist in a set of N components. Castle and Hendry (2010) also highlight the importance of including long and short-run common features restrictions in the individual models for the components, as proposed by Espasa and Mayo-Burgos (2013).

The problem of how to impose 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), the choice of the data from which the factors are extracted is not innocuous. Results are more accurate when factors are extracted from data that is 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), and Breitung and Eickmeier (2015), and Ando and Bai (2015). To the best of our knowledge, the only attempt to try to statistically discover the restrictions in DFM is the proposal by Ando and Bai, the other authors assume the restrictions to be known. Ando and Bai deal with stationary series with a grouped factor structure and develop a procedure to determine the series in each group and estimate the factors. The size of the groups is assumed to go to infinity. Instead, our approach deals with $I(1)$ variables and sizes of the groups which need not to go to infinity.

Bailey, Kapetanios, and Pesaran (2015) — BKP— and Bailey, Holly, and Pesaran (2015) — BHP— also deal with the issue of non-pervasive factors. BKP propose a measure for the degree of non-pervasiveness of the factors, and BHP develop a two stage procedure for dealing with pervasive and non-pervasive factors at the same time. Similarly to Ando and Bai (2015) these authors restrict their attention to stationary series and the cross-sectional dimension going to infinity.

A central aim of this paper — in which the consideration of common features is restricted to common trends — is to show that the pairwise strategy initially proposed by Espasa and Mayo-Burgos (2013) can be used as an objective method to discover blocks of components sharing single common trends. Our contribution in that respect consists of providing the statistical properties of the strategy using analytic and Monte Carlo procedures.

We show that the probability of finding cointegration between all possible pairs in a subset of series with a unique common trend tends to $(1-\varphi)$ in large samples, where φ is the nominal size used in the Johansen's trace test for each pair (theorem 1). Monte Carlo experiments confirm this result, and show that the pairwise strategy dominates standard Dynamic Factors Models in

several situations of empirical interest. This result only needs T going to infinity, but we argue that it extends to the case of both T and N going to infinity.

Another contribution of this paper is a comparison of the pairwise strategy's power for discovering sets of series that share a single common trend, with that of the Johansen's full vector approach. In our specific framework and for relatively short samples, the former improves the power with respect to the latter.

A related approach is that of Pesaran (2007), who develops a strategy for testing output and growth convergence across countries. For a group of N countries, the author performs unit root tests for all the $N(N-1)/2$ differences between pairs of log GDPs. Among other details, our approach differs in that we want to discover subsets of series with single common trends, not testing if the series in a specific subset share the trend.

The rest of the paper is organized as follows. In §2, we study the statistical foundations of the pairwise procedure. In §3, we perform a Monte Carlo experiment to confirm the results of previous section and compare the performance of our procedure with that of an alternative based on standard Dynamic Factor Models. This section includes an initial discussion for the case of '*general*' and '*sectoral*' trends. In §4 apply the pairwise procedure to the US CPI, and §5 is devoted to the conclusions.

2 Statistical foundations of the pairwise procedure

2.1 General framework and assumptions

The general framework for the models we work with is given by a VAR model for all 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 with noises; μ_t contains the deterministic components (constants and trends); $\Pi(z)$ 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 (VEqM):

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

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 three assumptions:

Assumption A *The N components are generated by the VEqCM in eq. (2).*

Assumption B *The N components are $I(1)$.*

Assumption C *There is, at least, one subset of n_1 components that share a unique common stochastic trend.*

Assumption D *The residuals of eq. (2) are iid and multivariate normally distributed.*

Assumption B avoids the complication of dealing with $I(2)$ patterns or seasonal unit roots in cointegration analysis, and rules out the 'trivial' cointegration relationships that will appear when there are $I(0)$ components. As we argued by Carlomagno and Espasa (2015) this is a sensible assumption for the components of the US CPI.

Assumption C gives relevance to our objective of discovering subsets of series with a unique common trend, and assumption D is necessary for the Johansen's maximum likelihood procedure.

Remark 1 *Instead of assumption B we could assume that at least n_1 components are $I(1)$ and the others are either $I(1)$ or $I(0)$. This new assumption would require testing the significance of the cointegration relationships' coefficients (β). This is how Carlomagno and Espasa (2015) proceed in an empirical application.*

Remark 2 *Assumption D is related to the residuals, not to the first differences of the components. Although this distinction is not relevant for the Monte Carlo experiments, it is important for the empirical applications. Carlomagno and Espasa (2015) develop a procedure for dealing with outliers and breaks in the framework of the pairwise approach. By allowing for outliers and location shifts in the distribution of the first difference of the components, they do not need to assume normality of ΔX_{it}. The only requirement is that normality can be achieved after correcting for a few outliers location shifts, which as argued by Juselius (2015) is a quite general assumption in VAR models.*

Remark 3 *Cheung and Lai (1993) show that the Johansen's trace test is not substantially affected by skewness and/or excess kurtosis of the residuals. Therefore, even if the assumption*

that after admitting outliers and location shifts ΔX_{it} is normal, is not valid, we do not expect a substantial deterioration of our procedure.

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 of ϵ_t has not special restrictions.*

The pairwise strategy looks for blocks of components that share just one common trend, and requires performing Johansen's cointegration tests between all possible pairs of components. For each pair, a bi-variate VAR model has to be estimated and the lag length determined in each case. Next, the procedure requires constructing subsets in which every series is cointegrated with all the others. We call these subsets *'fully cointegrated'* subsets.

After that, a single equation model for each component can be estimated, including as potential regressors all the possibly relevant cointegration relationships (if any) found in previous step, 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)). The resulting modeling procedure 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.

2.2 Asymptotic properties of the pairwise procedure

As we argue below, the asymptotic validity of the pairwise strategy only requires T going to infinity, N can remain fixed. In our view, this is a strength of the procedure for we do not need to assume that the number of components of an aggregate goes to infinity in order to achieve consistency. As also noted by Cubadda and Scambelloni (2015)), such an assumption could be considered as quite unrealistic. Nonetheless, since we want to deal with a 'large' number of components, we also study the behavior of the procedure when both T and N go to infinity.

Assume that we are dealing with a macro-variable composed by N basic components. The total number of pairs is $N(N-1)/2$, so that we initially need to perform this number of cointegration tests. Assume further that there are J subsets of much smaller dimension (n_j elements in each subset) such that the elements in each subset share a unique common trend, and the other $N - \sum_j n_j$ components have their own trends. Thus, there are $J + N - \sum_j n_j$ 'common' trends in the system. We will use the notation n_j both, to indicate the size and to label the 'fully cointegrated subsets'

The ideal properties of the procedure are: (1) Cointegration tests between all possible pairs in each n_j should indicate the existence of a cointegration relationship, and (2) no series outside n_j should be wrongly included in the estimated subsets of series that share a unique common trend. Abusing notation, we call those subsets $\hat{n}_j$.

Since we are performing Johansen's tests, the procedure inherits its asymptotic properties. There are, however, two specific features that deserve special attention: multiple testing and estimation of partial models.

2.2.1 The general problem of multiple testing

In the regular framework in which there is not repeated hypothesis testing, the probability of not false rejecting the null is $1-\varphi$ (with φ being the nominal size of the test). When m tests are performed, assuming that they are independent, the probability of not making any false rejection reduces to $(1-\varphi)^m$, and the probability of making at least one error is $1-(1-\varphi)^m$, which rapidly increases with m.

Several approaches have been proposed for controlling type I error rates in multiple testing frameworks, among which those that try to control either the Family Wise Error Rate ($FWER$) seem to be the most popular (see, e.g., Benjamini and Hochberg (1995)). Defining V as the number of true null hypotheses that were wrongly rejected $FWER = P(V \geq 1)$.

The Bonferroni correction adjusts all p-values in a single step for ensuring that $FWER \leq \varphi$. In the case of $m = 2$, with independent tests statistics we have $P(WR_1 \cap WR_2) = \varphi^2$ (where WR_j is the event of wrongly rejecting hypothesis j) and the Bonferroni correction delivers $FWER = \varphi$. When $P(WR_1 \cap WR_2) > \varphi^2$, $FWER$ is smaller than φ and the Bonferroni correction is too stringent, even if we are interested in testing whether at least one individual hypothesis is false (this is frequently called the *universal null hypothesis*).

A case of interest may be when $P(WR_1|WR_2) \simeq 1$ (or $P(WR_2|WR_1) \simeq 1$), such that $P(WR_1 \cap WR_2) \simeq \varphi$. In this case, $FWER \simeq \varphi$, and there is no need to adjust p-values, even if the relevant hypothesis is the *universal* one. This last argument can be generalized for the case of m tests. That is, if the probability of wrongly rejecting any combination of the m hypothesis at the same time is close to φ, it can be easily seen that $FWER \simeq \varphi$, and there is no need for correcting p-values.

2.2.2 Multiple testing in the pairwise approach

Since the pairwise procedure involves a large number of cointegration tests (e.g., 4950 for $N = 100$), it may be thought to raise the probability of false rejection. We analyze this issue for the three different types of pairs: i) Both series belong to the same n_j so that, in the context of Johansen's tests, the problem is rejecting the true hypothesis of $r = 1$ — one cointegration relationship[1]; ii.a) only one series belongs to some n_j, or both belong to different subsets, so that the problem now may be rejecting the true hypothesis $r = 0$; and ii.b) none of the series belongs to any n_j so that the problem again may be rejecting the true hypothesis $r = 0$.

False rejection of $r = 1$

In the Johansen procedure, the null hypothesis $r = 0$ 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 in large samples. 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* series in the same n_j would be $(1 - \varphi)^{n_j(n_j-1)/2}$, which quickly decreases with n_j. But clearly, tests are not independent. Theorem 1 below indicates that these tests are asymptotically equivalent in the sense that the probability of obtaining the same result in all of them tends to 1 as T goes to infinity.

Theorem 1 (Asymptotic equivalence of pairwise cointegration tests in a fully cointegrated subset). *Under assumptions A, C and D, given a subset of Q pairwise cointegrated series (i.e., there are $Q - 1$ cointegration relationships among them and a single common trend), the probability of obtaining the same result in all the $Q(Q-1)/2$ pairwise Johansen's trace tests tends to 1 as $T \to \infty$.*

Proof *See appendix A*

Another way to interpret the this theorem could be: *transitivity is a property not only of cointegration, but also of cointegration tests.* The intuition for this result is that, asymptotically, the $Q(Q-1)/2$ cointegration tests are tests for one versus no common trend, which can be seen as unit root tests for the estimated common trend. Since this trend is the same for all series, we have $Q(Q-1)/2$ estimations of the same trend, which tend toward the same process as T goes to infinity.

[1] Although Johansen's test is sequential, the probability of not rejecting $r = 0$ is asymptotically 0, for asymptotic power is 1.

Two relevant implications of theorem 1 are:

Implication 1 *Let WR_{ih} be the event in which the null of $r = 1$ is wrongly rejected for the pair (i, h). Theorem 1 implies that the joint probability for any combination of WR_{ih} (for any i, h belonging to n_j) tends to φ. Therefore, even in the case when the hypothesis of interest is the universal one defined in the Bonferroni approach (i.e., false rejecting at least one of the $n_j(n_j - 1)/2$ hypotheses), p-values need not be corrected.*

Proposition 1 *Theorem 1 works both when Q is fixed and when it goes to infinity.*

Proof *See appendix B.*

Implication 2 *A procedure for testing the hypothesis that a certain group of series share a single common trend (similar to, but less restrictive than, the hypothesis of interest in Pesaran (2007)) could be: i) Test $r = 0$ vs $r > 0$ in all possible pairs using regular critical values — asymptotic power is 1. ii) If all hypothesis are rejected, test $r = 1$ vs $r = 2$ in all pairs, store the maximum test statistic and compare it with regular critical values. Asymptotically, this procedure delivers the correct size. However, this procedure is not useful for discovering sub-groups of series sharing a unique common trend. To do this, we group series that are pairwise cointegrated ($r = 1$).*

False rejection of $r = 0$ when only one of the series belongs to a fully cointegrated subset, or each series belongs to a different subset.

Under assumption B, the true number of cointegration relationships between one series inside and one series outside a fully cointegrated subset is $r = 0$.

Two comments are relevant for this case. First, since to include a series in $\hat{n}_j$ we require all cointegration tests to find $r = 1$, it is evident that the *universal* null — relevant for Bonferroni corrections — is of no interest at all. What is relevant for the pairwise procedure is the probability of wrongly rejecting all $r = 0$ hypotheses, which, in any case, will be smaller than or equal to φ.

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

$$P(WR_1 \cap ...WR_{\hat{n}_j}) = P(WR_1|WR_2, ..., WR_{\hat{n}_j}) \times ... \times \ P(WR_{\hat{n}_j-1}|WR_{\hat{n}_j}) \times P(WR_{\hat{n}_j}), \quad (3)$$

where $P(WR_i)$ is the nominal size of the pairwise tests (φ). Using the extreme assumption that all the $\hat{n}_1 - 1$ conditional probabilities in eq. (3) 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 $\hat{n}_j$, $E[W]$, would be $(N - n_j)\varphi$. For $N - n_j = 100$ and $\varphi = 0.01$, $E[W] = 1$, which is quite tolerable.

Under this extreme assumption, the ratio of wrong over true elements in the estimated n_j would be:

$$\frac{(N - n_j)\varphi}{n_j} \tag{4}$$

Therefore, though we may have $N >> n_j$, when $N \to \infty$, we need n_j to grow at the same rate for avoiding the proportion of wrong elements to go to infinity. This implies a *pervasiveness* requirement similar to that of DFM (see, e.g., assumption B in Bai (2003)).

Discussion about the pervasiveness requirement

Interestingly, simulation results in §3 show that the actual figures are far below what eq. (4) suggests, indicating that the assumption that all conditional probabilities in eq. (3) are equal to one is quite extreme. This observation leads us to the relax the assumption as follows:

Assumption E *There is a proportion* δ $(0 < \delta \leq 1)$ *of the* $\hat{n}_j - 1$ *conditional probabilities in eq. (3) that do not exceed a fixed threshold* φ_{max}*, with* $0 \leq \varphi_{max} < 1$*.*

Note that we are not assuming the proportion δ to be fixed as N grows, so that the assumption does not seem too restrictive. Now, an upper bound for the expected ratio of wrong over true elements in $\hat{n}_j$ is:

$$\overline{E[W]} = \frac{(N - n_1)\varphi_{max}^{\delta(\hat{n}_j - 1)}\varphi}{n_j} \tag{5}$$

Still, when $N \to \infty$ we also need $n_1 \to \infty$. However, provided that $N/\delta > 1$, it can be shown that we can have both; $N/n_j \to \infty$, and eq. (5) fixed. Thus, we would not need a pervasiveness assumption in the sense of DFM; a necessary (though not sufficient) condition for eq. (5) to remain fixed is $n_j/log(N) \to \infty$. A formal proof of these statements is available upon request.

Although we do not have a formal proof to show that $N/\delta > 1$ as $N \to \infty$, a simple Monte Carlo study (available upon request) suggests it does.

False rejection of $r = 0$ when none of the series belongs to any fully cointegrated subset.

From Johansen's test properties, the probability of finding one cointegration relationship between two non-cointegrated I(1) series (recall assumption B) tends to φ as $T \to \infty$. Assuming that tests are independent, we can think of the subset of series not belonging to any n_j as a random graph with edge probability equal to φ. Then, a lower bound for the expected number of estimated fully cointegrated subsets composed by K series of the $N - \sum_j n_j \equiv \tilde{N}$ would be $E[K_{worng}] = C_K^{\tilde{N}} \varphi^{K(K-1)/2}$, which is almost zero for, say, $\varphi = 0.01$, $K > 3$ and moderately large $\tilde{N}$ (see Bollobás and Erdös (1976) for a discussion of *cliques*[2] in random graphs).

Additionally, Matula (1976) shows that the size of the *maximal clique*[3] in a random graph with M nodes and edge probability φ has a strong peak around $2log(M)/log(1/\varphi)$, which is 2 for $M = 100$ and $\varphi = 0.01$. Thus, selecting a low φ and disregarding estimated fully cointegrated subsets with fewer than three elements constitutes a *strong protection* against finding fully cointegrated subsets among these series.

Note that $E[K_{worng}]$ is a lower bound since we assumed that tests are independent. Notably, simulation results show that the actual number of wrong fully cointegrated subsets is close to this bound, meaning that, even though we are performing tests between all the pairs of a group of series, the independence assumption is sensible for this type of pairs (see appendix C and §3).

Under the independence assumption, we can use the result in Matula (1976) about the size of maximum fully connected sub-graphs to study the behavior of the procedure when $N \to \infty$. An additional assumption is required:

Assumption F $\frac{T}{N^{1/\kappa}} \to \geq c$, *when* $[T, N] \to \infty$, *for some* $c > 0$, $\kappa = -\frac{log(N)}{log(\varphi)}$, *and* φ *the nominal size of the pairwise tests.*

For our case of interest, and under assumption B, the size of this fully connected sub-graph has a strong peak at $2log(\tilde{N})/log(1/\varphi)$. Therefore, in order to avoid the maximum false fully cointegrated subset to go to infinity we need to choose φ as an inverse function of N. Using $\varphi = N^{-1/\kappa}$, for some $\kappa > 0$, an upper bound for the maximum size will be 2κ.

For being able to use such a significance level, we need assumption F to ensure that T grows at a rate larger than or equal to that of $N^{1/\kappa}$. Since κ can be larger than 1, we can deal with

[2] Cliques are sub-graphs in which all nodes are pairwise connected.
[3] The maximal sub-graph in which all nodes are pairwise connected.

the case of $N/T \to \infty$.

2.2.3 Summary and discussion of the problem of multiple testing in the pairwise approach

There are three different situations to distinguish: 1) Pairs inside some fully cointegrated subset; 2) pairs between a series inside and a series outside a fully cointegrated subset; and 3) pairs between two outsiders.

In the first case the problem could be wrong rejecting $r = 1$ too much, and include too few series in the estimated subset. Theorem 1 rules out this possibility by stating that the probability of including all the correct series is $1 - \varphi$, regardless the size of the true subsets.

In case 2, the problem could be wrong rejecting $r = 0$ and including too much wrong series in the estimated fully cointegrated subsets. We showed that the requirement of full cointegration provides a strong protection against this problem. Furthermore, we argued that we could have both; a fixed proportion of wrong series, and $N/n_j \to \infty$, meaning that we would not need a pervasiveness assumption in the sense of DFM. Though we did not derived a precise minimum requirement for the growth rate of n_j, a necessary (yet not sufficient) condition for keeping the proportion of wrong elements fixed is $n_j/log(N) \to \infty$.

The absence of a precise sufficient condition, softer than $N/n_j \to c$, may raise some concern, as this is a pervasiveness requirement in the sense of DFM. Thus, the fully cointegrated subsets we are looking for could, in principle, be discovered by Principal Components, which is a much simpler procedure than the pairwise approach. Two comments are worth in this respect; first and most important, when N is fixed we can deal with $N >> n_j$ and small n_j at the same time. In §3 we show that DFM fail in those situations. Second, as highlighted in remark 4 our procedure allows for quite rich 'idiosyncratic' dynamics while in DFM they are much more restricted (see, e.g., assumption B in Bai (2003), and assumptions A.2 and B in Doz et al. (2012)).

In the third case, the problem could also be wrong rejecting $r = 0$ too much and 'discover' false fully cointegrated subsets. We argued that choosing tight significance levels and disregarding estimated fully cointegrated subsets of small size is a strong protection against this issue. For the case of $N \to \infty$, we could use a significance level of $N^{1/\kappa}$ — for some chosen $\kappa > 0$ — and disregard estimated subsets with fewer elements than 2κ. This procedure requires $T/N^{1/k} > c$ as $[T, N] \to \infty$. Since κ can be larger than 1, previous condition includes the case $N/T \to \infty$.

2.2.4 Partial systems

The strategy of performing cointegration tests between all possible pairs of series is justified by the fact that in a set of n_j series that share a unique common trend, there are $n_j - 1$ cointegration relationships and the series are pairwise cointegrated.

This strategy requires partial systems' estimation in the sense that we assume the existence of a full VAR model for all the components but estimate several partial bi-variate systems. The models considered in the pairwise procedure are partial in the sense that we consider only a subset of variables, but not in the sense of Johansen (1992) and Harbo et al. (1998). That is, we are not seeking to estimate all the cointegration parameters from a bi-variate model (which is not possible). On the contrary, under full cointegration, since every pair of variables is cointegrated, the bi-variate VAR models are complete because all relevant variables are considered endogenous.

Interestingly, when the dynamic structure of each bi-variate model is selected using some information criteria, the power of the pairwise procedure for finding the true number of cointegration relationships is improved with respect to the traditional Johansen's trace test. In §3.2, we show this by means of a small simulation experiment.

3 Simulation results for the Pairwise strategy

In this section, we perform two different Monte Carlo experiments. The first one (in §3.2) is designed to compare the power properties of the pairwise approach to find the true number of cointegration relationships with respect to the traditional Johansen's trace test (see the discussion in §2.2.4). The second experiment is designed with two objectives: confirm the analytic results presented §2.2.2, and compare the performance of the pairwise approach with an alternative based on standard Dynamic Factor Models (§3.3). It is important to note that our objective is not making a general comparison between our approach and DFM; we do not want to extract general results. Our goal is much simpler, we just want to evaluate if the simple strategy of estimating dynamic factor models can be used in our framework of interest, namely, relatively large N and small n_j.

Before presenting the simulation results, in §3.1 we describe the general design of the experiments.

3.1 General design of the experiments

We consider two classes of DGPs: VECqM and Dynamic Factor Models. For space reasons we only describe the VECqM, full details of the DGP and the simulation results for the case of

DFM are available upon request.

DGP 1 - VEqM 1

The general expression for the VEqCM 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), and $\epsilon_t \sim N(0, I_N)$.

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, matrix α is set to have the following structure:

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\alpha_2 & 0 & 0 & \cdots & 0 \\ 0 & -\alpha_3 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & -\alpha_{n_1} \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{N \times r} = \begin{pmatrix} \alpha^U \\ \alpha^D \end{pmatrix}, \tag{6}$$

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 the specification of a DGP, given a selection of β, the choice of α does affect the properties of the process. We are assuming that the common trend among series in n_1 is driven by a single

shock, so that the first variable is exogenous and each cointegrating relation affects only one of the remaining variables.

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].

There are two important observations about this DGP. First, at first glance, the selected structure for matrix α^U could seem too simple to be realistic. However, the complexity of the system cannot be judged from matrix α alone. For example, it can be shown that this DGP is exactly equivalent to a DGP with a 'complete' matrix α^U in which cointegration relationships are normalized with respect to the sub-aggregate formed by the first n_1 series. Thus, with such a normalization we would have the same system but it would not be subject to the critique that α is too simple.

Although the system does not have short run interactions (both Φ_1 and the residuals' covariance matrix are diagonal), all bi-variate sub-systems that include at least one of the series in n_1 have invertible MA dynamics. Thus, the only bi-variate systems that can be written as pure VAR(1) models are those composed by two outsiders (for a discussion about linear transformations in VARMA models, see Lütkepohl (1984)).

DGP 2 - VEqM 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 $i = 1, ..., n_1$, where $\lfloor A \rfloor$ is the operator that takes the integer part of number A. 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 form the uniform distribution $U_{[0.05,0.1]}$, and

$$q = \begin{cases} 5 & \text{if } 5 < i < N-5, \\ 10-i & \text{if } i \leq 5, \\ 10-(N-i) & \text{if } i \geq N-5. \end{cases} \tag{7}$$

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). When generating DGP 1 and DGP 2 we should obtain systems with $N - n_1 + 1$ unit roots and all the other roots outside the unit circle. Since the way we generate the series does not ensure this, we order the roots from smallest to largest and disregard cases in which the root in position $N - n_1 + 2$ is smaller than 1.01.

This DGP generates quite complex short run dynamics, and there are no bi-variate subsystems with purely finite VAR structures, as all of them have invertible MA components.

3.2 Power comparisons: Pairwise vs. Johansen's trace test

The objective of this section is to compare the performance of the pairwise strategy with that of the full system Johansen's approach in situations in which this comparison makes sense and the later approach is feasible (see discussion in §2.2.4). To do so, we compare the power of the pairwise procedure for finding the true number of cointegration relationships ($n_1 - 1$) with that of the full system Johansen's trace test. For DGPs 1 and 2 described above, we consider the following possibilities for $[N, n_1]$: i) $[6, 2]$. There are six variables and one cointegration relationship between two of them (five common trends). ii) $[6, 3]$. There are six variables and two cointegration relationships between three of them (five common trends). iii) $[9, 2]$. There are nine variables and one cointegration relationship between two of them (eight common trends). iv) $[9, 4]$. There are nine variables and three cointegration relationships between four of them (six common trends).

For each of these four sub-DGPs, we perform the Johansen's trace test and the pairwise procedure. In the trace tests, we include only one lag, which is the true number. As discussed in §2.2.4, in the pairwise procedure, the lag structure depends on the *type of the pair* (i.e., both series have the common trend; one has it but the other does not; neither of the series has it). Thus, we try from one to five lags and select the optimal number according to the AIC and the BIC.

Cointegration tests are made at 1% of significance, and the number of Monte Carlo replications was 500. The experiments are performed with samples of 100, 200 and 400 data points.

The *Trace* columns in table 1 contain the probabilities of finding the correct number of cointegration relationships by means of the Johansen's trace test, when all the N variables are included in the model. The PW columns contain the probabilities of finding cointegration in all the pairs

that are truly cointegrated when the tests are done by the trace test, but in a pairwise fashion, and the lag length is selected according to the AIC (BIC). The preferred approach is marked in bold.

The table shows that nothing is lost by proceeding in a pairwise fashion. On the contrary, the pairwise procedure outperforms the regular trace test. For large samples and a small number of series (N), both procedures provide the same results (which coincide with the theoretical ones). However, as the number of series increases or the sample size decreases, the differences in favor of the pairwise procedure become remarkable (bold entries are only in the PW columns). The largest differences are for the case with $N = 9$ and $T = 100$.

This result is closely related to those obtained by Lütkepohl et al. (2003) and Johansen (1995). These authors find that the power of cointegration tests decreases with the number of stochastic trends in the system, so that, for instance, it would be more difficult to detect a single cointegration relationship in a three-dimensional system than in a bivariate one. Note, however, that our result is not exactly the same since in table 1 we are comparing estimation and testing cointegration in a single *full* model with several stochastic trends *vs.* doing it in several bivariate models (not one) with one stochastic trend. With our strategy we provide a better alternative to the Johansen's procedure for the case in which the cointegrated series constitute a fully cointegrated subset.

Note also the importance of lag selection for small sample sizes. The difference in the probabilities of finding all cointegration relationships with the pairwise procedure when $T = 100$ if we use the AIC or the BIC may be significant in favor of the latter. This is due to the efficiency losses generated by a larger number of regressors in small sample sizes (the BIC tends to select shorter lag lengths).

3.3 The behavior of the pairwise strategy and comparison with DFM

We now turn to the analysis of the pairwise strategy in the framework for which it has been designed: large N. As discussed in §2.2, the ideal procedure will: 1) from all the N series, identify a large proportion of those that truly share the trend (those in n_1); and 2) not include wrong series in the estimated fully cointegrated subset(s), $\hat{n}_1$.

Conditions 1 and 2 are closely related to what Castle et al. (2011) call *potency* and *gauge*. While *gauge* measures the retention frequency of irrelevant variables when selecting among a (potentially large) set of candidates, *potency* denotes the average retention frequency of relevant

Table 1: Probability of finding all cointegration relationships. Comparison between the Trace test and the Pairwise procedure

DGP 1: VEqCM 1 — with diagonal Φ

		T = 400			T = 200			T = 100		
			PW			PW			PW	
N	n_1	Trace	AIC	BIC	Trace	AIC	BIC	Trace	AIC	BIC
6	2	0.97	**0.99**	**0.99**	0.91	**0.99**	**0.99**	0.65	0.92	**0.96**
	3	**0.98**	**0.98**	0.96	0.88	**0.97**	0.95	0.53	0.62	**0.80**
9	2	0.89	**0.99**	**0.99**	0.57	**0.99**	**0.99**	0.07	0.92	**0.96**
	4	0.95	**0.98**	0.94	0.67	**0.92**	0.91	0.38	0.39	**0.64**

DGP 2: VEqCM 2 — with non-Diagonal Φ

		T = 400			T = 200			T = 100		
			PW			PW			PW	
N	n_1	Trace	AIC	BIC	Trace	AIC	BIC	Trace	AIC	BIC
6	2	0.96	**0.99**	**0.99**	0.91	**0.99**	**0.99**	0.68	0.88	**0.91**
	3	**0.98**	**0.98**	0.96	0.87	**0.94**	**0.94**	0.43	0.65	**0.72**
9	2	0.92	**1.00**	**1.00**	0.72	**0.99**	**0.99**	0.23	0.85	**0.90**
	4	0.96	**0.97**	0.95	0.57	0.87	**0.88**	0.38	0.26	**0.40**

Number of replications: 500. *Trace* columns contain the probabilities of finding the correct number of cointegration relationships by means of the Johansen's trace test, when all the N variables are included in the model. *PW* columns contain the probabilities of finding cointegration in all the pairs that are truly cointegrated when the tests are done by the trace test, but in a pairwise fashion and the lag length is selected according to the AIC (BIC) criteria.

variables.

For the two DGPs describe above and the two DFM (results for the later are available upon request), 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, to indicate the size of the 'fully cointegrated subset' and as its label. The three choices are $n_1 = 10$, $n_1 = 25$ and $n_1 = 40$. Scenarios 1 and 3 are motivated by results in Espasa and Mayo-Burgos (2013) about CPIs' components, and scenario 2 is just to have an intermediate structure. 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 we simulate a 100-dimensional model in which a subset of n_1 series share a single trend. The objective is to discover the series in n_1. To do that, we perform cointegration tests on all the 4950 bi-variate VAR sub-models that exist among the 100 series. Thus, for a particular DGP, scenario and sample size, we have 2.475 million sub-models (4950 for each replication). Since we have four DGPs (the two described above plus two DFM), three scenarios and three

sample sizes, we have $(4 \times 3 \times 3) \times 2.475 = 89.1$ million sub-models to estimate. Additionally, since the lag length for each of the 4950 sub-models of a particular replication is unknown, we select it with the AIC in a model with one cointegration relationship and admitting between one and five lags in the VEqM representation.

In order to compare our procedure with DFM, we consider two estimation strategies of DFM; Principal Components (PC) and the Quasi Maximum Likelihood (QML) procedure proposed by Doz et al. (2012). The later approach is an iterative procedure that combines the estimation simplicity of PC with the efficiency of the Kalman Filter Smoother.

In both cases we extract the factors from the whole data set and keep the number of factors suggested by the information criteria IC_k and the three penalty functions detailed in Bai (2004). When each penalty function suggests a different number of factors, we choose the minimum; otherwise, we choose the mode. This procedure implies that we are not always using the same penalty function in each experiment, but it artificially helps the dynamic factors methodology to pick the correct number of factors (which is always one by construction). For the three scenarios, we proceed as suggested by Bai and Ng (2004): extracting the factors from the differenced data and integrating the results to obtain estimates of the original factors. This seems the most sensible procedure when n_1 is small compared to N, as idiosyncrasies for series outside n_1 are $I(1)$.

In order to have a proper comparison with the pairwise procedure, we compute confidence bands for the factor loadings and identify those series with statistically significant factor loadings (at 0.5% of significance). We call the subset formed by those series as the *DFM counterpart of the fully cointegrated subset*. Using those series, we can compute the gauge and potency of the DFM approaches. To compute the variance of the loadings estimated by PC we consider the asymptotic variance derived by Bai (2003). For the QML procedure, and also as a second alternative for PC, we use a HAC estimator with data driven bandwidth and quadratic spectral kernel (see Andrews (1991)) in the regression $\Delta X_t = \Lambda \hat{F}_t + \epsilon_t$ estimated by OLS.

Therefore, we end up with three alternative DFM estimation procedures: $PC1$ is PC with variances computed as in Bai (2003); $PC2$ is PC with the HAC estimator of the variance; and the QML procedure of Doz et al. (2012). Since basic conclusions do not change substantially, we only report results for $PC1$ (the other are available upon request). Another interesting procedure that we do not explore could be that proposed by Ando and Bai (2015).

Tables 2 and 3 include the *gauge* and *potency* of the pairwise strategy (PW) and $PC1$ for

DGPs 1 and 2.

Focus first on table 2 (DGP 1). The pairwise procedure 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 states that the asymptotic probability of finding cointegration in all the true cointegrated pairs is close to $(1 - \varphi)$, with φ being the nominal size of the tests (recall that 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 $0.002 \times (100 - 40) = 0.12$. Recall from §2.2 that an upper bound for the expected number of wrong series in $\hat{n}_1$ is $E[W] = (N - n_1)\varphi$ —0.6 in scenario 3. Therefore, this result shows that the actual $E[W]$ is far from this upper bound, meaning that the assumption that all conditional probabilities in eq. (3) are equal to one is quite extreme. Importantly, we did not find any other fully cointegrated set composed by outsiders, in accordance with the DGP used to generate the data.

Finally, as table 2 shows, 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 comparison with DFM. Table 2 shows that for DGP 1, the DFM procedure performs better than the pairwise approach in terms of *potency*. The probability of including a large proportion of the true series is larger in the DFM approach (except for scenario 3 and $T = 200, 400$). However, *gauge* is substantially larger in scenarios 1 and 2, even for large sample sizes (except for scenario 3 and $T = 400$).

The DFM approach fails to isolate the series in n_1 for scenario 1 for all sample sizes, and in scenario 2 for $T = 100$. In scenario 1 with $T = 400$, on average, we include $0.082 \times (100 - 10) = 7.4$ wrong series. This bad performance substantially deteriorates as the sample size decreases. For T = 100, on average, 24.2 wrong series are included in $\hat{n}_1$.

In summary, the main conclusion from table 2 may be that the pairwise procedure is preferred for situations of relatively small n_1 — a conclusion that seems more evident for relatively small sample sizes. When n_1 and T become larger, DFM may be preferred. Note, however, that even in those situations (large n_1 and T) the pairwise procedure also shows a very good performance.

In DGP 1, the only source of correlation between the series in n_1 is the common trend, and

the series outside n_1 are independent between each other and with respect to those inside n_1. This restriction is removed in DGP 2, as series have non-zero short run interactions with series inside and outside n_1. Table 3 includes the results for this DGP. The performance of PW is somewhat worse than in table 2, except for scenario 3 with $T = 100$ when the deterioration in terms of power 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 $(0.004 \times (100 - 40) = 0.24)$. On the contrary, the DFM approach, for $T = 100$, completely fails in all scenarios.

As the sample size increases, both procedures improve their performance in terms of *gauge* and *potency*. For $T = 400$ PW almost recovers its performance of table 2 in all scenarios, in contrast, the DFM procedure shows a good performance only for scenario 3. Thus, the main conclusion form table 3 is that PW dominates in all scenarios and sample sizes.

Overall, PW dominates DFM in almost all situations. In the simple $DGP1$ DFM could be preferred when both n_1 and T are not small, but performances of both procedures are very similar. In more complex DGPs (DGP 2) the DFM procedure fails in almost all situations and PW is clearly preferred. This conclusions remain valid when the DGP is an exact DFM of a DFM with auto-correlated and cross-correlated idiosyncrasies.

Table 2: Comparison of Gauge and Potency of the Pairwise procedure with DFM. $DGP1$ ($VEqCM1$ - diagonal Φ_1)

	Pairwise						**PC 1**					
	Sce 1		Sce 2		Sce 3		Sce 1		Sce 2		Sce 3	
	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot
T=100	0.4	77.0	0.2	67.0	0.2	63.2	26.9	94.1	5.1	97.8	1.0	87.3
T=200	0.3	96.4	0.2	95.3	0.1	94.5	18.3	99.5	0.7	99.5	0.0	93.0
T=400	0.3	99.2	0.2	99.0	0.2	98.6	8.2	100.0	0.0	99.8	0.0	96.0

- $Gauge = \frac{100}{(N-n_1)Nexp}\sum_{i=1}^{Nexp} Z_{2,i}$. - $Pot = \frac{100}{n_1 Nexp}\sum_{i=1}^{Nexp} Z_{1,i}$. - 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: $n_1 = 40$.

A final comparison between the two procedures could be their forecasting performance, but the one that performs better in grouping the components with common features is, in principle, expected to dominate the forecasting exercise.

3.4 Robustness check

In previous sections 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

Table 3: Comparison of Gauge and Potency of the Pairwise procedure with DFM. $DGP2$ ($VEqCM2$ — non-diagonal Φ_1)

	Pairwise						**PC 1**					
	Sce 1		Sce 2		Sce 3		Sce 1		Sce 2		Sce 3	
	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot
T=100	2.2	67.5	1.1	60.2	0.4	26.0	42.2	95.7	17.4	98.7	6.3	93.9
T=200	0.2	90.6	0.2	84.2	0.1	80.5	39.5	98.6	12.1	99.6	2.3	96.4
T=400	0.2	98.7	0.2	98.2	0.1	96.9	37.1	99.5	8.9	99.8	1.2	97.2

See notes to table table 2.

have only one common trend. As argued by Espasa and Mayo-Burgos (2013) this is a sensible assumption when dealing with all the components of a macro variable. In fact, they show that the pairwise procedure leads to more accurate forecasts of different CPIs than 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 *sectoral* 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 sectoral one) we will not discover the 'true' data structure. Thus our approach needs to be modified.

Provided that there is a subset of series that have just the *general* trend, we can proceed as follows: i) Apply the pairwise procedure described above. This will lead to discover the subset of series that have only the *general* trend — call it n_1. ii) Test for cointegration in all the triplets formed by one series inside $\hat{n}_1$ and a pair of outsiders. For the triplets that contain pairs of the same *sector* we will find one cointegration relationship — two common trends. iii) Construct an *adjacency matrix* for the series outside $\hat{n}_1$. In each cell of this square and symmetric matrix put a 1 if all the $\hat{n}_1$ triplets that contain the corresponding pair have just one cointegration relationship, otherwise put a 0. iv) Look for maximal cliques in the previous *adjacency matrix*. This will lead to discover the series in each sector.

Remark 5 *By theorem 1, in point iii above it would be irrelevant if we test cointegration with all the series in $\hat{n}_1$, with some of them, or with the estimated common trend of $\hat{n}_1$. When dealing*

with small samples, requiring to find one cointegration in all the $\hat{n}_1$ triplets may be too stringent. Instead, we could relax this requirement by allowing a few triplets to fail in showing that result. In the simulations below we use the stringent requirement.

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

3.4.1 Simulation results for the case of *'general'* and *'sectoral'* trends

As argued above, the generalization for the case of *general* and *sectoral* 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 $n_1 = 15$ (now n_1 is the subset of series that have only the *general* trend). Assume also that $\hat{n}_1 = 15$. After testing cointegration in all the 4950 pairs, the procedure requires making other $15 \times 85(85-1)/2 = 53550$ cointegration tests. As highlighted in remark 5, this issue could be mitigated by testing cointegration only with the estimated common trend of $\hat{n}_1$, so that the additional tests in previous example would be only 3570. We do not explore this possibility.

Since the objective of this section is just having a *robustness check* for our procedure, not producing new results, we drastically simplify the simulation design. We consider only one *type* of DGP with two scenarios. The DGP is $DGP1$ described in §3.1, modified to have *general* and *sectoral* trends. Again, the number of replications is 500.

Let n_1 be the number of variables that have only the general trend, and s_i the number of variables that in addition to the general, have the trend of the i^{th} sector. Using the same normalization for matrix β as in $DGP1$, without loss of generality, we normalize all cointegration relationships with respect to one of the variables in n_1. To get a simple visual example of $\beta's$ structure, assume $N = 10$, $n_1 = 3$, $s_1 = 3$, $s_2 = 3$, and that the remaining series has its own trend. In this case we can set β such that:

$$\beta' = \begin{pmatrix} \beta_{11} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{21} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{31} & 0 & 0 & \beta_{34} & 1 & 0 & 0 & 0 & 0 & 0 \\ \beta_{41} & 0 & 0 & \beta_{44} & 0 & 1 & 0 & 0 & 0 & 0 \\ \beta_{51} & 0 & 0 & 0 & 0 & 0 & \beta_{57} & 1 & 0 & 0 \\ \beta_{61} & 0 & 0 & 0 & 0 & 0 & \beta_{67} & 0 & 1 & 0 \end{pmatrix} \tag{8}$$

An important difference with respect to $DGP1$ is that we cannot set the coefficients β_{ij} equal to -1 because the series in n_1 would be cointegrated with all 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 $[-5, -0.1]$. Matrices α and Φ_1 are the same as in $DGP1$.

We consider two scenarios just for $T = 400$. In both of them 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 thirty. Thus, in this second scenario $N = 60$.

Table 4 includes the *gauge* and *potency* of the modified pairwise procedure for discovering the *general* and the *sectoral* trends. Figures under 'Sectors' columns are averages for the two sectors. As the table shows, the procedure has high potency for discovering the true series in each sector with little costs in terms of gauge.

Table 4: Gauge and potency of the 'pairwise' procedure for the case of *general* and *sectoral* trends

	Scenario 1		Scenario 2	
	n_1	*Sectors*	n_1	*Sectors*
Gauge	98.6	90.9	98.3	87.3
Potency	2.0	1.1	1.3	0.9

- $Gauge = \frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$. - $Pot = \frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$. - 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 1: $N = 60$, $n_1 = 10$, $s_1 = s_2 = 10$. - Figures in 'Sectors' columns are averages for the two sectors.

4 Empirical application

In this section we apply the pairwise procedure with outliers correction to the US CPI. The absence of economic theory linking disaggregated prices in the long run could make the concept of *cointegration* to sound inadequate for this application. However, this observation does not preclude the existence of linear combinations between the CPI components that cancel unit roots and could be useful to obtain better forecasting results. The absence of theory only implies that these relationships may not be expected to be *permanent* as, for example, is the relationship between income and consumption. For this reason, in this section we substitute the concept of cointegration by *common unit roots restrictions*[4].

4.1 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 $1999.1 - 2014.12$ (192 observations). The total number of components is 181. After dropping those with less than 162 valid observations we keep 172 basic components. From these series we exclude eight that evolve by steps (regulated prices) so that we end up with 164 series which, considering the weights of 2014, represent 91% of the CPI [5].

4.2 Results of the pairwise tests

Since the pairwise approach does not deal with seasonal unit roots, we performed previous OSCB (see Osborn et al. (1988)) tests to all the components. Results indicate that they do not show seasonal unit roots in general and that the assumption of only one regular unit root and linear growth seems sensible (details are available upon request).

We use the outliers correction procedure detailed in Carlomagno and Espasa (2015), perform Johansen's tests at the 5% of significance, and select the number of lags for each pair the *AIC* in a model without trend in the 'long run' and one common unit root restriction. Centered seasonal dummies are included in all models.

In order to obtain economically and statistically sensible cointegration restrictions we propose to consider only those that satisfy the following conditions: i) the 'long-run' relationship does not require a deterministic trend; ii) coefficients of both components are statistically significant;

[4] We are grateful to David Hendry for this observation.

[5] The eight excluded series are: College tuition and fees, Elementary and high school tuition and fees, Child care and nursery school, Postage, Delivery services, Wireless telephone services, Food at employee sites and schools, and Housing at school excluding board.

iii) the second largest root of the bivariate VAR's characteristic polynomial is not close to one; and iv) the 'long-run' relationship is stable over time.

As argued by Carlomagno and Espasa (2015), the pairwise approach can be extended with exogenous variables in the bi-variate models; we include the Real Effective Exchange Rate ($REER$).

We found ten fully cointegrated subsets, with a total of 42 series that represent 11.2% of the CPI's weight (full details of the composition of these subsets are available upon request).

4.3 Forecasting all the components of the US CPI

Using the results of the pairwise tests we construct single equation models that include the cointegration restrictions derived from the fully cointegrated subsets. The final forecasting models are selected using the model selection algorithm *Autometrics*. Starting from a General Unrestricted Model (GUM) and using a multiple path search, this algorithm reduces the GUM to a simpler model that encompasses it and passes a battery of diagnostic tests. We also use Impulse Indicator Saturation (IIS) for correcting data irregularities (see Santos et al. (2008)).

Although we consider more than one possible GUM, a representative example for a series that belongs to a 'fully cointegrated' subset of size m, would be:

$$\Delta x_t = c + \sum_{i=1}^{m-1} \alpha_i CR_{i,t-1} + \sum_{i=1}^{11} \gamma_i S_{it} + \sum_{i=1}^{K} \phi_i \Delta x_{t-i} + \epsilon_t,$$

where S_{it} represents the i^{th} centered seasonal dummy, and $CR_{i,t}$ a cointegration relationship.

Since for each component, the process of building the econometric model is subject to a set of diagnostic tests included *Autometrics*, we can conclude that they are reasonable for empirical applications. Additionally, as the basic components aggregate to the CPI, we can apply another test to the models for the disaggregates. It consists of comparing the forecast of the aggregate obtained by aggregating the forecasts of the components with the forecasts from a scalar model for the aggregate. We denote the indirect approach by *I-PW* (PW stands for *pairwise*) and the direct one by *D*. The latter is our baseline model.

In the pairwise approach, series with no common unit roots restrictions can be modeled individually or using the methodology proposed by Guerrero and Peña (2003), which allows to obtain individual forecasts consistent with the forecast of the sub-aggregate obtained with a scalar model. We apply this procedure for the series with no common unit root restrictions, and denote this alternative as $I - PW - GP$ (for Guerrero and Peña).

Given our interest in forecasting all the components, the comparison with direct approaches

should not be used as a definitive criterion for assessing the forecasting performance of our procedure, we should use some disaggregated baseline. Therefore, we also compare the forecasting performance of *I-PW* with the disaggregated forecasts using univariate models for each basic component, denoted as I.

As other forecasting possibilities, we enlarge the baseline and the basic indirect approaches with dynamic factors extracted from all the components, and denote those procedures as $B - DFM$ and $I - DFM$ respectively.

Table 5 includes a summary of the forecasting results. The first row includes the root mean squared forecast error in the direct baseline for $\Delta_{12} log(CPI)$ for horizons $H = 1$ to $H = 12$. All the other entries in the table are ratios with respect to the baseline. The evaluation period is $2011.1 - 2014.12$. This evaluation period may seem rather short, specially when analyzing medium term RMSFEs. For example, for $H = 12$ we have thirty six forecasts, but only three of them are independent. Therefore, results for horizons larger than four or five steps ahead must be interpreted with extreme caution.

The selection of the sample size was related to the availability of disaggregated data. We could have chosen a smaller number of disaggregates, extend our sample from the beginning, and enlarge the evaluation period of the forecasting exercise. As discussed in the introduction, our procedure requires working with high levels of disaggregation since intermediate sub-aggregates may add up components that do not share common unit roots, precluding the formation of the subsets we are looking for. In the trade-off between more disaggregation vs. larger evaluation period, we gave more importance to the former. Although the maximum publicly available disaggregation level — the one we are considering — still contains aggregated series, this is the best we can do.

Forecasting scalar models enlarged with Dynamic Factors (D-DFM)

The best results are obtained with only one factor. The inclusion of the factor produces some minor gains over the baseline in all horizons except in 7 to 9, where results are indistinguishable.

Forecasting results for indirect approaches

As a brief summary; we found that the baseline option is hard to beat in short horizons (1-5). Although when adding a dynamic factor, *D-DFM*, there are some improvements in the forecasting accuracy, they are only minor with ratios between 0.99 and 0.96.

Univariate models for all the components (I-B)

This is the simplest disaggregated approach. In short horizons (1-6) it clearly deteriorates with respect to the baseline, as the RMSFE are around 8% larger than in the baseline. From horizons 7 to 12 this approach considerably outperforms the baseline. For $H = 12$ the RMSFE is 60% of the one in the baseline.

Univariate models for all components enlarged with Dynamic Factors (I-DFM)

The best results are obtained with two factors. The advantage of the indirect forecast completely disappears if univariate models are enlarged with Dynamic Factors, and the bad performance for short horizons is even worst. In this application the use of DFM to forecast the components is not useful at all, univariate models do much better.

Pairwise procedures

In short horizons $I - PW - GP$ provides slightly better forecast for the aggregated CPI than the corresponding case in which basic components outside subsets with common trends are forecast individually. For longer forecast the conclusion is the opposite.

The results using models which include common unit roots restrictions are better than in the basic indirect procedure $(I - B)$. In particular the approach $I - PW - GP$ clearly dominates $I - B$ in the first four horizons[6]. In long horizons, when $I - B$ clearly dominates the baseline, the inclusion of common unit roots delivers some further improvements.

As a conclusion, this exercise shows that the modeling of the basic components by single-equation methods taking into account the common unit roots restrictions between them, identified by pairwise methods, can be considered as adequate in the sense that the forecasts for the aggregate are quite good. This is an interesting indirect test of the common-feature disaggregated approach.

[6]Recall that the total weight of the series in some set with common unit roots is only 11%.

Table 5: Relative RMSE $\Delta_{12} log(CPI)$. Evaluation period: 2011.1-2014.12 (first row: RMSE 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
B	**0.22**	**0.39**	**0.50**	**0.52**	**0.56**	**0.59**	**0.67**	**0.74**	**0.80**	**0.83**	**0.88**	**0.96**
B-DFM	0.98	0.97	0.96	0.97	0.98	0.99	1.01	1.00	1.00	0.98	0.96	0.97
I	1.07	1.05	1.08	1.12	1.10	1.05	0.96	0.89	0.78	0.66	0.62	0.61
I-DFM	1.43	1.30	1.24	1.26	1.23	1.22	1.18	1.18	1.16	1.11	1.06	1.00
I-PW	1.07	1.05	1.08	1.12	1.08	1.04	0.95	0.87	0.75	0.65	0.61	0.59
I-PW-GP	1.06	0.99	0.98	0.99	1.01	1.02	0.97	0.92	0.92	0.91	0.90	0.89

- Dark red entrances highlight the loser procedure.
- Light red indicates procedures that are, at most, 5 points smaller than the worst one.
- Dark green indicates the best procedure.
- Light green indicates procedures that are, at most, 5 points larger than the best one.

5 Concluding Remarks

In this paper, we studied the properties of a pairwise procedure for testing cointegration between all the possible pairs of an aggregate's components at the maximum level of disaggregation. This procedure allows to discover blocks of series that share a unique common trend (*fully cointegrated blocks*). The main theoretical result is that cointegration tests inside those blocks are asymptotically equivalent, in the sense that the probability that all tests deliver the same conclusion tends to 1 as T goes to infinity independently of the number of series. Thus, multiple testing is not an issue for pairs of components inside a fully cointegrated block. This result is valid both when N is fixed and when it goes to infinity. We also showed that the pairwise approach can be extended for cases with *general* and *sectoral* trends.

In a Monte Carlo experiment, we confirmed the asymptotic results and compared the performance of the pairwise approach with that of a standard DFM alternative. This comparison showed that the pairwise procedure dominates in situations in which the number of series that share the trend (n_1) is relatively small with respect the total number of components, N. The DFM alternative fails in those situations.

Finally, we found that, in moderately short samples, the pairwise strategy leads to power improvements with respect to a regular Johansen's test applied to a (reduced) group of series that share a common trend. These improvements are remarkable in many situations.

Acknowledgments

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

References

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

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 817–858.

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

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data. *Journal of Econometrics 122*(1), 137–183.

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

Bai, J. and S. Ng (2004). A panic attack on unit roots and cointegration. *Econometrica 72*(4), 1127–1177.

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*.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 289–300.

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

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

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

Carlomagno, G. and A. Espasa (2015). Forecasting a large set of disaggregates with common trends and outliers. *WP, UC3M, Statistics Department*.

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

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

Cheung, Y.-W. and K. S. Lai (1993). Finite-sample sizes of johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics 55*(3), 313–328.

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

Cubadda, G. and E. Scambelloni (2015). Index-augmented autoregressive models: Representation, estimation, and forecasting. *Working Paper*.

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

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

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

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

Giacomini, R. and C. Granger (2004). Aggregation of space-time processes. *Journal of Econometrics 118*(1), 7–26.

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.

Harbo, I., S. Johansen, B. Nielsen, and A. Rahbek (1998). Asymptotic inference on cointegrating rank in partial systems. *Journal of Business & Economic Statistics 16*(4), 388–399.

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. (1992). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics 52*(3), 389–402.

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

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

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

Lütkepohl, H. (1984). Linear transformations of vector arma processes. *Journal of Econometrics 26*(3), 283–293.

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.

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

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

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, H. (2007). A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics 138*(1), 312–355.

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.

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

Appendix A Relationship between pairs' test statistics under the null of 'full cointegration'

The first step in Johansen's procedure is to concentrate the model with respect to $\alpha\beta'$, what is done by regressing ΔY_t and Y_{t-1} on $(\Delta Y_{t-1}, ..., \Delta Y_{t-k+1})$. These auxiliary regressions give the residuals R_{0t} and R_{1t}, respectively, and the matrices S_{ij} are defined as $T^{-1}R_iR_j'$, where R_i is a $n \times T$ matrix. For $n = 2$, 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:

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

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

Let X_t be the vector containing the series in n_1. From the Granger Representation Theorem, disregarding deterministic terms, the cointegrated VAR can be written as:

$$X_t = X_0 + C(1)\sum_{i=1}^{t}\epsilon_t + C^*(L)\epsilon_t,$$

where $C(1) = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp'$, has rank $n_1 - r$, and $C^*(L)$ is a stationary lag polynomial matrix. Under full cointegration, $r = n_1 - 1$ and the rank of $C(1)$ is 1. Therefore, individual series in n_1 can be written as:

$$X_{mt} = \delta_m CT_t + w_{mt}; \quad m = 1, ..., n_1, \tag{A.2}$$

where CT_t is a $I(1)$ process and w_{mt} are stationary ones. From eq. (A.2), any series in n_1 can be expressed as $X_{mt} = \gamma_{mq}X_{qt} + \eta_{mr,t}$, with $\eta_{mt} \sim I(0)$, and X_{qt} being other series in n_1.

Assume that the aforementioned matrices S_{ij} refer to the vector , $Y_t = [X_{1t}, X_{2t}]'$. We now derive the test statistic for any other pair in n_1 given the one for Y_t. Let

$$\begin{aligned} X_{it} &= \gamma_{i1}X_{1t} + \eta_{i1,t}, \\ X_{jt} &= \gamma_{j2}X_{2t} + \eta_{j2,t}, \end{aligned} \tag{A.3}$$

call $Y_t^* = [X_{it}, X_{jt}]'$ and assume, without loss of generality, that $p^* \geq p$ (where p^* is the lag

length of the model for Y_t^*, and p is the lag length of the model for Y_t). Writing the auxiliary regressions for ΔY_t^* and Y_{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:

$$S_{ij}^* = T^{-1}(\Gamma R_i + \epsilon_i)(\Gamma R_j + \epsilon_j)', \tag{A.4}$$

where $\Gamma = \begin{bmatrix} \gamma_{i1} & 0 \\ 0 & \gamma_{i1} \end{bmatrix}$, and ϵ_i and ϵ_j are stationary processes. Then;

$$S_{11}^* = T^{-1}[\Gamma R_1 R_1' \Gamma' + \Gamma R_1 \epsilon_1' + \epsilon_1 R_1' \Gamma' + \epsilon_1 \epsilon_1'] \tag{A.5}$$

In eq. (A.5), all terms inside the brackets are $Op(T)$ except for $\Gamma R_1 R_1' \Gamma'$, which is $Op(T^2)$. Thus, S_{11}^* is $Op(T)$ and its long-run behavior is dominated by $\Gamma R_1 R_1' \Gamma'$. That is, $S_{11}^* \to T^{-1}(\Gamma R_1 R_1' \Gamma')$ as $T \to \infty$. The remaining S_{ij}^* are $Op(1)$ and can be written as

$$S_{ij}^* = \Gamma S_{ij} \Gamma + \Omega_{ij}, \text{ for } (i,j) \neq (1,1), \tag{A.6}$$

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

The new eigenvalue problem is: $(S_{10}^* S_{00}^{*-1} S_{01}^*) v^* = \lambda^* S_{11}^* v^*$. Using eq. (A.5) and eq. (A.6), we get:

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

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:

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

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

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

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

Comparing eq. (A.10) with eq. (A.1), we can make three considerations:

i) If $X_i \equiv X_1$ and $X_j \equiv X_2$, we get $\Psi = 0$, $c = 1$ and $\Gamma = I$, so we recover the original problem. ii) In the extremely unlikely case that $\Omega_{ij} = 0$ — for $(i,j) \neq (1,1)$ —, we get $\Psi = 0$ 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 the one for the pair (X_1, X_2). iii) In the general case that $\Omega_{ij} \neq 0$, we will have $\Psi \neq 0$ and $c \neq 1$. Note that the eigenvalues of the problem eq. (A.10) are the solutions of the second-order polynomial in λ^* $|\lambda^* S_{11} - (cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi})| = 0$.

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, and $G \to 0$.

Now, replace θ_{ij} by θ_{ij}^* in eq. (A.11) to get B^*, C^* and G^*. Since θ_{ij} and θ_{ij}^* are $Op(1)$, the asymptotic behavior of G^* is the same as that of G, for 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 Proof of proposition 1

When Q is fixed, the proof follows directly from the proof of theorem 1. For the case of $Q \to \infty$ we need a little more elaboration. Consider first our argument of the asymptotic power

for ruling out wrong not rejections of $r = 0$. Define Z_i as the random variable that takes the value 1 if the null of $r = 0$ is wrongly not rejected for the ih pair among the Q series, and zero otherwise. The expected proportion of not rejections $Q^*E(Z_i)/Q^*$ where Q^* is the number of pairs between the Q series. Since asymptotic power of Johansen's test is 1, $E(Z_i)$ and the expected proportion of not rejections converge in probability to 0 as $T \to \infty$.

Consider now the test $r = 1$ vs $r = 2$. Choose a pair among the Q series (pair 0), and define W_i as the random variable that takes the value 1 if the result of the test for the i^{th} pair is different form that of pair 0. Using theorem 1 and the same argument as above, it follows that the expected proportion of tests results which are different form that of pair 0 goes to zero as T goes to infinity. ■

Appendix C Monte Carlo evidence for false rejections of $r = 0$ when none of the series belongs to the fully cointegrated subsets

We generate $N + 1$ random walks, pick one of them and perform Johansen's cointegration tests between the selected series and the remaining N. Call these tests $-2lnQ(H^*(0)/H^*(1))$. We replicate this experiment 1000 times. The DGP is:

$$Y_{t_{(N+1)\times 1}} = Y_{t-1} + e_t, \tag{C.1}$$

with $e_t \sim N(0, \Sigma)$. We set $N = 1000$, and the sample size is $T = 400$. The structure of Σ is not relevant. We consider $\Sigma = I$, and $\Sigma = 0.95(\mathbb{1} - I) + I$, where $\mathbb{1}$ is a $(N + 1) \times (N + 1)$ matrix full of ones. This second option is a matrix with ones in the main diagonal and 0.95 elsewhere.

If the N tests statistics of each replica where independent, they should be follow the Johansen's distribution. To assess if this is the case, for each replica, we compute the cumulative probability at the Johansen's quantiles and take the mean and the median across experiments.

Table C.1: Quantiles comparison. Cumulative probabilities of $-2lnQ(H^*(0)/H^*(1))$ at the Johansen's quantiles.

	50%	**75%**	**80%**	**85%**	**90%**	**95%**	**97.5%**	**99%**
				$\boldsymbol{\Sigma = I}$				
Mean	0.48	0.73	0.79	0.84	0.89	0.94	0.97	0.99
Median	0.53	0.79	0.84	0.89	0.93	0.97	0.98	0.99
				$\boldsymbol{\Sigma = 0.95(\mathbb{1} - I) + I}$				
Mean	0.47	0.72	0.78	0.83	0.88	0.94	0.97	0.99
Median	0.50	0.80	0.85	0.90	0.94	0.98	0.99	0.99

The slight differences between the cumulative probabilities of $-2lnQ(H^*(0)/H^*(1))$ the those of the Johnasen's distribution confirm that the assumption of independence outside all n_j is sensible.