UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 14-13
Statistics and Econometrics Series 09
May 2014

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

THE PAIRWISE APPROACH TO MODEL A LARGE SET OF DISAGGREGATES WITH COMMON TRENDS

Guillermo Carlomagno*, Antoni Espasa**

Abstract

The objective of this paper is to model and forecast all the components of a macro or business variable. Our contribution concerns cases with a large number (hundreds) of components where multivariate approaches are not feasible. We extend in several directions the pairwise approach originally proposed by Espasa and Mayo-Burgos (2013) and study its statistical properties. The pairwise approach consists on performing common features tests between the N(N-1)/2 pairs of series that exist in a group of N of them. Once this is done, groups of series that share common features can be formed. Next, all the components are forecast using single equation models that include the restrictions derived by the common features. In this paper we focus on discovering groups of components that share single common trends. The asymptotic properties of the procedure are studied analytically. Monte Carlo evidence on the small samples performance is provided and a small samples correction procedure designed. A comparison with a DFM alternative is also carried out, and results indicate that the pairwise approach dominates in many empirically relevant situations. A relevant advantage of the pairwise approach is that it does not need common features to be pervasive. A strategy for dealing with outliers and breaks in the context of the pairwise procedure is designed and its properties studied by Monte Carlo. Results indicate that the treatment of these observations may considerably improve the procedure's performance when series are 'contaminated'.

Keywords: *Common trends, Cointegration, Factor Models, Disaggregation, Forecast model selection, Forecast Combination, Outliers treatment.*

* Departamento de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), e-mail: gcarloma@est.econ.uc3m.es
** Departamento de Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), and Instituto flores de lemus, Universidad Carlos III, Madrid, Spain, e-mail: espasa@est.econ.uc3m.es

1 Introduction

This paper focuses on how to model all the components of a macro or business variable when the number of components is large (hundreds). The main objective is to identify and estimate relevant relationships between the components, develop a single-equation econometric modeling procedure which incorporates those relationships and could provide accurate forecasts of the components. At the same time, this strategy should be valuable for a better understanding of the aggregate. Thus, we may end up with a procedure that is valuable for policy and business decision making. This procedure will also provide an indirect forecast of the aggregate which, hopefully, could be more accurate than alternative forecasting methods or, at least, not worst.

There are at least four relevant reasons for disaggregating a macro-variable. One is the knowledge of disaggregated data on themselves. Another, very often more relevant, comes from acknowledging the importance of understanding the micro heterogeneity and the statistical effects of the aggregation process in order to explain the dynamic properties of aggregated variables (see e.g. Bils and Klenow (2004), Lunnemann and Mathä (2004), Imbs et al. (2005), Clark (2006), Altissimo et al. (2007), Boivin et al. (2009), or Beck et al. (2011)). Third, the use of disaggregated information may lead to better forecasts of the aggregate (see Espasa et al. (2002), Giacomini and Granger (2004), Hendry and Hubrich (2005), Hendry and Hubrich (2011), Espasa and Albacete (2007) and Espasa and Mayo-Burgos (2013)). Finally, disaggregation is relevant for comparative analysis of the components, what is clearly important when dealing with prices since relative prices' dynamics may be a key ingredient for decision making.

When working with a large set of disaggregates one of the main challenges in econometric modeling is how to deal with the trade-off between informational losses (due to disregarding some disaggregate information) and estimation uncertainty (due to the increased amount of parameters to be estimated when disaggregates are considered). There does not seem to be an agreement on how to tackle this problem, which in any case is an empirical issue. In this paper we are interested in modeling and forecasting the disaggregates (not only the aggregate), so that a great amount of estimation must be inevitably carried out. An indirect way of testing the quality of this large set of estimated models is to test if forecasting the aggregate by aggregating the forecasts from the individual models of the components, the result is not significantly worse, hopefully better, than the forecasts from alternatives approaches.

Giacomini and Granger (2004) propose a way for dealing with the informational losses vs. estimation uncertainty trade off when variables are spatially correlated. Basically, they show that using a restricted multivariate model may solve the problem by not considering all the disaggregates in every equation, but only the relevant ones. In their case of interest (spatially correlated variables) the restrictions are known beforehand.

When restrictions are unknown, they may be determined by statistical procedures at the cost of introducing an additional source of uncertainty in the model. Depending on the procedure's ability to detect the restrictions, it may be the case that the procedure still outperforms other possible alternatives.

Hendry and Hubrich (2011) stress the relevance of the informational losses vs. estimation uncertainty trade-off by showing that estimation uncertainty is one of the main causes of the forecast error difference between direct and indirect methods. The authors develop a methodology for imposing the restrictions when they are unknown. The cornerstone of the proposal is that including disaggregated information in the model for the aggregate cannot lower but may improve the predictability of the aggregate (see also Hendry and Hubrich (2005)). Basically, their proposal consists on including some (selected) components

in a single-equation model for the aggregate, that is; making use disaggregated information instead of using disaggregated forecasts.

Interestingly, it can be shown that the conditions under which all components' coefficients in the equation for the aggregate are equal to zero (the direct univariate method is optimal) are exactly the same as the Conditions for the Efficiency of the Direct Forecast established by Kohn (1982). Since Hendry and Hubrich's proposal only requires some of those coefficients being different from zero, in this sense, it may be seen as an intermediate alternative between the direct univariate and the indirect multivariate approaches.

Another alternative for dealing with the informational-loses vs. estimation uncertainty trade off is the consideration of common features as proposed by Espasa and Mayo-Burgos (2013). The procedure consist on trying to discover blocks of components that share unique common features (trends, cycles, breaks), and then include the restrictions implied by those commonalities in single equations models to forecast all the components. Specifically, the search of those blocks is carried out by performing common features tests between all possible pairs of components (which are $N(N-1)/2$). Note that the level of disaggregation used in this procedure must be the maximum possible since *ad-hoc* sub-aggregates may add up series that do not share common features.

This strategy may have three advantages with respect to component selection proposed by Hendry and Hubrich (2011). First, restrictions implied by the existence of common features are explicitly tested, from where structural economic interpretations may be derived, putting the economic theory closer to the forecasting procedure. Besides, the pairwise approach takes into account possible cointegration relationships which, as shown by Clark (2000), should not be neglected in forecasting exercises. Third, this procedure allows modeling and forecasting all the disaggregates (not only the aggregate), generating valuable information for economic and business decision making. Finally, Espasa and Mayo-Burgos's strategy does not exclude the component selection approach of Hendry and Hubrich (2011). After specifying the common features, component selection can be used to see which common features and stationary-transformed disaggregates are relevant in each component's equation.

The problem of how to impose unknown restrictions in multivariate models is also present in the Dynamic Factors literature. Boivin and Ng (2006) find that if the data contain non-pervasive factors (factors that are common only to a reduced subset of series)[1] the choice of the data from which factors are extracted is not innocuous; results are more accurate when factors are extracted from data which is informative about them. In the same line, Espasa and Mayo-Burgos (2013) find that the forecasting performance of DFM is improved if the factor is estimated from series which have it. As a third example, Beck et al. (2011), working in the context of price setting, find that if the loading matrix has a block structure, factors are more precisely estimated if this structure is considered in the estimation process.

Boivin and Ng (2006) propose an ad-hoc procedure which consists of pre-grouping the series in broad categories and extract the factors form those groups. Results in Espasa and Mayo-Burgos (2013) suggest that this strategy will hardly work for disaggregated prices because common features are not distributed by 'broad categories'. Beck et al. (2011) do not need to estimate the block structure because they assume they know it, but this may not be a realistic assumption in many cases.

Note that the issues about non-pervasive common features have not only theoretical interest; a proper

[1]Strictly speaking, the condition for the factors to be *pervasive* is that $\Lambda'\Lambda/N \to \Sigma_\Lambda$ as $N \to \infty$, where Λ is the loading matrix and Σ_Λ is a positive definite matrix. This condition requires that the factors must have infinitely many non zero loadings as $N \to \infty$.

estimation of the factors may change dramatically the conclusions about the estimated dynamic properties of aggregate and sectoral prices, and therefore the implications for economic policy and business decisions. In fact, from their empirical application to European inflation Beck et al. (2011) find that previous conclusions on sectoral components' properties should be modified. Specifically, they find the sectoral component to be less volatile than previously estimated, to explain much lees variance of inflation series, and to show little persistence. Then, sector specific components are not the main driving force of inflation as previous studies suggested. Country and regional factors are the main drivers of inflation.

One aim of this paper is to show that the pairwise strategy initially proposed by Espasa and Mayo-Burgos (2013) can be used as an objective method to determine the block structure. That is, blocks can be constructed by looking for common features between all the $N(N-1)/2$ pairs of series and then grouping those that share a unique common feature. For that purpose, the knowledge of the statistical properties of the testing procedure is crucial, and their study by analytical methods and simulation experiments is one of the main contributions of this paper.

Restricting our attention to common trends, we show that the probability of finding cointegration between all possible pairs in a group of series that share a unique common trend does not depend on the number of series and tends to $(1-\alpha)$ as T goes to infinity, where α is the nominal size used in the Johansen's trace test. This result is confirmed by Monte Carlo. In the simulation experiments we also study the procedure's small samples properties and, given the deterioration it suffers as T decreases, a small samples correction is proposed and studied by Monte Carlo. In the Monte Carlo experiments we also compare the performance of the pairwise approach with an alternative based on Dynamic Factor Models.

The other contributions in this paper are related to the specification of the bi-variate models for the $N(N-1)/2$ pairs, to the specification of the individual forecasting equations and to the treatment of outlying observations. Concerning the bi-variate models, we analyze the importance of a correct specification of their dynamic structure for different 'types of pairs' and conclude that it may be critical for cointegration tests results and for the performance of pairwise strategy. Furthermore, we find that the result in Johansen (1995) and Lütkepohl et al. (2003) about the negative dependence of the Johansen's test's power with respect to the number of common trends in the system, derives -for relatively short samples- in a power improvement of the pairwise strategy with respect to the regular Johansen's test when applied to a -reduced- group of series that share a single common trend.

In relation with the forecasting equations we study how different normalizations of the same cointegration relationships, by implying different α adjustment matrices, may help to improve the forecasting accuracy of our procedure. In particular, we study two 'extreme' alternative normalizations with simple structural interpretations and conclude that using the 'correct' one may lead to a reduction in the number of regressors in the individual equations without cost in terms of informational losses. Thus, estimation uncertainty reductions and forecasting accuracy improvements may come up from a simple change in the normalization of the cointegration relationships, what can be exploited in applied work.

One of the two 'extreme' normalizations we study is expressing all the $n_1 - 1$ cointegration relationships that exist in a group of n_1 components which share a single stochastic trend as deviations of $n_1 - 1$ series with respect to the sub-aggregate formed by all of them -nomalization *(b)*-. We argue that the strategy in Espasa and Mayo-Burgos (2013) of including in each forecasting equation only the cointegration relationship of the dependent variable with the sub-aggregate is correct only in the special

case that the matrix α in a model with normalization *(b)* has a diagonal structure. In more general cases, the Espasa and Mayo-Burgos (2013) approach omits relevant long-run information. We generalize the forecasting strategy by considering all cointegration relationships as potential regressors for all the individual equations. This extension solves the problem.

Regarding the problem of outliers, although there is plenty of evidence in the literature that outlying observations may seriously distort estimates and inferential conclusions, due to the difficulties associated to their treatment, in many empirical applications this problem is neglected. We provide a review of the alternatives for dealing with outliers and breaks in cointegrated systems and propose a strategy that combines the IIS methodology (see Santos et al. (2008)) with the feasible GLS procedure proposed by Saikkonen and Lütkepohl (2000). Shortly the outlier's dates are first estimated in single equations using IIS and then, these dates are used in the GLS procedure. The performance of this proposal is studied by Monte Carlo experiments, which highlight the importance of correcting for outliers and breaks when data are 'contained'.

A related approach to the pairwise procedure is the proposed by Pesaran (2007), who develops a strategy for testing output and growth convergence across countries. Output convergence between two countries implies that log GDPs are cointegrated with cointegrating vector $[1, -1]$ and without trend in the cointegration relationship. This can be tested by performing unit root tests on the log difference of the two GDPs and checking the significance of the deterministic trend. For a group of N countries the testing strategy consists on performing unit root tests for all the $N(N-1)/2$ differences between pairs of GDPs. Pesaran (2007) shows that for a given group of N countries, under the null of convergence, the fraction of false unit root conclusions tends to α as N and T go to infinity (where α is the nominal size used in the unit root tests). This is an important result since it holds even though the $N(N-1)/2$ tests are not independent between each other.

There is however an important difference between Pesaran (2007)'s objective and ours. Pesaran (2007) is interested in testing the *universal* null hypothesis that (almost) all GDP differences in the data set are stationary without trend. In contrast, our objective is discovering the groups of series that share a common stochastic trend. Another difference is that we do not want to restrict the cointegration vectors to be $[1 - 1]$, so that the use of log differences will not work in our case. Even though the result about the proportion of wrong unit root conclusions can be extended to more general types of cointegration, it would not be relevant for our purposes since it does not allow to form the groups we are looking for. Indeed, this potential extension could be useful for testing whether or not a group of series share a single trend, but if the series a grouped by sub-sets with a single trend, we would not be able to identify those sub-sets.

The rest of the paper is organized as follows. In section 2 we study the statistical foundations of the pairwise procedure, including a justification for the forecasting strategy. Next, in section 3 we perform a Monte Carlo experiment to confirm the results of previous section, analyze the small sample properties of the pairwise approach and compare its performance with other alternatives based on Dynamic Factor Models. In section 4 we give a brief review on alternative methodologies for dealing with outliers and breaks in cointegrated systems and propose an outlier correction procedure for the pairwise strategy. The performance of this procedure is analyzed in section 5 in a Monte Carlo experiment. Before the concluding remarks, in section 6 we give the detailed steps for the pairwise procedure, and finally section 7 is devoted to the conclusions.

2 Some details about the pairwise procedure

The pairwise strategy for discovering blocks of components that share single common features consists on performing common features tests between all possible pairs of components and use those results to construct subsets in which all components share a unique common feature. For the case of common trends, the procedure 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, as we argue below, the lag length determined in each case. Then, the procedure requires constructing subsets in which every series is cointegrated with all the others. We call these subsets as *Fully Cointegrated* sets (the detailed algorith is describe in 6).

Next, in the forecasting stage, a single equation model for each component is estimated including as regressors the cointegration relationships found in previous step.

In section 2.1 we give some details about the estimation and testing procedures in the bi-variate sub-models and derive a strategy for the forecasting stage. Next, in section 2.2 we analyze the statistical properties of the procedure.

2.1 On the specification of the models for the pairs and the forecasting equations

To illustrate our arguments about the estimation and testing process of the bi-variate sub-models (section 2.1.2) and about the forecasting equations (section 2.1.3) we make use of two general DGPs, which are also used in the simulation experiments of section 3. Thus, we devote section 2.1.1 to describe these DGPs in detail.

The general framework for the models we work with is given by a VAR model whose series are assumed to be at most $I(1)$:

$$X_t = \mu_t + \Pi_1 X_{t-1} + ... + \Pi_k X_{t-k} + \epsilon_t \Rightarrow (I_n - \Pi_1 L - ...\Pi_k L^k)X_t = \mu + \epsilon_t \Rightarrow \Pi(L)X_t = \mu + \epsilon_t, \quad (2.1)$$

where X_t is a $N \times 1$ vector, Π_i are $(N \times N)$ coefficient matrices, ϵ_t is a white noise vector (its elements are iid), 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:

$$\Delta X_t = \mu_t + \alpha\beta' X_{t-1} + \Phi_1 \Delta X_{t-1} + ... + \Phi_{k-1}\Delta X_{t-k-1} + \epsilon_t, \quad (2.2)$$

where α and β are $N \times r$ matrices, with $0 < r < N$, r being the number of cointegration relationships, $\alpha\beta' = -I_n + \Pi_1 + ... + \Pi_k$, *and* $\Phi_i = -\sum_{j=i+1}^{k} \Pi_j$.

In what follows, for saving notation we set $\mu_t = 0$, but all derivations can be generalized for the case of $\mu_t \neq 0$.

2.1.1 Two alternative DGPs

The cointegrated VAR in expression (2.2) remains unchanged if we change the original matrices α and β' by $\alpha^\circ = \alpha H^{-1}$ and $\beta'^\circ = H\beta'$, with H being a $r \times r$ normalizing matrix. Although these changes will not have any consequences on the system's dynamic properties, they may change and eventually ease its interpretation [2].

[2] This *re-normalization* can be done in the estimation process or after it. To see this, note that in the usual the ML estimator of α is $\hat{\alpha} = S_{01}\hat{\beta}(\hat{\beta}' S_{11}\hat{\beta})^{-1}$. But the usual normalization for β is to assume that $\beta' S_{11}\beta = I$ (conditional

As we argue below, this discussion is relevant for two main reasons, first it helps to decide the structure of the simulated data, and second it has important implications on the specification of the forecasting equations in empirical applications (when sample sizes are finite).

Two normalizations of interest in a system with a unique common trend between a subset (n_1) of the disaggregates of a macro variable may be to express the $n_1 - 1$ cointegration relationships either as deviations with one selected series in n_1 -normalization *(a)*-, or as deviations form the sub-aggregate formed by the series in n_1 -normalization *(b)*-. These two simple normalizations will not change the system but may help to interpret the cointegration relationships form an economic view point.

Note that for the simulation exercises a matrix α has to be selected, what will affect the system's properties.

In this section we consider two 'extreme' models; in both of them we use a 'simple' matrix α and they differ in the normalization used. We use normalization *(a)* for the first model and normalization *(b)* for the second. We derive an equivalence condition for the two models and show that model *(a)* can be transformed to model *(b)* by selecting a proper matrix α, which turns out to be more 'complex' than the original one (and the other way around).

Is this last observation what makes the discussion about the normalization to be relevant for the forecasting stage of the procedure. The fact that a system with a 'complex' matrix α may be transformed to other with a 'simpler' one implies a reduction in the number of variables of the individual forecasting equations without cost in terms of informational losses. We may normalize the system in this two alternative ways and select the cointegration relationships relevant for each individual equation. Proceeding in this way we may get a 'free' reduction in estimation uncertainty, what would translate in an improvement in the forecasting accuracy. Clearly, this will not necessarily be the case but it may be.

The procedure proposed by Espasa and Mayo-Burgos (2013) for the forecasting stage was to include in each individual equation the cointegration relationship of the corresponding component with the sub-aggregate, and no other long-run information. This strategy, though simple, will be correct only in the specific case that the true DGP has the structure of model *b*, in more general situations this approach will usually omit relevant long-run information.

We now turn to the details of the two mentioned models. In both cases we assume the existence of a unique fully cointegrated set of size n_1 ($0 < n_1 < N$), and that the rest $N - n_1$ series are not cointegrated with any other. Thus, while the whole system contains $N - n_1 + 1$ unit roots, the sub-model formed by the series in n_1 has only one. We will use n_1 both, to denote the number of series in the fully cointegrated set and as the name of the set.

i. **Model (a):** In this model cointegration relationships are expressed as deviations of each component in n_1 with respect to one of them (say, x_1 with $x_1 \in n_1$). This can always be done because any fully cointegrated system (a system like (2.2) with $r = N - 1$) can be normalized such that $\beta' = \tilde{\boldsymbol{\beta}}'_{r\times(N-r)}, I_r$ and the new model will be $\Delta X_t = \mu_t + \tilde{\alpha}[\tilde{\beta}', I_r]X_{t-1} + \Phi\Delta X_{t-1} + ... + \Phi_{k-1}\Delta X_{t-k-1} + \epsilon_t$ with $\tilde{\alpha} = \alpha\beta_c'$, and β_c' being the $r \times (N-1)$ matrix that excludes the first column of β'. To see this, assume without loss of generality, $N = n_1$, $k = 2$ and $\mu_t = 0$ (these assumptions are just to save notation and have no consequences on the argument) so that the system is $\Delta X_t = \alpha\beta' X_{t-1} + \Phi\Delta X_{t-1}$

independence). Thus, $\hat{\alpha} = S_{01}\hat{\beta}$. Now if we re-normalize β such that $\beta'^{\circ} = H\beta'$, then $\hat{\alpha}^{\circ} = S_{01}\hat{\beta}H'(H\hat{\beta}'S_{11}\hat{\beta}H')^{-1} = S_{01}\hat{\beta}H'(HIH')^{-1} = S_{01}\hat{\beta}H^{-1} = \hat{\alpha}H^{-1}$.

where β' is a $r \times N$ full row rank matrix and $r = N - 1$. To obtain the normalization $\beta' = [\tilde{\boldsymbol{\beta}}'_{r\times(N-r)}, I_r]$ write;

$$H_{r\times r}\beta'_{r\times N} = [\tilde{\boldsymbol{\beta}}'_{r\times(N-r)}, I_r]$$

Call now β'_c the $r \times (N-1)$ matrix that excludes the first column of matrix β'. Then,

$$H\beta'_c = I_r,$$

so that,

$$H = \beta_c'^{-1} \tag{2.3}$$

and the new model will be $\Delta X_t = \tilde{\alpha}[\tilde{\beta}', I_r]X_{t-1} + \Phi\Delta X_{t-1}$, with

$$\tilde{\alpha} = \alpha H^{-1} \tag{2.4}$$

Note that this argument does not require $r = N - 1$; it is valid for any $(0 < r < N)$.

Thus, if the system has n_1 series that share one common trend and $N - n_1$ that has their own trends, without loss of generality we set matrix β such that:

$$\tilde{\beta}' = \begin{pmatrix} \tilde{\beta}_2 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \tilde{\beta}_3 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \tilde{\beta}_4 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & & & & & \\ \tilde{\beta}_N & 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}_{r\times N} = \begin{pmatrix} \tilde{\beta}'^L & \tilde{\beta}'^R \end{pmatrix},$$

where sub-matrix $\tilde{\beta}'^L$ has dimensions $r \times n_1$, $\tilde{\beta}_i < 0$ for $i = 2, .., N$, and $\tilde{\beta}'^R$ is a $r \times N - n_1$ matrix of zeros. Hence, $\tilde{\beta}'^L$ has the following structure:

$$\tilde{\beta}'^L = \begin{pmatrix} \tilde{\beta}_2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \tilde{\beta}_3 & 0 & 1 & 0 & 0 & \cdots & 0 \\ \tilde{\beta}_4 & 0 & 0 & 1 & 0 & \cdots & 0 \\ \cdots & & & & & & \\ \tilde{\beta}_N & 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix}_{r\times n_1}$$

This normalization was also suggested by Clements and Hendry (1995). Different normalization 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.

For the sake of simplicity, matrix $\tilde{\alpha}$ is set to have the following structure:

$$\tilde{\alpha} = \begin{pmatrix} 0 & 0 & \cdots & 0 & \\ -\tilde{\alpha}_2 & 0 & 0 & \cdots & 0 \\ 0 & -\tilde{\alpha}_3 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & -\tilde{\alpha}_{n_1} \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{N \times r} = \begin{pmatrix} \tilde{\alpha}^U \\ \tilde{\alpha}^D \end{pmatrix},$$

where sub-matrix $\tilde{\alpha}^U$ is $n_1 \times r$, sub matrix $\tilde{\alpha}^D$ is a matrix of zeros with dimensions $(N - n_1) \times r$, and the values $\tilde{\alpha}_i$ are taken from the uniform distribution with parameters $[0.15, 0.3]$ -these parameters are motivated by results in EM for CPI series-. Hence, $\tilde{\alpha}^U$ has the following structure:

$$\tilde{\alpha}^U = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\tilde{\alpha}_2 & 0 & 0 & \cdots & 0 \\ 0 & -\tilde{\alpha}_3 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & -\tilde{\alpha}_{n_1} \end{pmatrix}_{n_1 \times r} \tag{2.5}$$

In contrast to the election of $\tilde{\beta}$, the choice of $\tilde{\alpha}$ 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. [3]

With these structures for the long run matrices, and setting the $r \times 1$ vector $\tilde{\boldsymbol{\beta}}$ equal to $[-1, ..., -1]'$, series in n_1 follow a process similar to the one specified by Banerjee and Marcellino (2009) in their *DGP1*, with two differences[4]. First we do not assume that the $\tilde{\alpha}_i$s are equal, and second we add some short-run structure ($\Phi \Delta X_{t-1}$).

The objective of including the lagged difference in the model is to avoid the presence of common cycles. Given that $\tilde{\alpha}$ has reduced rank there exists a matrix $\tilde{\alpha}'_\perp$ such that $\tilde{\alpha}'_\perp \tilde{\alpha} = 0$, so that not including the term $\Phi \Delta X_{t-1}$ would lead to $\tilde{\alpha}'_\perp \Delta X_t = \tilde{\alpha}'_\perp \epsilon_t$ which implies a common cycle structure (the fact that cointegrated VAR(1) processes always present a common cycle structure was firstly highlighted by Vahid and Engle (1993)). Including the full rank matrix Φ avoids this issue. For simplicity, it is a diagonal matrix whose diagonal elements are drawn form the uniform distribution with parameters [0.5, 0.8].

ii. **Model (b):** In this model the cointegration relationships are expressed in terms of deviations form the sub-aggregate formed by the components in n_1 (see Appendix A.1 for a proof that all the series in n_1 are cointegrated with this sub-aggregate).

As aforementioned, the choice of matrix α made for model *(a)* does affect the long-run adjustment process of the series. The assumptions that each variable reacts only to one cointegration relationship

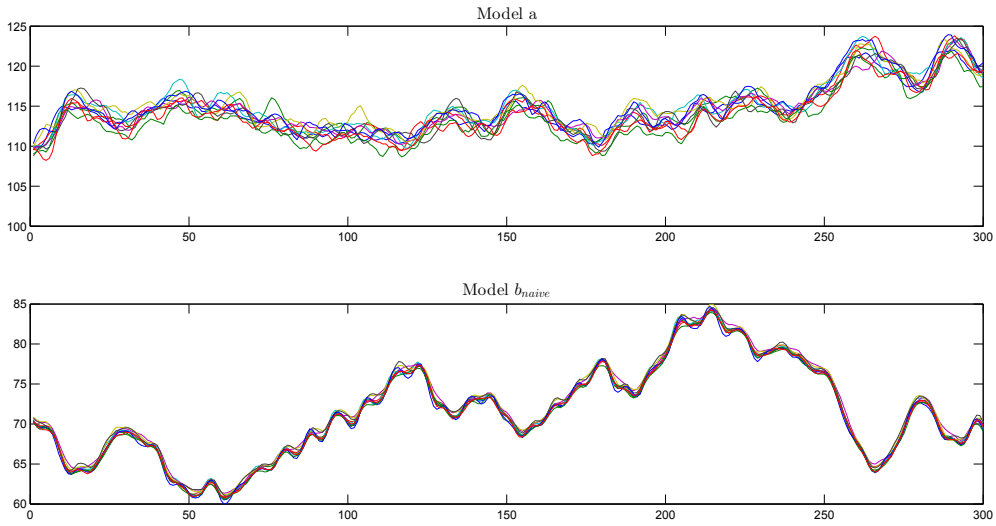
[3] This choice does not affect cointegration tests properties.

[4] The authors propose a generalization the Factor Augmented VAR models to consider cointegration and call the new model FECM. They use this DGP to show an analytical example of the model's properties and to obtain Monte Carlo results.

and there is one exogenous series which drives the long run movements, make the system simple and easily readable. Nonetheless, due to this simplicity, the model may result inadequate to describe real data dynamics. A more realistic structure may be to assume that all variables react to all cointegration relationships, in which case matrix α^U would have the following structure:

$$\alpha^U = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1r} \\ -\alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2r} \\ \alpha_{31} & -\alpha_{32} & \alpha_{33} & \cdots & \alpha_{3r} \\ \cdots & & & & \\ \alpha_{n_1 1} & \alpha_{n_1 2} & \alpha_{n_1 3} & \cdots & -\alpha_{n_1 r} \end{pmatrix}_{n_1 \times r} \tag{2.6}$$

As we did for model *(a)*, when simulating the model, we could initially assume that values α_{ij} are drawn form the uniform distribution with parameters $[0.15, 0.3]$ (call this model as b_{naive}). Note however that this new assumption does not seem realistic either. Figure 2.1 shows in its first panel an example of model *(a)* and in its second panel an example for model b_{navie} (in both cases we set with $N = 10$ and $n_1 = 9$, and used the same cointegration relationships and the same shocks to generate the processes).



Note: In model *(a)* matrix α^U is set as in expression 2.5 while in model b_{naive} it is set as in expression 2.6

Figure 2.1: Examples of series following model *(a)* and model b_{naive}

As the figure shows, in model b_{naive} series's movements seem too close to be realistic. This pattern comes from the fact that all series adjust quite rapidly and in the 'same direction' to shocks in any equation, so that desequilibria last for very short time periods[5]. Thus, if we want a 'complete' α^U we somehow need to select smaller entries for adjustment parameters to get a more realistic pattern. Instead of selecting the entries of α^U in an *ad-hoc* fashion, we show that setting matrix β^L such that cointegration relationships are deviations of each component with respect to the sub-aggregate and α^U such that each component reacts only to the desequilibria between itself and the aggregate (model *(b)*) does the trick. That is, model *(b)* is also an easily interpretable system, as was model

[5]Note also that, as may be inferred from the plot, model b_naive generates larger roots than model *(a)*. Additionally in some cases it generates roots that are larger than 1 (explosive patterns) and multiple unit roots

(a), but it neither relies on oversimplifying assumptions about α^U nor generates the pattern of model b_{naive}. Figure 2.2 shows an example to compare the patterns generated by models a and b respectively, the unrealistic pattern of model b_{naive} is not anymore present.

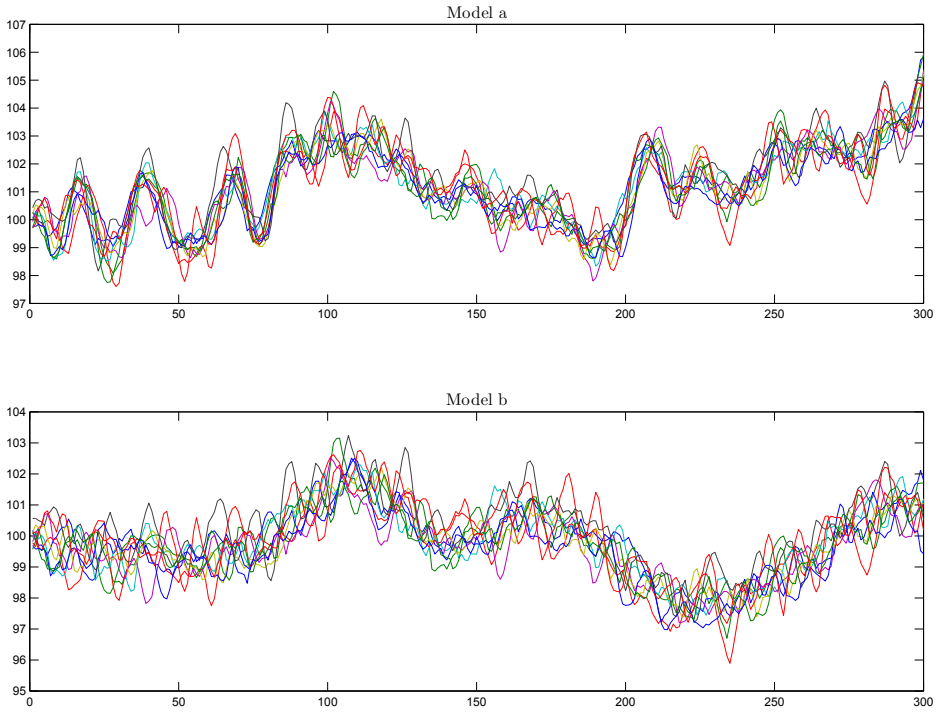


Figure 2.2: Examples of series following model *(a)* and model *(b)*

A simple way to write model *(b)* may be (to save notation, we make abstraction of the components not having the common trend -i.e, $n_1 = N$-, but the analysis keeps exactly the same when the other components are included):

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \vdots \\ \Delta X_{Nt} \end{bmatrix} = \begin{bmatrix} a_{11}^* & \dots & a_{1r}^* \\ \alpha_2^* & \dots & 0 \\ \dots & & \\ 0 & \dots & \alpha_N^* \end{bmatrix} \begin{bmatrix} -w_1\beta_2^* & 1-w_2\beta_2^* & \dots & -w_N\beta_2^* \\ \vdots & & & \\ -w_1\beta_N^* & -w_2\beta_N^* & \dots & 1-w_N\beta_N^* \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \\ \vdots \\ X_{N,t-1} \end{bmatrix} + \Phi_{n\times n} \begin{bmatrix} \Delta X_{1,t-1} \\ \Delta X_{2,t-1} \\ \vdots \\ \Delta X_{N,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{Nt} \end{bmatrix}, \tag{2.7}$$

but we need to determine the first row of matrix $\boldsymbol{\alpha^*}$ (call it $\boldsymbol{a_1^*}$). For doing this define W_{nf} as a row vector containing all weights except the first one, $\Delta X_{nf,t}$ as a column vector containing the first differences of all variables except the first one and $\boldsymbol{\alpha}_{nf}^*$ as a $(r \times r)$ diagonal matrix, whose diagonal is the vector $[\alpha_2^*, ...\alpha_N^*]$. As is clear from expression (2.7), vector $[\alpha_2^*, ...\alpha_N^*]$ contains the adjustment speeds of series X_2 to X_N to their *'own'* cointegration relationship with the sub-aggregate (call these cointegration relationships as $e_{i,t-1}^*$, for $i = 2, .., N$). Similarly, we define α_1^* as the adjustment speed of the first component to its *'own'* cointegration relationship with the aggregate ($e_{1,t-1}^*$, see Appendix A.1 for an expression of this cointegration relationship).

Define $e_{nf,t-1}^* = [e_{2,t-1}^*, ..., e_{N,t-1}^*]'$, and $\Delta^E X_{it}$ as the dynamics of series X_{it} that correspond to equilibrium correction movements.

Note now that,

$$w_1\Delta^E X_{1,t} = \Delta^E A_t - W_{nf}\Delta^E X_{nf,t} = W_{nf}(\boldsymbol{\alpha}^*_{nf} - \alpha^*_1 I_r)e^*_{nf,t-1} - W_{nf}\boldsymbol{\alpha}^*_{nf}e^*_{nf,t-1},$$

where we used the result that $\Delta^E A_t = W_{nf}(\boldsymbol{\alpha}^*_{nf} - \alpha^*_1 I_r)e^*_{nf,t-1}$ (see Appendix A.2 for a proof of this result).

Therefore:

$$\boldsymbol{a^*_1}e^*_{nf,t-1} = \frac{1}{w_1}W_{nf}(\boldsymbol{\alpha}^*_{nf} - \alpha^*_1 I_r - \boldsymbol{\alpha}^*_{nf})e^*_{nf,t-1}$$

From where;

$$\boldsymbol{a^*_1} = -\frac{1}{w_1}W_{nf}\alpha^*_1 \tag{2.8}$$

Then, expression 2.7 becomes:

$$\begin{bmatrix}\Delta X_{1t}\\ \Delta X_{2t}\\ \vdots\\ \Delta X_{Nt}\end{bmatrix} = \begin{bmatrix}-\alpha^*_1\frac{w_2}{w_1} & \dots & -\alpha^*_1\frac{w_N}{w_1}\\ \alpha^*_2 & \dots & 0\\ \dots & & \\ 0 & \dots & \alpha^*_N\end{bmatrix}\begin{bmatrix}-w_1\beta^*_2 & 1-w_2\beta^*_2 & \dots & -w_N\beta^*_2\\ \vdots & & & \\ -w_1\beta^*_N & -w_2\beta^*_N & \dots & 1-w_N\beta^*_N\end{bmatrix}\begin{bmatrix}X_{1,t-1}\\ X_{2,t-1}\\ \vdots\\ X_{N,t-1}\end{bmatrix} + \Phi_{n\times n}\begin{bmatrix}\Delta X_{1,t-1}\\ \Delta X_{2,t-1}\\ \vdots\\ \Delta X_{N,t-1}\end{bmatrix} + \begin{bmatrix}\epsilon_{1t}\\ \epsilon_{2t}\\ \vdots\\ \epsilon_{Nt}\end{bmatrix}, \tag{2.9}$$

Another possibility could be to write a model that includes the aggregate and all components except one, but in this case we should also change matrix Φ.

iii. **Equivalence condition for normalizations a and b:** given model *(a)* -model *b*- one can obtain exactly the same model by specifying the cointegration relationships with respect to the aggregate -one exogenous component- and selecting an appropriate matrix $\boldsymbol{\alpha^*}$ ($\tilde{\boldsymbol{\alpha}}$). We keep on making abstraction of the components not having the common trend.

Let model *(a)* be:

$$\begin{bmatrix}\Delta X_{1,t}\\ \Delta X_{2,t}\\ \vdots\\ \Delta X_{N,t}\end{bmatrix} = \begin{bmatrix}0 & 0 & 0 & \dots & 0\\ \tilde{\alpha}_2 & 0 & 0 & \dots & 0\\ 0 & \tilde{\alpha}_3 & 0 & \dots & 0\\ \vdots & & & & \\ 0 & 0 & 0 & \dots & \tilde{\alpha}_N\end{bmatrix}\begin{bmatrix}\tilde{\beta}_2 & 1 & 0 & 0 & \dots & 0\\ \tilde{\beta}_3 & 0 & 1 & 0 & \dots & 0\\ \vdots & & & & & \\ \tilde{\beta}_N & 0 & 0 & 0 & \dots & 1\end{bmatrix}\begin{bmatrix}X_{1,t-1}\\ X_{2,t-1}\\ \vdots\\ X_{N,t-1}\end{bmatrix} + \begin{bmatrix}\phi_1 & 0 & 0 & \dots & 0\\ 0 & \phi_2 & 0 & \dots & 0\\ \vdots & & & & \\ 0 & 0 & 0 & \dots & \phi_N\end{bmatrix} + \begin{bmatrix}\epsilon_{1,t}\\ \epsilon_{2,t}\\ \vdots\\ \epsilon_{N,t}\end{bmatrix} \tag{2.10}$$

Define now $e_{nf,t-1}$ and $e^*_{nf,t-1}$ as $r \times 1$ vectors containing the cointegration relationships in models *(a)* and *(b)* respectively. Then, given the adjustment matrix in one of the models we want to find the other one subject to the restriction:

$$\boldsymbol{\alpha^*}e^*_{nf,t-1} = \tilde{\boldsymbol{\alpha}}e_{nf,t-1} \tag{2.11}$$

As shown in Appendix A.1, $e^*_{it} = -\tilde{\beta}_i e^*_{1t} + e_{it}$, for $i = 2,..,N$, and $e^*_{1t} = -\frac{1}{\beta^*_1}\sum_{i=2}^n w_i e_{it}$. Thus, the two normalizations are equivalent when:

$$\boldsymbol{\alpha}^{*} = \tilde{\boldsymbol{\alpha}} H^{-1}, \tag{2.12}$$

$$H = -\boldsymbol{\beta}^{*}_{\boldsymbol{nf}} W_{nf} + I_r \tag{2.13}$$

and we used that $\beta_i^* = -\frac{\tilde{\beta}_i}{\beta_1^*}$ (see Appendix A.1).

2.1.2 The bi-variate sub-models

The strategy of performing cointegration tests between all possible pairs of series is justified by the fact that in a set of n_1 series that share a unique common trend there are $n_1 - 1$ cointegration relationships and the series are pairwise cointegrated. Note that this strategy requires partial systems' estimation in the sense that we assume the existence of a full VAR model for all the components and we estimate several partial bi-variate systems. Johansen (1992) state the conditions under which inference for a cointegrated VAR model can be conducted from partial models. Basically, if the partial model excludes only weakly exogenous variables for β -but condition on them- it is a valid device to make inference on the full system (see also Hendry (1995)).

When weak exogenity holds Harbo et al. (1998) show how to modify the regular Johansen's cointegration test in order to make inference form the partial model. Basically, an additional regressor (the difference of the excluded vector) has to be added to the auxiliary regression for the not excluded vector, and a new set of asymptotic tables have to be used.

The models considered in the pairwise procedure are partial in the sense that we only consider 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 form a bi-variate model (which is impossible). On the contrary, under the null of full cointegration, since every pair of variables is cointegrated, the bi-variate VAR models are complete because all relevant variables are considered as endogenous.

Note however that the lag length in the bi-variate models will be larger than or equal to that of the full model. The lag length of the partial models will depend on the series we are considering. Specifically, there are three different cases to distinguish; i) both series have the common trend of n_1, ii) only one of them have it, or iii) none of them have the common trend. It can be shown that the lag length is non-increasing form cases i to iii (for a discussion on linear transformations of VAR processes see Lütkepohl (1984)).

Additionally, it can be shown that while the AR polynomials of the sub $VARMA$ model for a pair of components that have the common trend are of the same order for the two DGPs considered, the MA polynomial is more complex in the case of model *(b)* and its complexity increases with n_1. This implies that the VAR model to approximate the true VARMA would require more lags.

Thus, the lag length has to be selected for each pair using some information criteria.

Interestingly, when the lag length is selected for each pair, the power of the pairwise procedure for finding the true number of cointegration relationships $(n_1 - 1)$ is improved with respect to the traditional Johansen's trace test. We show this with a small simulation experiment.

We consider the two alternative DGPs analyzed in pooint 2.1.1 (models *(a)* and *(b)*). For each of the two models we consider the following possibilities for $[N, n_1]$:

i. $[6, 2]$. There are 6 variables and one cointegration relationship between two of them.

ii. $[6, 3]$. There are 6 variables and two cointegration relationships between three of them.

iii. $[9, 2]$. There are 9 variables and one cointegration relationships between two of them.

iv. $[9, 4]$. There are 9 variables and three cointegration relationships between four of them.

In sum, we have 8 alternative DGPs. For each of these 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 above, for the pairwise procedure, the lag structure depends on the *type of the pair* (i.e, both series have the common trend, one have it but the other does not, neither of the series have it). Thus, we try from 1 to 5 lags in this case, and select the optimal number according to the AIC and BIC criteria.

Cointegration tests are made at 1% of significance and the number of Monte Carlo replicas was 1000.

Detailed tables opened by the number of lags and the *type of pair* are available upon request. Here, we include a summary of the results. Columns *Trace* in table 2.1 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. Columns PW 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. 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 in both models, regardless the sample size. For large samples and small number of series both procedures provide the same results (which coincide with the theoretical ones). However, as the number of series increases or the sample size reduces 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 the one obtained by Lütkepohl et al. (2003) and Johansen (1995). These authors find that cointegration test's power decreases with the number of stochastic trends in the system, so that, for instance, it would be more difficult detect a single cointegration relationship in a three dimensional system than in a bi-variate one. Note however that it is not exactly the same result since in table 2.1 we are comparing estimation and testing cointegration in a single model *vs.* doing it in several models (not one) with fewer stochastic trends.

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 loses generated by a larger number of regressors in small sample sizes (the BIC tends to select shorter lag lengths). This observation may not remain valid for model *(b)* and larger n_1 as the lag length for the pairs in which at least one of the series has the common trend depends on n_1.

2.1.3 The forecasting equations

As showed above, cointegration relationships in a model with only one common trend can always be normalized in two readily interpretable ways:

- **Normalization (a):** cointegration relationships are expressed as deviations of each component with respect to one of them.

Table 2.1: Probability of finding all cointegration relationships. Comparison between the Trace test and the Pairwise procedure

			T = 400			T = 200			T = 100		
				PW			PW			PW	
N	*model*	n_1	Trace	AIC	BIC	Trace	AIC	BIC	Trace	AIC	BIC
6	*a*	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**
	b	2	0.97	**0.99**	**0.99**	0.84	**0.99**	**0.99**	0.54	0.82	**0.86**
		3	**0.98**	**0.98**	0.97	0.84	0.95	**0.96**	0.39	0.56	**0.73**
9	*a*	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**
	b	2	0.89	**0.99**	**0.99**	0.58	0.98	**0.99**	0.12	0.82	**0.88**
		4	**0.96**	**0.96**	0.90	0.58	**0.91**	**0.91**	0.35	0.29	**0.53**

Number of replicas: 1000. Columns *Trace* 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. Columns *PW* 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.

- **Normalization (b):** cointegration relationships are expressed as deviations of each component with respect to the aggregate.

We also showed that a simple model that uses normalization *(a)* -model *(a)*- is equivalent to a more complex model that uses normalization *(b)* -i.e, while in model *(a)* the zeros in the matrix $\tilde{\alpha}$ make adjustments to long run equilibrium readily interpretable, to use normalization *(b)* for model *(a)* we need a more complex equilibrium adjustment mechanisms). The reverse is also true, a simple model with normalization *(b)* -model *(b)*-, is equivalent to a more complex model that uses normalization *(a)*.

This means that it could be possible to transform a complex model into a simple one just by changing the normalization of the cointegration matrix. Thus, trying a normalization that leads to simple α matrix may be worthwhile. Two normalizations that could be appealing form their interpretability are the cases *(a)* and *(b)*.

This result is relevant for the forecasting stage of the procedure when dealing with finite samples since we may be able to reduce the number of regressors in the individual equations (reducing the estimation uncertainty) without cost in terms of informational losses. Indeed, for each forecasting equation we may proceed as follows:

i. From all the cointegration relationships with the aggregate, select the significant ones.

ii. From all the cointegration relationships with a single component, select the significant ones.

iii. Keep the simpler model.

This strategy allows to capture all the long-run information in the system, with the minimum amount of parameters, which cannot reduce the forecasting accuracy and may improve it.

Note that point ii will work properly regardless the variable we select to normalize the cointegration matrix. Assume for example that the simplest possible specification is the model *(a)* described in previous

section, but we choose other component (say x_3) on which normalize the cointegration relationships. The only effect of this change is to add only one cointegration relationship to each forecasting equation.

To see this, let matrices $\tilde{\alpha}$ and $\tilde{\beta}$ be the ones corresponding to model *(a)* in equation 2.10. Normalizing in variable x_3 implies changing the original matrix $\tilde{\beta}'$ for:

$$\underline{\tilde{\beta}'} = \begin{bmatrix} 1 & 0 & b_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & b_2 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & b_3 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & b_4 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & & & \\ 0 & 0 & b_r & 0 & 0 & 0 & \dots & 1 \end{bmatrix}_{r \times n_1}$$

To find the new matrix $\underline{\tilde{\alpha}}$ corresponding to the new normalization write:

$$\begin{aligned} \underline{\tilde{\beta}'} &= H_{r\times r}\tilde{\beta}', \\ \underline{\tilde{\beta}'_c} &= H_{r\times r}\tilde{\beta}'_c \end{aligned} \tag{2.14}$$

where $\tilde{\beta}'_c$ is the matrix that remains after removing the first column of matrix β', so that $\tilde{\beta}'_c = I_r$. Then, we have:

$$\underline{\tilde{\alpha}} = \tilde{\alpha}H^{-1}, \tag{2.15}$$

where $H = \underline{\tilde{\beta}'_c}\tilde{\beta}'^{-1}_c = \underline{\tilde{\beta}'_c}$. Given the structure of matrix $\underline{\tilde{\beta}'_c}$, matrix $\underline{\tilde{\alpha}}$ will be equal to matrix $(\tilde{\alpha})$ except for the first column that will contain non-zero values for the rows 2 to N.

Then, choosing a 'wrong' component on which normalize the cointegration relationships will add only one term to each forecasting equation. This suggests that the effort of normalizing in all possible components to find the one that implies fewer terms in the forecasting equations does not seems to be worth given the little cost of choosing a random one. This result also implies that for each equation, cointegration relationships may be normalized in the dependent variable, what may ease the structural interpretation of each equation. Note that this is not possible when working with the full system but we could do it because we estimate the system equation by equation.

2.2 Asymptotic properties ($T \to \infty$) of the pairwise procedure

In this section we study the cost, if any, of proceeding by pairs instead of specifying the full multivariate model and perform regular Johansen's cointegration tests.

To fix ideas assume that we are dealing with a macro-variable composed by $N = 100$ basic components. The total number of pairs is $N(N-1)/2 = 4950$, so that we initially need to perform this amount of cointegration tests. Assume further that $n_1 = 40$ within the 100 components share a unique common trend and the other $N - n_1 \equiv n_1^c = 60$ have their own trends. Ideally, the procedure should identify only one subset of size 40 containing just the correct series.

Hence, the ideal properties of the procedure are:

1. Cointegration tests between all possible pairs in n_1 (which are $n_1(n_1-1)/2 = 780$) should indicate the existence of a cointegration relationship.

2. No series in n_1^c should be -wrongly- included in $\hat{n}_1$.

Condition 2 seems the most relevant since including wrong series in a fully cointegrated set will bias the forecasts of the components in that set. Additionally, if the proportion of wrong series is large the procedure will collapse. On the other hand, not fulfilling condition 1 will generate efficiency losses but will not bias the forecasts.

In this section we focus on the analytical derivations of condition 1 and give some comments for condition 2.

As aforementioned, the pairwise strategy consists of performing Johansen's cointegration tests between all possible pairs of components, so that it inherits the asymptotic properties of Johansen test. However it has two specific features which deserve special attention, namely:

- Estimation of partial models.
- Multiple testing.

The issue of partial models estimation was already addressed in point 2.1.2 above, so here we focus on the problem of multiple testing.

2.2.1 The problem of multiple testing

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

Several approaches have been proposed for controlling type I error rates in multiple testing frameworks (see e.g Benjamini and Hochberg (1995)), among which those that try to control the Family Wise Error Rate ($FWER$) or the False Discovery Rate (FDR) seem to be the most popular. Define V as the number true null hypotheses that were -wrongly- rejected, and R as the total number of rejections. Then;

$$FWER = P(V \geq 1),$$
$$FDR = E(V/R)$$

When there are some true null hypothesis any procedure that controls $FWER$ also controls FDR, that is, controlling $FWER$ requires tighter adjustments.

The Bonferroni correction adjust all p-values in a single step for ensuring that a *universal*[6] error rate of at most α is maintained. The principle is very simple; if m tests are performed each of them has to be rejected whenever its p-value is smaller than α/m (strictly speaking the corrected p-value is $1-(1-\alpha)^{1/m}$, but this magnitude is often approximated by α/m). This ensures $FWER \leq \alpha$.

The argument for such a correction is as follows. Assume that we have performed two tests whose null hypothesis are H_0^1, H_0^2, and call A, B the events of wrongly rejecting H_0^1, H_0^2 respectively. Then:

$$FWER = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 2\alpha - P(A \cap B) \tag{2.16}$$

[6]The *universal* error rate is the probability of making at least one wrong rejection.

If the events are mutually exclusive, then $P(A \cap B) = 0$ and the Bonferroni correction will deliver $FWER = \alpha$. When $P(A \cap B) > 0$ (which is the most relevant case), $FWER < \alpha$ and Bonferroni corrections will be too stringent, even when the relevant hypothesis is the *universal* one (which is seldom the case).

A case of interest may be when $P(A|B) = 1$ (or $P(B|A) = 1$), such that $P(A \cap B) = \alpha$. In this case $FWER = \alpha$ and there is no need of adjusting p-values, even if the relevant hypothesis is the *universal* one.

Using the principle of inclusion and exclusion for probability, this last argument can be generalized for the case of m tests. That is, if the probability of wrongly rejecting any combination of hypothesis at the same time is equal to α, it can be easily seen that:

$$FWER \equiv P(\bigcup_{i=1}^{m} E_i) = \alpha \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} = \alpha, \tag{2.17}$$

where we used the Binomial theorem to set $\sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} = 1$.

Since the pairwise procedure makes a large amount of cointegration tests (4950 for $N = 100$) it may be thought to raise the probability of false rejection. In this subsection we analyze this issue for the three different types of pairs:

i. Both series belong to n_1 so that the problem is rejecting the true hypothesis of $r = 1$, (although Johansen's test is sequential, the asymptotic probability of not rejecting $r = 0$ tends to 0).

ii. Only one series belongs to n_1 so that the problem rejecting the true hypothesis of $r = 0$.

iii. None of the series belong to n_1 so that the problem rejecting the true hypothesis of $r = 0$.

As we will argue, multiple testing does not occur for pairs in i. For pairs in ii. multiple testing may occur, but the procedure includes an automatic control for this issue. Still, even if is this automatic control does not work, the possible inflated rejection frequency of true null hypothesis ($r = 0$) is tolerable for widely general data sets configurations. Finally, for pairs in iii. multiple testing occurs, but the procedure automatically controls this issue and false null rejections of $r = 0$ are not an issue for these group of series.

i. **False rejection of $r = 1$**

Given the way that Johansen's trace test statistics are constructed, the probability of finding no cointegration between two cointegrated series goes to zero as T goes to infinity, so that finding no cointegration between pairs in n_1 is not an issue in large samples. The problem is therefore 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 n_1 would be $(1 - \alpha)^{n_1(n_1-1)/2}$, which quickly decreases with n_1. But, clearly, tests are not independent. Theorem 2.1 shows that they 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 ∞.

Theorem 2.1 (Asymptotic equivalence of pairwise cointegration tests in a fully cointegrated set) *Given a set of N pairwise cointegrated series (i.e, there are $N-1$ cointegration relationships among*

them) the probability of obtaining the same result in all the $N(N-1)/2$ pairwise Johansen's trace tests tends to 1 as T goes to ∞.

Proof: See Appendix B.

The intuition for this result is that, asymptotically, the $N(N-1)/2$ cointegration test are tests for one versus none common trend, which are in turn unit root tests for the estimated common trend. Since this trend is the same for all series we have $N(N-1)/2$ estimations of the same trend, which tend to the same process as T goes to ∞.

Let WR_j be the event in which the null of no cointegration is wrongly rejected for the pair j. Theorem 2.1 implies that:

$$\sum_{\substack{j_1, j_2, ..., j_k \\ 1 \leq j_1 < j_2 ... < j_k \leq N(N-1)/2}} P(WR_{j_1} \cap WR_{j_2} \cap ... \cap WR_{j_k}) = \alpha$$

Therefore, even in the rare case when the hypothesis of interest is the *universal* one defined in the Bonferroni approach (i.e, false rejecting at least one of the $N(N-1)/2$ hypothesis, which clearly is not the case of the pairwise procedure), p-values should not be corrected.

ii. **False rejection of $r = 0$**

We need to distinguish two cases:

a. One of the series belongs to the fully cointegrated set.

Two comments are relevant for this case. First, as for including a series in $\hat{n}_1$ we are requiring 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 $r = 0$ hypothesis, which in any case will be smaller than or equal to α.

Second, let X_{out} be a series outside n_1 and X_{in} a series inside n_1. The worst possible world is that finding cointegration between X_{out} and X_{in} ensures that, asymptotically, cointegration between X_{out} and any other series in n_1 will be found with probability one.

In this worst possible case the probability of including X_{out} in $\hat{n}_1$ would be α, so that the expected number of wrong series ($E[W]$) in $\hat{n}_1$ will be $(N - n_1)\alpha$. For $(N - n_1) = 100$ and $\alpha = 0.01$, $E[W] = 1$. It is in this sense ($E[W]$ is small) that we tolerate the increased false rejection probability for pairs in 2.

The proportion of wrong series in n_1 will therefore be $\frac{(N-n_1)\alpha}{n_1}$. As N increases we need to assume that n_1 increases proportionally in order to avoid $E[W]$ going to infinity. Let $n_1 = \mu N$, then the upper bound for expected proportion of wrong series in n_1 is $\frac{(1-\mu)N\alpha}{\mu N} = \frac{(1-\mu)\alpha}{\mu} = 4\%$ for $\alpha = 0.01$ and $\mu = 0.2$.

Finally the lower bound probability of not including any wrong series is $(1 - \alpha)^{N-n_1}$ which for $\alpha = 0.01$ and $(N - n_1) = 100$ is 0.37.

This was a 'worst case' analysis so the figures are upper (lower) bounds for the true ones. Simulation results show that actual figures are far from these bounds (see section 3).

b. None of the series belongs to the fully cointegrated set.

From the Johansen test properties the probability of finding one cointegration relationship between two non-cointegrated I(1) series tends to α as $T \to \infty$. Assuming that tests are independent (which seems a sensible assumption for these pairs), we could think the set $N - n_1$ as a random graph with edge probability equal to α. In this case, the expected number estimated fully cointegrated sets composed by K series of the $(N - n_1)$ would be $E[K^{lb}_{worng}] = C_K^{(N-n_1)} \alpha^{K(K-1)/2}$, what is almost zero for $\alpha = 0.01$, $K > 3$ and moderately large $(N - n_1)$ (see Bollobás and Erdös (1976) for a discussion *cliques*[7] in random graphs).

Additionally Matula (1976) showed that the size of the *maximal clique*[8] in a random graph with edge probability p has a strong peak around $2log(n)/log(1/p)$ which is 2 for $n = 100$ and $p = 0.01$. Thus, selecting a low α and disregarding estimated fully cointegrated sets with less than 3 elements constitute a *strong protection* against finding fully cointegrated sets among these series.

Note that $E[K^{LB}_{worng}]$ is a lower bound since we assumed that tests are independent. A -quite loose- upper bound can be defined by assuming that once we find one common trend in the pairs (X_{1t}, X_{2t}) and (X_{2t}, X_{3t}), we will find one common trend in the pair (X_{1t}, X_{3t}) with probability 1. Under this assumption we get $E[K^{ub}_{worng}] = C_K^{(N-n_1)} \alpha^{K-1}$, the term α^{K-1} comes form the fact that, under this assumption, in a set of K series it is enough to find cointegration in $K - 1$ pairs to find it in all of them.

Figure 2.3 illustrates $E[K^{lb}_{worng}]$ and $E[K^{ub}_{worng}]$. Notably, simulation results show that the actual number of wrong fully cointegrated sets is close to the lower bound, meaning that the independence assumption is sensible for this type of pairs (see section 3).

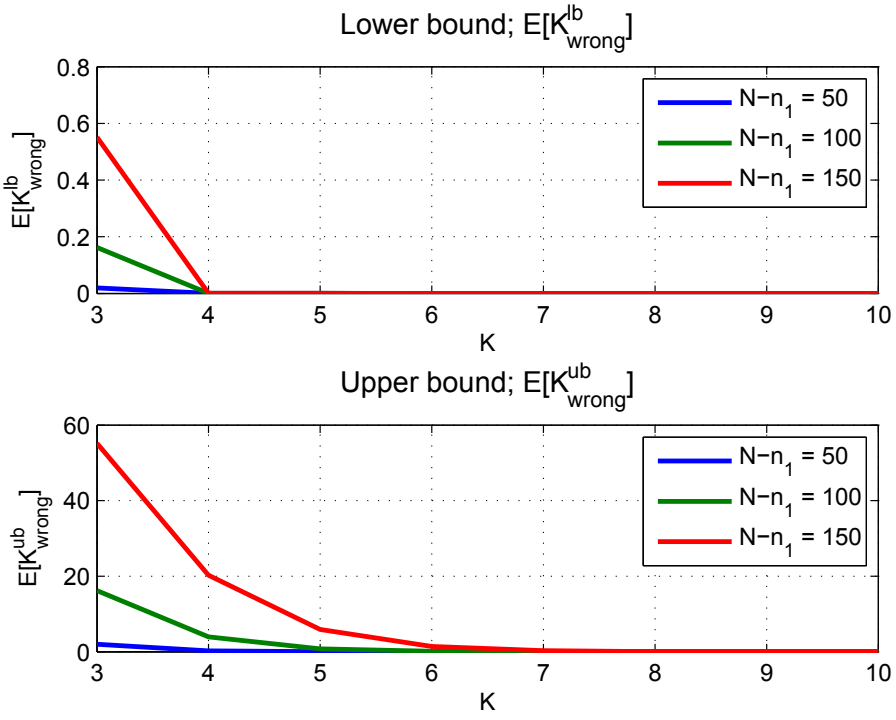


Figure 2.3: Lower and upper bounds for the expected number of estimated fully cointegrated sets composed by series that do not share any common trend ($\alpha = 0.01$)

[7] Sub-graphs in which all nodes are pairwise connected.
[8] The maximal sub-graph in which all nodes are pairwise connected.

3 Simulation results for the Pairwise strategy

In this section we perform a Monte Carlo experiment to fulfill three objectives: confirm the analytical results of section 2.2, analyze the small sample properties of the pairwise strategy, and compare its performance with other alternatives (the detailed algorithm for the pairwise strategy is included in section 6). These three objectives are covered in sections 3.2, 3.3, and 3.4 respectively.

As discussed in section 2.2, the ideal procedure will:

1. identify a large number of the series that truly share the trend
2. not include wrong series in the fully cointegrated set

Conditions 1 and 2 are 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 variables (see also Castle et al. (2012)). As an extension of condition 2, we may consider the possibility of finding 'extra' fully cointegrated sets.

In section 2.2 it was also argued that condition 2 is the most critical one since including a large amount of wrong series would have disastrous consequences for the procedure. Hence, we analyze the two conditions giving special attention to the second one.

To study these points, we simulate a N-dimensional process in which a subset of n_1 series share a single stochastic trend, and the other $N - n_1$ variables have their own trends (i.e, they are not cointegrated with any other series). We consider four scenarios that differ in the size of n_1 -we use the notation n_1 both, to indicate the size of the 'common trend subset' and as its label-.

3.1 General design of the experiments

The following Data Generating Processes are considered:

1. VEqCM structure 1: Model *(a)* analyzed in section 2.1.1, with $\tilde{\boldsymbol{\beta}} = [-1, ..., -1]$, and $\tilde{\alpha}_i$ taken from the uniform distribution with parameters $[0.15, 0.30]$. In this model cointegration relationships among series in n_1 are expressed as deviations from the first variable, which is assumed to be weakly exogenous. Each of the other variables is assumed to react only to its 'own' cointegration relationship.

2. VEqCM structure 2: Model *(b)* analyzed in section 2.1.1. In this model cointegration relationships are expressed in terms of the deviations of each variable in n_1 with respect to their sub-aggregate. Each variable in n_1 is assumed to react only to its 'own' cointegration relationship. Recall that in this model the lag length of the bi-variate sub-models which include at least one variable of n_1 will be increasing in n_1 and larger than in model *(a)*. This is because the MA polynomial of the $VARMA$ bi-variate sub-model that comes into a pair of variables which -at least one- has the common trend is more complex for model *(b)* and this complexity increases with n_1. Table 3.1 illustrates this issue.

3. The following factor structure:

$$\begin{aligned} Y_t &= \Lambda F_t + \epsilon_t, \\ F_t &= F_{t-1} + \eta_t, \end{aligned} \tag{3.1}$$

Table 3.1: Median Optimal lag length for Models *(a)* and *(b)* for the different types of pairs ($T = 400$)

	Model *a*			Model *b*		
type of pair	*both*	*one*	*none*	*both*	*one*	*none*
$n_1 = 10$	3	3	1	4	5	1
$n_1 = 15$	3	3	1	4	5	1
$n_1 = 25$	3	3	1	5	6	1
$n_1 = 40$	3	3	1	6	6	1

- We perform 1000 replicas. For each type of pair (both series belong to n_1, only one of them does, none belong to n_1) we compute the median number of lags selected by the AIC criteria in each experiment. So that for each experiment we have three medians. Call them m_{both}, m_{one} and m_{none}. This table reports the median across experiments of the m_i for each experiment and model (results for the mean are the same).
- In all cases $N = n_1 + 2$ and $T = 400$.

where Λ is a $n \times 1$ vector of factor loading whose firsts n_1 elements are taken form the uniform distribution with parameters $[0.1, 0.8]$, ϵ_t is a $n \times 1$ vector of *iid* $N(0, \Sigma)$ with $\sigma_{ii}^2 = (1 - \lambda_i^2)/2$, and η_t is a $n \times 1$ vector of *iid* $N(0, 1)$ processes independent of ϵ_t. This structure implies that the first difference of each series has unit variance, and the signal to noise ratio is $\lambda_i^2/(1 - \lambda_i^2)$.

For the three models we consider four scenarios. In all off them we set $N = 100$ and they differ in the choice of n_1. The four choices are $n_1 = 10$, $n_1 = 15$, $n_1 = 25$ and $n_1 = 40$. Scenarios 1 and 4 are motivated by results in Espasa and Mayo (2013) about CPIs' components, the other two scenarios are just to have intermediate structures. Additionally, we consider three possible sample sizes: $T = 100$, $T = 200$ and $T = 400$.

Since results are similar for the three data structures, we focus on the results for model *(a)*. Results for model *(b)* and for the factor structure are available upon request.

The number of replicas was 1000.

3.2 The behavior of the pairwise strategy

Let Z_1 be the number of correct series included in $\hat{n}_1$ (the estimated sub-set of variables that share a unique common trend). In addition to computing the potency, we compute $P(Z_1 \geq x)$ for $x \in [0, 1, ..., n_1]$. The larger this probability for each x, the better the procedure's performance. Given the results presented in section 2.2, we expect $P(Z_1 = n_1)$ to be close to $1 - \alpha$ in large samples.

Let Z_2 be the number of incorrect series included in the estimated fully cointegrated set. Apart from evaluating the gauge, we compute $P(Z_2 \geq x)$ for $x \in [1, ..., n - n_1]$. The larger this probability the worst the performance of the procedure.

Figure 3.1 includes the plots for $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$ for scenario 3 (since the basic conclusions do not change plots for the other scenarios are not included for saving space but are available upon request). Gauge and potency measures are included in table 3.4 of section 3.4 where a comparison of the pairwise strategy with an alternative approach is carried out.

As the figure shows, the procedure performs reasonably well for all sample sizes. For $T = 400$ the probability of including all the 25 series is close to 99%. On average, for this sample size, we include 24.6 correct series in the fully cointegrated subset. This result is in line with theorem 2.1, which states that the asymptotic probability of finding cointegration in all the true cointegrated pairs tends to $(1 - \alpha)$ as

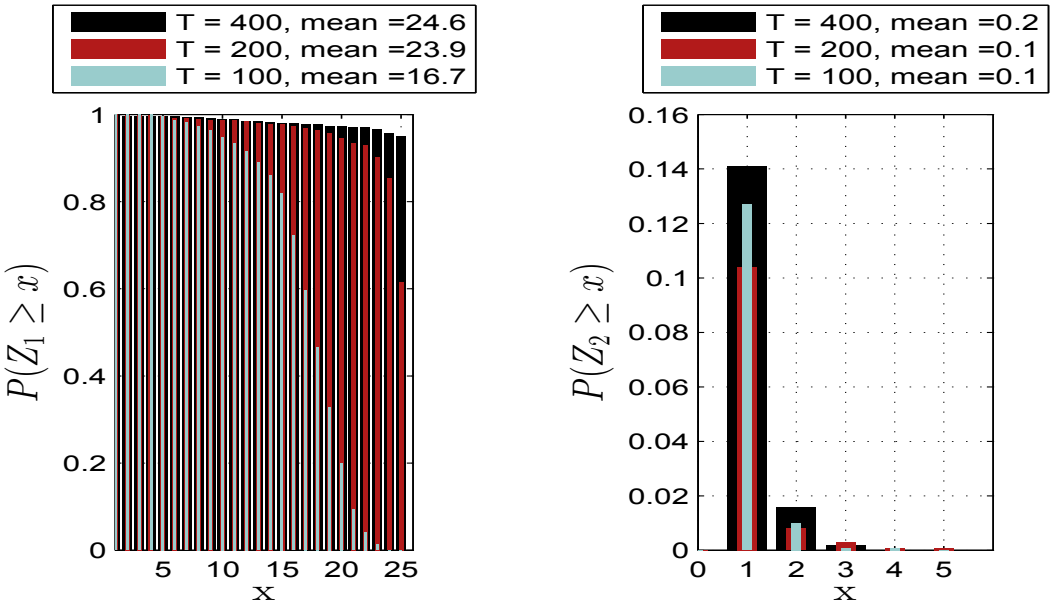


Figure 3.1: Pairwise approach. $P(Z_i \geq x)$, Scenario 3 ($n_1 = 25$)

$T \to \infty$, with α being the nominal size of the tests [9].

On the other hand, the probability of including wrong series is quite low and fast decreasing with x. The probability of including at least one wrong series is 0.14. In other words, we are including no wrong series in 86% of the experiments. Moreover, this probability goes to zero very fast, indeed including more than two wrong series has a probability close to zero.

On average, we are including just 0.2 wrong series. Recall form section 2.2 that an upper bound for the expected number of wrong series in $\hat{n}_1$ is $E[W] = (N - n_1)\alpha$ (0.75 in scenario 3). Therefore, this result shows that the actual $E[W]$ is far form this upper bound, meaning that the 'equivalence' assumption used to compute this bound is far form being correct. This performance is appealing given that including wrong series is the worst possible error.

In terms of condition 3, we did not find any fully cointegrated set composed by outsiders.

Finally, as figure 3.1 shows, although gauge remains quite stable when the sample size varies, potency deteriorates as T decreases. For instance in scenario 3 (the one included in figure 3.1) we go from $\bar{Z}_1 = 24.6$ with $T = 400$ to $\bar{Z}_1 = 16.7$ with $T = 100$. In next section we propose a modification of the original procedure that tackles this issue.

3.3 Small samples correction

There are two main reasons which could explain the deterioration of the potency in small samples:

i. The Johansen's trace test properties deteriorates in small samples.

ii. The the equivalence of the tests showed in theorem 2.1 is valid only asymptotically, and also deteriorates in small samples.

[9] The reason for not finding exactly the 99% probability is that we do not know the true lag length of the sub-models for the pairs.

To mitigate this issue, we propose a slight modification of the procedure: relax the '*full cointegration*' requirement to '*almost full cointegration*'. The relaxation consists in allowing to enter in $\hat{n}_1$ those series for which cointegration with at most λ series in the initially estimated fully cointegrated set was not found, at the original 99% confidence level, but it was at the 95% (see section 6 for the details).

This strategy will lead to increase the potency of the procedure but will also increase the risk of including wrong series. Although theorem 2.1 is only valid for a group of fully cointegrated series, it gives an intuition for the following case.

Assume that we have three series (S_1, S_2 and S_3) such that (S_1, S_2) is the unique truly cointegrated pair. Assume further that we wrongly find cointegration for the pair (S_1, S_3). The probability of finding cointegration for the pair (S_2, S_3) given that we found cointegration in (S_1, S_3) would be larger than or equal to α (the unconditional asymptotic probability); call this probability α^*.

Let Z_2^λ be the number of series which do not belong to the true fully cointegrated set but cointegration tests indicate cointegration with all but λ of the series in the original $\hat{n}_1$. That is, Z_2^λ represents the number of potential candidates to enter the *almost fully cointegrated set* which we do not want to include. For a given initial $\hat{n}_1$, the larger α^*, the larger Z_2^λ would be. Similarly, for a given α^* the smaller $\hat{n}_1$ the larger Z_2^λ would be; so that Z_2^λ is a decreasing function of the original $\hat{n}_1$.

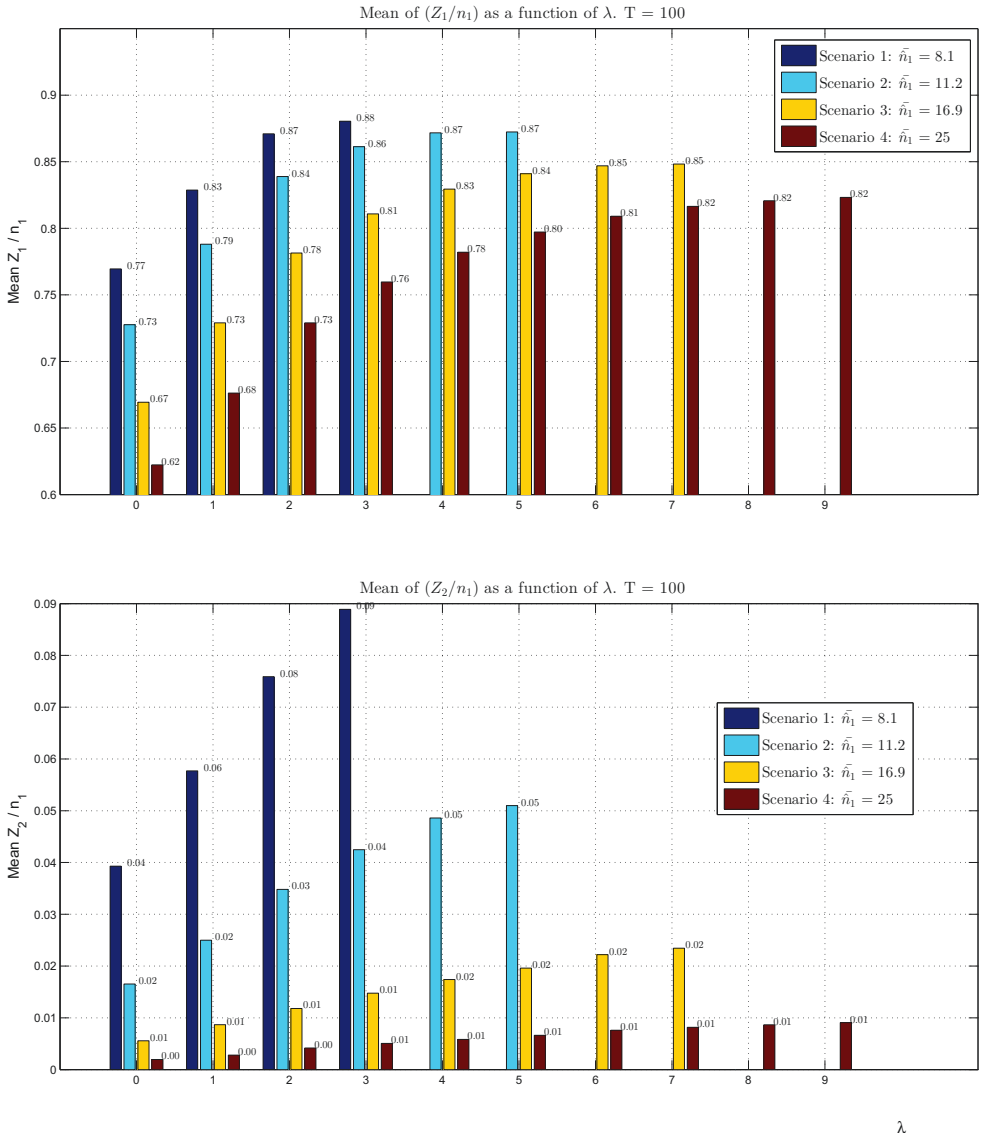


Figure 3.2: Mean of the ratios Z_1/n_1 and Z_2/n_1 as a function of the relaxation parameter λ.

Table 3.2 illustrates these arguments. It shows the mean number of potential candidates for each scenario, $T = 100$, and relaxation parameter up to $\lambda = 3$. While column (a) contains the mean number of series that have between 1 and λ holes in the current *(almost) fully cointegrated set*, column (b) includes the series of column (a) whose holes were filled after relaxing cointegration tests to the 5%. Columns (c) and (d) are analog to (a) and (b) but wrong candidates are excluded.

As the columns *Ratios* show, while for scenarios 3 and 4 (large n_1) almost all the potential candidates are correct series, this is not true for scenarios 1 and 2. This difference between scenarios becomes more evident for larger relaxation parameters (λ). For instance, with $\lambda = 1$ in scenario 1, 75% of the candidates are correct series, whereas in scenario 4, 98% of the candidates are correct series. With $\lambda = 3$, while in scenario 1 only 40% of the candidates are correct series, in scenario 4, 97% of the candidates are correct series.

Table 3.2: Statistics of the Relaxation process. Mean number of potential candidates ($T = 100$)

Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 1$

	All Candidates (Z^*)		Correct Candidates (Z_1^*)		Ratios Z_1^*/Z^*	
	(a) No Rest	(b) pval>0.05	(c) No Rest	(d) pval>0.05	c/a	d/b
Sce 1	1.06	0.92	0.76	0.68	0.75	0.75
Sce 2	1.37	1.26	1.18	1.10	0.87	0.88
Sce 3	1.96	1.83	1.85	1.72	0.95	0.95
Sce 4	2.63	2.51	2.58	2.46	0.98	0.98

Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 2$

	All Candidates (Z^*)		Correct Candidates (Z_1^*)		Ratios Z_1^*/Z^*	
	(a) No Rest	(b) pval>0.05	(c) No Rest	(d) pval>0.05	c/a	d/b
Sce 1	1.24	0.76	0.73	0.51	0.65	0.70
Sce 2	1.60	1.17	1.26	0.98	0.82	0.85
Sce 3	2.26	1.80	2.08	1.70	0.94	0.94
Sce 4	3.30	2.77	3.19	2.69	0.97	0.97

Maximum Number of holes admitted to consider a sereis to enter in $\hat{n_1}$: $\lambda = 3$

	All Candidates (Z^*)		Correct Candidates (Z_1^*)		Ratios Z_1^*/Z^*	
	(a) No Rest	(b) pval>0.05	(c) No Rest	(d) pval>0.05	c/a	d/b
Sce 1	0.85	0.26	0.29	0.10	0.42	0.40
Sce 2	1.27	0.56	0.86	0.43	0.72	0.75
Sce 3	1.75	1.00	1.57	0.91	0.91	0.91
Sce 4	2.48	1.60	2.39	1.56	0.96	0.97

- Number of experiments: 1000. For scenarios 1 to 4, n_1 is $10, 15, 25$ and 40 respectively.
- The $\hat{n}_1$ subset is updated in each step.
- Column (a) contains the mean number of series that have between 1 and λ holes in the current *(almost) fully cointegrated set* (cointegration at the 1% was rejected with at least λ series in $\hat{n}_1$).
- Series in column (b) and (d) are those of column (a) and (d) whose holes were 'filled' after relaxing cointegration tests to the 5%.
- A comparison between columns (a) and (b) or columns (c) and (d) gives an idea of the effects of requiring cointegration at the 5% for the holes to be filled vs. no requiring anything.
- Columns (c) and (d) are analog to (a) and (b) but only truly correct series are considered.
- Note that Z_2^λ does not explicitly appears in this table, it can be obtained by subtracting column (c) to column (a), or column (d) to column (d).
- All figures (including the *Ratios*) are averages across experiments. Then, figures in column *Ratios* are not necessarily equal to c/a and d/b because they are the mean across experiments.

Thus, the results in table 3.2 confirm that it is for situations with 'large' initial $\hat{n}_1$ that the relaxation is less risky, i.e, Z_2^{λ} is a decreasing function of n_1. Notably, it also happens that is precisely for those scenarios that improving the original results is most needed. Recall that the equivalence of cointegration tests proved in theorem 2.1 is valid only asymptotically; as T decreases, the asymptotic equivalence deteriorates and the probability of finding cointegration between all the correct pairs moves away form $(1-\alpha)$ and becomes a function of the number of pairs. The larger the number of pairs, the lower the probability of finding cointegration between all (or a high proportion) of them.

In line with this argument, results in section 3.4 show that the original potencies for $T = 100$ are decreasing in n_1: $0.77, 0.73, 0.67$ and 0.62 for scenarios 1 to 4 respectively.

Figure 3.2 adds more evidence for the two arguments made above, namely, while the risk of relaxing the full cointegration requirement is decreasing in n_1, and the potential benefit is increasing. The figure contains the ratios Z_2/n_1 and Z_1/n_1 as a function of λ for the four scenarios. As it shows, in scenarios 3 and 4 we can increase the ratio Z_1/n_1 (the potency) by 20 percentage points with almost no cost in terms of Z_2/n_1. This is not the case for scenarios 1 and 2, for which the benefits are lower and the costs higher.

Thus, the relaxation parameter λ (which indicates the maximum number of 'holes' that a candidate series can have to enter the *almost fully cointegrated* subset) has to be defined as a function of the original n_1. However, as we have no prior rules to define that function, we perform a simulation exercise to decide on the appropriate λ given the initial $\hat{n}_1$ (because the true n_1 is unknown in empirical applications).

Using the same simulated series as in section 3.2 we run again the pairwise procedure but instead of requiring *full cointegration* we consider the relaxation to *almost full cointegration* using alternative relaxation parameters λ. The alternative values of λ considered were; $[1, 2, ..., 9]$.

To decide on the 'optimal' λ we consider the following - arbitrary- criteria. For each experiment we count the number of incorrect series (Z_2) included in the *estimated almost full cointegrated subset* and compute the ratio $(Z_2/\hat{n}_1)$, where $\hat{n}_1$ is the originally estimated *strict* full cointegrated subset[10]. Next, we average this ratio over all the 1000 experiments. Finally, a decision rule to choose the optimal λ has to be defined.

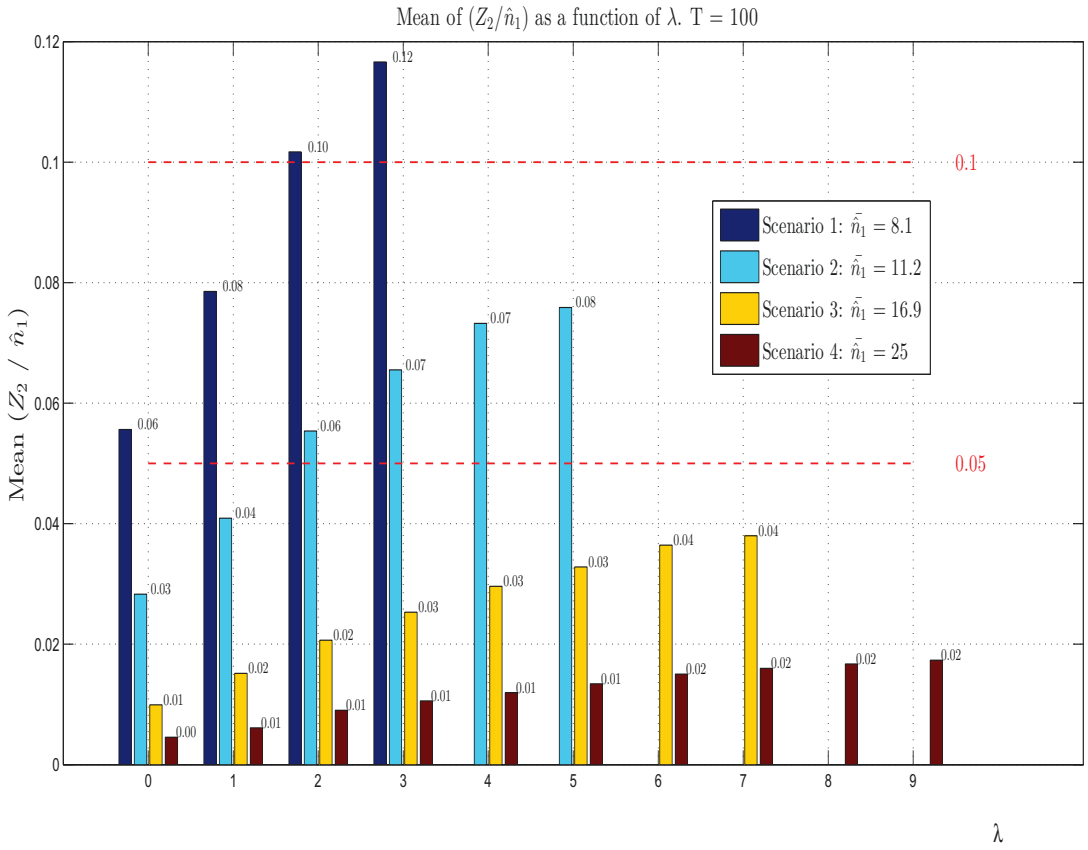
We define the optimal λ as the maximum one such that the mean ratio $Z_2/\hat{n}_1$ does not exceed a certain threshold.

Figure 3.3 includes the simulation results for the four scenarios and $T = 100$. Dashed red lines represent two arbitrary decision rules to choose the optimal λ given the initial $\hat{n}_1$. We are requiring the expected value for the ratio $Z_2/\hat{n}_1$ to be 0.05 (0.1).

With the criteria of 0.05 the optimal λ for scenarios 1 and 2 would be 0 and 1 respectively. However for scenarios 3 and 4 this rule is not operative since we never reach the 0.05 threshold. In these cases we set λ equal to 5 and 7 respectively since these are the $\lambda's$ for which the ratio $Z_1/\hat{n}_1$ stabilizes (see first panel in figure 3.2). This sort of 'jump' in the maximum number of holes admitted (from zero and one in scenarios 1 and 2, to five and seven in scenarios 3 and 4) confirms once again the argument discussed above; the probability of having a wrong series with few holes is a decreasing function of n_1.

To show an example of how this procedure works, figure 3.4 we reproduces the probability functions of figure 3.1 just for the selected λ for scenario 3.

[10] Another alternative could be to compute Z_2/n_1, but as in practice the only possible baseline is $\hat{n}_1$ we prefer the previous criteria which turns out to be more conservative since we found $\hat{n}_1 < n_1$ in all experiments.



Note: Dashed red lines represent two arbitrary decision rules to choose the optimal λ given the initial $\hat{n}_1$. We are requiring the expected value for the ratio $Z_2/\hat{n}_1$ to be 0.05 (0.1).

Figure 3.3: Mean of the ratio (Num of incorrect series / Num of series in $\hat{n}_1$ in the strict full cointegration framework) as a function of the relaxation parameter λ.

At least two relevant messages can be taken form figure 3.4. First, the *almost full cointegration criteria* may increase $P(Z_1 \geq x)$ significantly. For example in scenario 3 ($n_1 = 25$) the probability of including at least 20 of the correct series increases form 0.2 to 0.8. Second, this improvement has an associated cost in terms of $P(Z_2 \geq x)$.

If the user considers that the cost is too high, a smaller λ can be chosen. Detailed figures for all the λ's considered for scenario 3 are included in Appendix C (for the other scenarios they are available upon request). Looking at these figures the user can select the λ that matches her willingness to increase the gauge in exchange of increasing potency.

3.4 Comparison with DFM

The third objective of the Monte Carlo experiments was to compare the pairwise procedure's performance with other alternative approaches. In this section we apply DFM and compare the results of both strategies.

For implementing the DFM, we apply the traditional Principal Components procedure to 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 suggest a different number of factors we choose the minimum, otherwise we chose the mode. This procedure implies that we are not always using the same penalty function in each experiment, but artificially helps the dynamic factors methodology to

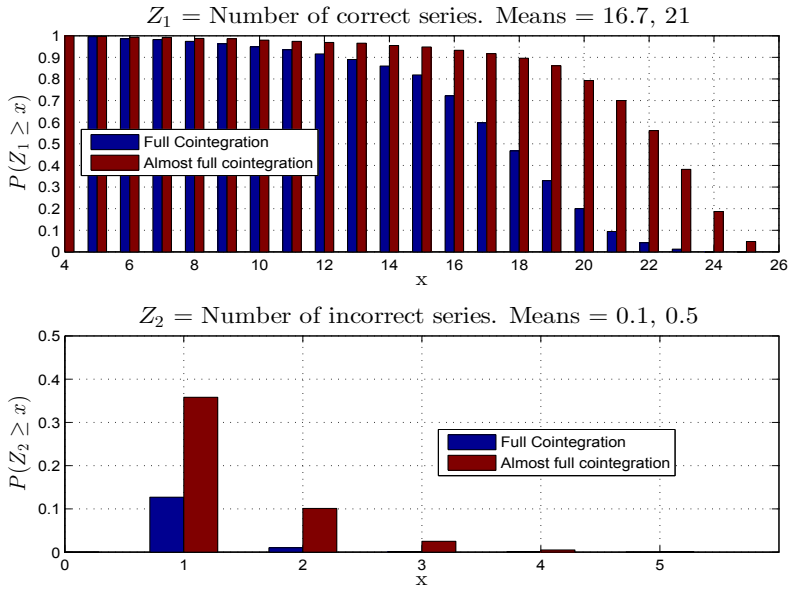


Figure 3.4: $P(Z \geq x)$. $T = 100$. Scenario 3 for $\lambda = 5$.

pick the correct number of factors (which is always one) [11].

To analyze DFM performance on conditions 1 and 2 we consider two criteria. First, factors have to be cointegrated with the series in n_1, this is an initial 'minimum quality requirement'. To analyze this issue we perform Engle-Granger cointegration tests between each of the n_1 series and each of the estimated factors[12].

Second, in order to have a proper comparison with the pairwise procedure procedure, we compute confidence intervals for the factor loadings and identify those series with statistically significant factor loadings [13]. We consider the subset formed by these series as the *DFM counterpart of the fully cointegrated subset*. Using those series we can compute the gauge and potency of the DFM approach and the probability functions, $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$, we used to study the performance of the pairwise approach.

Table 3.3 contains the mean rejection frequencies of E-G tests at 10% of significance between the true series in n_1 and the estimated first factor for the four scenarios and different sample sizes. Figures in the table are averages across experiments and series in n_1.

For instance, for $T = 400$, in scenario 1 the average probability of rejecting the null of no cointegration between one of the 10 series in n_1 and the estimated factors is 0,66. For $T = 100$ the traditional DFM procedure fails in the three scenarios. Moreover, in scenario 1 ($n_1 = 10$) DFM also fails even for moderately large samples ($T = 200$).

It is worth to mention that these conclusions are not true if the Data Generating Process is a factor

[11] For the four scenarios we proceed as suggested by Bai and Ng (2004); extracting the factors form 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 showed by Bai (2004), computing the non-stationary dynamic factors from the levels of the variables is also a correct procedure under the assumption that the unique source of non-stationarity is the factors (idiosyncrasies must be stationary). This assumption ensures that the series are cointegrated and the spurious regression problem will not be an issue. We also consider this option and, as expected, it is a worst option than extracting the factors form the differences

[12] Note that the factors are not observed series, so that a generated regressor problem may appear in these tests. The consideration of this problem will increase (in absolute value) the critical values of the tests. Since we are not dealing with this issue, we are being conservative in the sense that we will over reject the null of no cointegration.

[13] We use the asymptotic variance of the factor loadings provided by Bai (2003).

model instead of a VEqCM. In this case, all figures in table 3.3 are close to one.

Table 3.3: Mean rejection frequencies of Engle-Granger tests between series in n_1 and the estimated factors from the whole data set (H_0 : *No cointegration*)

Scenarios	T = 100	T = 200	T = 400
1	0.00	0.13	0.66
2	0.01	0.43	0.94
3	0.03	0.73	0.99
4	0.07	0.81	1.00

The null hypothesis is no cointegration.
Nominal size of the test: 10%.
Sce1: $n_1 = 10$, Sce2: $n_1 = 15$, Sce3: $n_1 = 25$, Sce4: $n_1 = 40$.

For space reasons the probability density functions $P(Z_1 \geq x)$ and $P(Z_2 \geq x)$ for the DFM procedure are not included here. Instead, table 3.4 includes the *gauge* and *potency* of the pairwise strategy and its DFM counterpart.

A general conclusion from this table is that 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 when the DFM procedure is used (except for scenario 4 and $T = 200, 400$) . However, it is also the case the worst error risk is substantially increased in scenarios 1 to 3, even for large sample sizes, i.e, the *gauge* is significantly larger in the DFM approach for scenarios 1 to 3 (relatively small n_1).

The DFM counterpart of the pairwise procedure procedure fails to isolate the series in n_1 for scenarios 1 to 3, even for large samples. In scenario 1 with $T = 400$, on average we include $0.81 \times (100 - 10) = 7,3$ wrong series. This bad performance substantially deteriorates as the sample size decreases. For T = 100, on average 24,2 wrong series are included in $\hat{n}_1$. For scenarios 2 and 3 these comments are also valid, except for scenario 3 and $T = 400$.

Table 3.4: Comparison of Gauge and Potency of the Pairwise procedure with its DFM counterpart

Pairwise

	Sce 1		Sce 2		Sce 3		Sce 4	
	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot
T=100	0.4	77.0	0.3	72.8	0.2	66.9	0.1	62.2
T=200	0.3	96.9	0.2	96.5	0.2	95.6	0.1	94.9
T=400	0.3	98.5	0.2	98.2	0.2	98.3	0.2	98.0

DFM counterpart

	Sce 1		Sce 2		Sce 3		Sce 4	
	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot
T=100	26.8	94.0	16.0	98.7	5.1	97.6	1.0	87.3
T=200	18.3	99.7	6.7	99.9	0.7	99.4	0.0	93.0
T=400	8.1	100.0	1.4	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

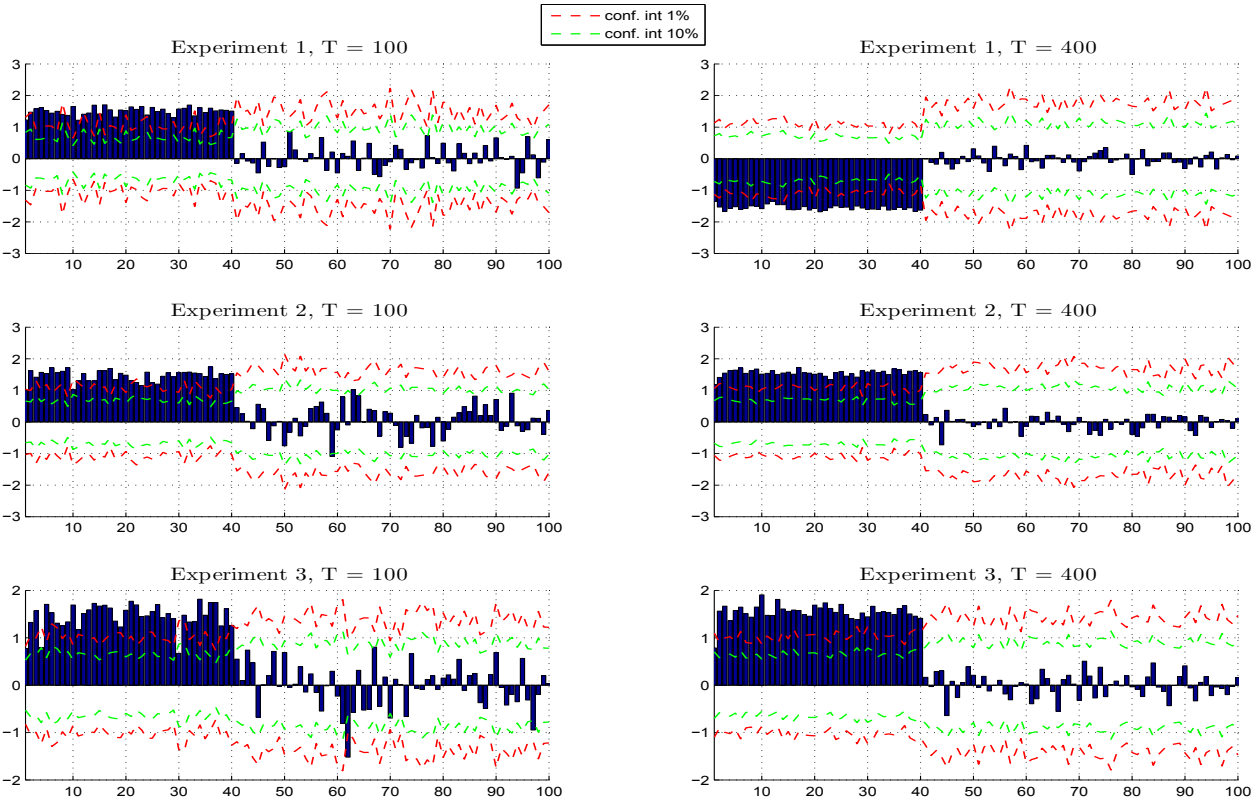


Figure 3.5: Three random examples to compare the factor loadings in scenario 4 for $T = 100$ and $T = 400$

Overall, the main conclusion form table 3.4 may be that the pairwise procedure is preferred for situations of relatively small n_1 -with respect to n-; 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 performs very good (these conclusions are also valid when the DGP is a Factor model instead of a VEqCM).

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

To close this section it is worth to highlight an apparent contradiction between the good performance of DFM in scenario 4 for $T = 100$ showed in table 3.4, and its bad performance in terms of cointegration showed in table 3.3. According to table 3.4 the procedure successes in identifying the correct series in n_1 giving significant loadings to the correct series and not significant ones to the series outside n_1. However, according to table 3.3 the estimated factor (that aggregates the series using the estimated loadings) is not cointegrated with the series in n_1, which are precisely those whose loadings are significant.

Figure 3.5 shows that this is not necessarily a contradiction. It includes the estimated factor loadings in three random experiments for $T = 100$ and $T = 400$. As it shows, although for both sample sizes the procedure correctly identifies the 40 series, for $T = 100$ the remaining loadings are substantially more noisy than for $T = 400$. This explains the apparent contradiction between tables 3.3 and 3.4, and suggests -in line with Boivin and Ng (2006) and Beck et al. (2011)- that results could be improved by extracting the factors only form the series that show significant loadings (at least for $T = 100$).

4 Outliers and breaks

The presence of outlying observations can generate devastating effects on parameter estimates and inferential conclusions if not adequately treated. Dealing with this issue in non-stationary data is specially troublesome since results on unit root and cointegration rank tests are affected by the presence of outliers and breaks, and the other way round; tests for the presence of outliers and breaks will also be affected by the presence of unit roots (see *inter alia* Perron (1989); Perron and Vogelsang (1992); Doornik et al. (1998); Johansen et al. (2000), Perron and Rodríguez (2003), Perron (2006) and Juselius (2006)).

We use the term *break* for changes in the deterministic components of the model (i.e, a level shift), and with the term *outlier* we refer to other types of atypical observations in the mean -of the stationary transformation- of a series with non-permanent effects.

When trying to distinguish between a unit root and a (trend-) stationary process, traditional tests will tend to keep the null of unit root when the process suffer mean shifts but is stationary within regimes. Additionally, when trying to detect a structural break, most tests will reject the null of no break when the process has a unit root but with constant parameters. Similarly, as noted by Quintos (1998) and stressed by Perron (2006), tests for structural breaks on cointegrated systems will over-reject the null of no break when the cointegrating rank is over specified (when the number of unit roots in the system is under-specified). Furthermore, cointegration rank tests will under estimate the number of cointegration relationships if the data is subject to breaks. Therefore, a circular problem exists when dealing with non-stationary series that may be subject to structural breaks.

On the other hand, an additive outlier (AO) has the opposite effect on unit root tests. As noted by Franses and Haldrup (1994), the presence of AOs induce a negative MA component in the residuals making traditional unit root tests to over-reject the null of unit roots.

These facts make the assessment of cointegration rank (and/or integration order) in the presence of outliers and breaks to be difficult because the appropriate treatment of these observations and the cointegration rank should, in principle, be decided simultaneously.

Although the pernicious effects of outlying observations and breaks in cointegration analysis is very well documented in the literature, the question of how to deal with these issues has not clear and generally accepted answer. In this section we include a brief review of the proposals available in the literature to tackle this problem (subsection 4.1), and propose an empirical strategy for dealing with these issues in the framework of the pairwise approach (subsection 4.2).

4.1 Alternatives for dealing with outliers and breaks in cointegration

4.1.1 Robust methods

The robust alternatives to the traditional cointegration (Johansen's) tests are based on fat tailed distributions of the errors.

There are two possibilities; the pseudo likelihood LR test proposed by Lucas (1997) and Lucas (1998), and the pseudo likelihood LM test proposed by Lucas (1998). In both alternatives the distribution of the test statistic depends on nuisance parameters (canonical correlation between linear combinations of the innovations and the pseudo-score), so that critical values should be simulated for each case. This drawback is minimized by the fact that the nuisance parameters can be estimated.

For the LR test Lucas (1997) and Lucas (1998) consider two alternatives. In the first one the critical

values are simulated in each case estimating the nuisance parameter (proposal of Lucas (1997) and the LR1 in Lucas (1998)). The second alternative is the conservative approach in which the critical values are simulated using Gaussian errors in the DGP. That is, the critical values of non-Gaussian tests are generated using Gaussian errors. This approach is conservative since critical values are always to the right of the 'true' ones. This procedure is labeled as LR2 by Lucas (1998).

The Pseudo likelihood LM test proposed by Lucas (1998), is an alternative procedure that the author considers that could be relevant since, as he shows, its distribution depends on less nuisance parameters than the LR test. This does not means that will perform better but it could be the case. The procedure for generating critical values is as in LR1.

These alternatives (LR and LM) are evaluated by Lucas (1997), Lucas (1998) and Franses and Lucas (1998) under different frameworks. Lucas (1997) and Lucas (1998) consider the tests' performance under different possibilities of innovations' distributions. Franses and Lucas (1998) consider the performance of the LR2 approach under the presence outliers (additive and level shifts). Lucas (1997) considers the performance of LR1 for a DGP without systematic growth and Lucas (1998), considers the performance of LR1, LR2 and LM for DGPs with and without systematic growth. Additionally, Lucas (1998), analyses the possibility of GARCH residuals.

The main results of these papers are the following; in terms of size: the LM procedure is conservative, LR2 is conservative somewhat more than the LM (clearly for t-student errors with $v \leq 5$, i.e, when the 'distance' to normality is larger) and LR1 is somewhat oversized. In terms of power: if innovations are Gaussian, Gaussian based tests are the better, thus Johansen's test is the preferred option, for non-Gaussian innovations, any option (LM, LR1, LR2) is better than Johansen's test, LR1 is the most powerful (but has also the worst size performance), except for Gaussian innovations and LM is always slightly better than LR2.

In summary: *"In terms of overall performance, we advocate the use of either the LM-test with simulated p-values, or the LR-test with conservative critical values. The former is preferred in situations with conditional leptokurtosis, the latter in cases with conditional normality."* (Lucas, 1998 *pp* 207) .

For empirical applications of these robust methods see Franses et al. (1998), Bosco et al. (2007), and Nielsen (2008).

4.1.2 Modeling outliers and level shifts with artificial variables

In a simulation exercise Nielsen (2004) finds that the aforementioned pseudo likelihood approach is not generally robust to AOs in small samples, which contradicts Franses' and Lucas' (1998) results. As Nielsen argues "*additional simulations suggests that the pseudo likelihood approach performs better in simple models where AOs can more easily be approximated by sequences of IOs"*, which is in fact the case of the DGP considered by Frances and Lucas (1998).

Nielsen (2004) analyzes a procedure based on the inclusion of dummy variables in the Gaussian model. He focus on analyzing the effects of AOs and IOs in the cointegrated VAR model. Specifically, his major concern is to what extent the usual practice of including unrestricted dummy variables to correct large residuals is indeed justified.

In line with other author's results (e.g, Doornik et al. (1998), and Franses and Lucas (1998)) Nielsen (2004) finds that the presence of IOs do not generate relevant distortions on cointegration's rank inference, so that he focus the analysis on the effects of AOs.

Nielsen's simulation results indicate that the common practice of including unrestricted impulse dum-

mies "*is not well suited; and to remove isolated outlying observations the estimation model introduces shifts in the levels of the non-stationary directions. The test are highly over sized...*". For solving this issue he proposes an iterative two steps estimation method, and a testing strategy to decide whether or not to include an intervention in the model (see the details of the procedure in Nielsen, 2004). The author shows that his suggested procedure gives correct size and power for the cointegration rank tests when the process contains AOs.

However, as Nielsen (2004) recognizes, his proposal implies highly non standard procedures that significantly complicates estimation. For this reason, he explores by simulation other estimation alternatives that could achieve similar size and power properties but avoiding the trickiness of his proposal.

As the first alternative Nielsen (2004) considers the estimation of the model with interventions but not imposing the non-linear restrictions that appear in a model with interventions, so that the model can be estimated by the standard Reduced Rank Regression procedure. Results do not show a significant improvement with respect to the unrestricted approach.

Given that the outliers' locations are initially considered as known, another possibility is to use univariate procedures to correct the series for the outlying observations in a first step and then testing for cointegration in the 'cleaned' series. Nielsen (2004) shows that this procedure results in tests with sizes close to the correct ones (although some power loss is also found).

When outliers' locations are unknown, Nielsen (2004) considers two possibilities. In the first approach the author generalizes to the multivariate case the Perron and Rodríguez (2003) procedure and apply his estimation procedure that explicitly considers the non-linear restrictions. The other approach is looking for outliers in univariate ARIMA models using the program TRAMO (see Gómez and Maravall (1997)) and test for cointegration in the corrected series.

Extension of Perron and Rodriguez (2003)

By extending the work of Vogelsang (1999), Perron and Rodríguez (2003) propose a procedure to estimate the outliers' locations in possibly non-stationary time series which asymptotic properties do not depend on the actual integration order (I(1) or I(0)) of the process. Such a procedure is very appealing since it would solve the circular problem described at the beginning of this section.

The procedure follows the tradition of testing for outliers in a sequential fashion (see e.g Tsay (1986) and Chen and Liu (1993), for some examples, or Perron (2006) for a complete survey on methods for determining outliers' and breaks' locations). The authors estimate the outliers' dates by looking for the largest t-statistics in a univariate model for the differenced data. The generalization to the multivariate case applied by Nielsen (2004) consist of changing the tests statistic used by Perron and Rodríguez (2003) by the maximum -over t- of a LR test that compares compares the likelihood of the VEqCM model without interventions and a model with interventions.

As aforementioned this procedure has the important drawback of requiring non-standard estimation procedures of the cointegration model. This fact leads Nielsen (2004) to explore a simpler alternative.

Univariate corrections of AOs

In a first step outliers are corrected in a univariate model (using TRAMO or Perron and Rodriguez, 2003 procedure), and then standard cointegration tests are performed over the 'corrected' series. Monte Carlo simulations show correct size results for this procedure, but some power loss with respect to the multivariate procedure.

Although not considered by Nielsen (2004), a third possibility closely related with the previous one would be to extend the Saikkonen and Lütkepohl (2000) feasible GLS procedure to the case of unknown break dates. This extension is carried out by Lütkepohl et al. (2004).

Feasible GLS procedure

As analyzed by Johansen et al. (2000), in the traditional Gaussian approach, the asymptotic distribution of the cointegration rank test changes due to the presence of structural breaks, and it depends on the breaks' dates. These new distributions can be approximated by *Gamma* functions whose parameters (mean and variance) can also be approximated by certain functions of the number of non-stationary relations and the location of break points. However, Johansen et al. (2000) restrict their attention to the case of a broken level in a model without deterministic trend, and a broken linear trend in a model with linear trend only outside the cointegration relationship, in both cases they consider one or two breaks.

The feasible GLS estimation procedure of the coefficients associated to the deterministic parameters proposed by Saikkonen and Lütkepohl (2000) and Saikkonen and Lutkepohl (2002) has the virtue that test statistic's distribution does not depend on the break dates.

The authors propose a two step procedure for dealing with deterministic components and interventions in the cointegrated VAR model. Specifically, Saikkonen and Lütkepohl (2000) generalize the proposal of Saikkonen and Lutkepohl (2002) for dealing with deterministic components (constants and trends) to the case of a single level shift at a known date. The starting point of these authors is the following unobserved components model:

$$X_t = \mu_0 + \mu_1 t + \theta DS_t + Y_t, \quad t = 1, 2, ..., \tag{4.1}$$

where $DS_t = 0$ if $t < T_1$, $DS_t = 1$ if $t \geq T_1$, and it is assumed that $\lambda = T_1/T$ remains fixed as T grows. Y_t is unobserved, assumed to be at most I(1) and to follow a cointegrated VAR(p) process, whose VEqCM representation is:

$$Y_t = \alpha\beta' Y_{t-1} + \sum_{j=1}^{k-1} \Phi_j \Delta Y_{t-j} + \epsilon_t, \quad t = 1, 2, ..., \tag{4.2}$$

where ϵ_t is a Gaussian white noise.

The proposal of Saikkonen and Lütkepohl (2000) is to obtain initial estimators for α, β, Φ_i and Γ (the covariance matrix of ϵ_t) in a VEqCM with deterministic terms and interventions, and then estimate μ_0, μ_1 and θ by multivariate GLS. Once this is done the unobserved process Y_t can be obtained as $\hat{Y}_t = X_t - \hat{\mu}_0 - \hat{\mu}_1 t - \hat{\theta} DS_t$.

The authors show that the asymptotic distribution of the traditional LR test applied to $\hat{Y}_t$ is not affected by the inclusion of impulse or step dummies in the original model. That is, they prove that the asymptotic distribution of the test is exactly the same as the one obtained by Saikkonen and Lütkepohl (2000) for a model without outliers or breaks. This feature constitutes an important advantage over traditional procedures (see Johansen et al. (2000)) since for the latter specific asymptotic tables need to be generated in each case as critical values depend on the break dates, what is specially undesirable for applied work since new tables are needed whenever new data points become available.

Given this advantage of the Saikkonen and Lütkepohl (2000) procedure over Johansen et al. (2000), the next step is to compare the asymptotic and small sample properties of both testing approaches. The comparison is carried out by Lütkepohl et al. (2003). About asymptotic properties, the authors conclude

that in the case of no deterministic trend the feasible GLS procedure is clearly superior. When the the process has a trend, it is also superior but there are some special cases when the traditional approach has more local power.

Regarding the small sample properties Lütkepohl et al. (2003) conclude that "... *the tests perform rather similar with respect to power although each of the test proposals has relative advantages in specific situations. Generally, the Saikkonen and Lütkepohl (2000) tests have more favorable size properties, however. Therefore, we recommend to use them in applied work.*"

This procedure is extended by Lütkepohl et al. (2004) for the case of a unique level shift at an unknown date. The main difference with respect to Saikkonen and Lütkepohl (2000) is that the extended procedure includes an initial step in which the break date is estimated based on a VAR in the levels of the variables.

Once the break date has been estimated, Lütkepohl et al. (2004) proposal is to apply the same feasible GLS procedure as Saikkonen and Lütkepohl (2000) for determining the cointegration rank. The authors derive its asymptotic distribution and show that it is the same as the one derived by Saikkonen and Lütkepohl (2000) for the case of a known break date, which was in turn the same as the obtained by Saikkonen and Lutkepohl (2002) for the case of no breaks.

4.1.3 Impulse Indicator Saturation

The procedures for multiple outliers and breaks correction considered so far follow the tradition of estimating the location of atypical observations in a sequential fashion by including artificial variables (one by one) and selecting the model with the maximum (minimum) t-statistic (sum of squared residuals) associated with the artificial variable. Once one date is selected, the procedure is repeated until no outlying observations are found.

As argued by Peña et al. (2001) sequential procedures suffer from three main drawbacks. First, they will often wrongly identify level shifts as innovative outliers (see also Balke (1993)). Second, initial parameter estimates are biased so that the outlier search is done based on a wrong model, what could make the procedure to fail. Finally, when the series have patches of additive outliers and level shifts, the procedure may also fail due to the fact that the correlation between the effects of consecutive outliers can be very high.

Recent developments on automatic model selection procedures applied to fully saturated regressions with impulse indicators seem to provide a general method to identify outlying observations that do not suffer from those drawbacks.

To determine where atypical observations may situate, the impulse indicator saturation (IIS) methodology requires the inclusion of T indicator variables $d_{j,t} = 1_{\{j=t\}}$ for $j = 1, .., T$ (one indicator for each observation), in the regression model. Since a perfect fit would turn out in such a model, the indicators must be included in groups.

As described *inter alia* by Santos et al. (2008), in the first step only half of the indicators are included ($d_{j,t} = 1_{\{j=t\}}$ for $j = 1, .., T/2$), and those that are statistically significant at a predetermined significance level α ($|t_j| < c_\alpha$) are recorded. Next, the first $T/2$ indicators are dropped and the ones for the remaining observations are included. Finally, the significant indicators in each step are included altogether and those that are non-significant dropped.

Santos et al. (2008) derive the asymptotic distribution of the mean and variance estimators computed after IIS under the null of normality and no outliers. They also conduct Monte Carlo studies to assess the small sample properties of the estimators and check the performance of the procedure in a highly

non-normal case (t_4-distributed error). The authors conclude that Monte Carlo simulations match the theoretical analysis and that the algorithm is also relevant for fat-tailed distributions.

Johansen and Nielsen (2009) analyze the properties of robust regression parameter estimators and consider IIS regressions as a special case of M-robust estimators. In impulse saturated regressions, parameters are firstly estimated without indicators and the IIS procedure aforementioned is then applied over the residuals of this first step regression. The authors derive asymptotic distributions under stationary and trend stationary autoregressive processes under the null of no outliers. They show that the efficiency loss due to testing the significance T indicators is almost nonexistent for low α ($\alpha \leq 1/T$). Indeed, in the case of no outliers and with $\alpha = 1/T$ the procedure will, on average, retain only one indicator. This has the negligible negative effect of dropping just one non-outlying observation.

Along the lines of Johansen and Nielsen (2009), Castle et al. (2012) study the characteristics of IIS but in a framework in which the other regressors are also selected with a general to specific methodology. The authors use the algorithm *Autometrics*. Since the null-distribution theory of *Autometrics* has not been developed, Castle et al. (2012), perform several simulation studies to analyze its operational characteristics. Specifically, they focus in the empirical null retention frequency ('gauge') and the no-null retention frequency ('potency') of indicator variables, and conditional and unconditional mean standard errors (MSEs) of exogenous' variables' parameters. These exercises are conducted not only for normally distributed residuals but also for fat-tailed ones (Student-t with 3 degrees of freedom). Their results show that non-relevant efficiency loss is caused by the inclusion of T impulse indicators when they are irrelevant, but "*helps to correct in fat tailed distributions*".

Castle et al. (2012) also study the performance of *Autometrics* with IIS under the presence of outliers. Their results show that *Autometrics* with IIS performs well in jointly selecting variables and detecting breaks for all the specifications they tried: location-scale, location-scale-trend, stationary autoregressions and unit root processes.

4.2 Empirical strategy

As aforementioned the problem we are dealing with requires cointegration tests that consider the possibility of multiple outliers and breaks. Since providing new theoretical results on these areas is out of the scope of this research, we propose an empirical strategy and check its properties by Monte Carlo experiments.

Our proposal is to jointly select the dynamic structure and the interventions applying *Autmetrics* with IIS in univariate models for the first differences of the variables (the *DGUM* in Castle et al. (2012)) [14]. Next, after the interventions are identified, we may apply three alternative procedures:

i. Estimate a VEqCM model including the interventions found for both variables, drop the insignificant and test for cointegration with the Johansen test. Note that this procedure may require simulating critical values for each test

ii. Estimate uni-equational models including the interventions for both variables, drop the insignificant and test for cointegration using the *PcGive* approach. This does not require the simulation of new critical values, but requires exogenous variables.

iii. Use the estimated dates and apply the GLS procedure described above.

[14] This strategy implies the assumption that series are at most I(1).

We disregard alternatives i and ii. The former due to the complexity of simulating new critical values for each test, the latter due to the absence of guarantees about the exogeneity of the variables. Then we focus on the third alternative. [15]

4.2.1 The problem of consecutive outliers

Let y_t be one component of the aggregate and assume that it is an $I(1)$ process with some level breaks such that;

$$y_t = x_t + \gamma LS_t \tag{4.3}$$

$$\Phi(L)x_t = \epsilon_t \tag{4.4}$$

where $\Phi(L)$ is a polynomial in L with one unit root, ϵ_t is a Gaussian white noise, γ is $1 \times k$ vector of coefficients and LS_t a $k \times 1$ vector of level shifts, and x_t is the uncontaminated (unobserved) process.

From (4.3) and (4.4)

$$\Phi(L)y_t = \Phi(L)\gamma LS_t + \epsilon_t \tag{4.5}$$

If $\Phi(L) = (1 - L)$, then (4.5) becomes;

$$\Delta y_t = \gamma \Delta LS_t + \epsilon_t \tag{4.6}$$

In this simple case, the IIS methodology will, hopefully, find the correct break dates and the strategy outlined in section 4.2 will, hopefully, work.

However if, as is usually the case, $\Phi(L)$ has a more complex structure the strategy previously outlined needs more elaboration. Let, for instance, $\Phi(L) = (1 - L)(1 - \phi_1 L)$, then, the 'true' model for Δy_t becomes;

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \gamma \Delta LS_t - \phi_1 \gamma \Delta LS_{t-1} + \epsilon_t \tag{4.7}$$

and in the more general case that $\Phi(L)$ has a larger structure;

$$\Delta y_t = \phi_1 \Delta y_{t-1} + ... + \phi_p \Delta y_{t-p} + \gamma \Delta LS_t - \phi_1 \gamma \Delta LS_{t-1} - \phi_2 \gamma \Delta LS_{t-2} - ... - \phi_p \gamma \Delta LS_{t-p} + \epsilon_t \tag{4.8}$$

Therefore, applying IIS we may find more than k impulses, so that the break date does not emerges directly since we cannot distinguish between lagged values of a break and different breaks. This is a relevant issue since the GLS procedure requires the inclusion of LS_t only.

[15] We are not considering the possibility of seasonal unit roots, so if the series contain seasonal unit roots, the aforementioned procedures could be applied to seasonal adjusted series. Although seasonal adjusting procedures may artificially generate common dynamics in the series, adjusted series are sometimes used by econometricians, possibly as a way the avoid complex estimation problems. Just to cite some articles, Hendry and Hubrich (2011), Stock and Watson (2007) and Trenkler et al. (2007) are relevant examples. Alternatively, the procedure can be easily extended for the inclusion of seasonal dummies and stationary seasonal dynamics.

One possible solution for this issue may be to apply IIS in a model with AR residuals (instead of the lagged dependent variable). However, such a model includes non-linear restrictions in the parameters, what would make the estimation procedure highly complex given the large amount of indicators we need to include.

In order to keep the estimation procedure simple we proceed as follows. First apply IIS with the lagged dependent variable as regressors and store the retained impulses. Next, if there are potential lagged impulses, estimate a model with AR residuals including the retained impulses as regressors, test the significance of the potential lagged impulses' parameters and drop the insignificant ones.

With this simple procedure we are testing if the lagged impulses' coefficients satisfy the restrictions implied by the dynamics of the model, or are *true* impulses. To see this assume that we are dealing with an AR(1) model so that the true model is given by expression (4.7). Assume also that after IIS we retained LS_t and LS_{t-1}, so that the procedure requires estimating:

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \lambda_1 \Delta LS_t + (\lambda_2 - \phi_1 \lambda_1)\Delta LS_{t-1} - \phi_1 \lambda_2 \Delta LS_{t-2} + \epsilon_t,$$

and testing $(\lambda_2 - \phi_1 \lambda_1) = 0$. If the true model is (4.7) we will not reject the hypothesis and keep only ΔLS_t as the *true* impulse. Note that this argument also works for a general $AR(p)$ model and any distribution of the impulses.

5 Simulation results for the Pairwise strategy with outlier correction

To analyze the performance of the strategy proposed in section 4.2 we perform two different simulation experiments. First we study the performance of IIS and the procedure for correcting consecutive interventions in a simple framework such that we do not need to simulate a large number of components. Then, we hash up the simulation exercise of section 3.2 but applying the outlier correction strategy described in section 4.2.

5.1 IIS and consecutive outliers correction

With the aim of keeping things simple we apply IIS without *Autometrics* - what Castle et al. (2011) call the *1-cut approach*- and check its properties by Monte Carlo. Additionally, in this experiments we also study the performance of the procedure for correcting consecutive interventions outlined above (results for these two analyses are summarized in tables D.2 and D.1 of Appendix D).

We simulate three alternative un-contaminated DGPs and then contaminate them with five alternative level breaks configurations. The DGPs are the following:

i. DGP1: a random single series taken form *model a* with $n_1 = 25$ (see section 2.1.1) and belonging to the fully cointegrated set. Note that the univariate AR model for such series is un-known (we include three lags in the equation for IIS).

ii. DGP2: a random single series taken form *model a* with $n_1 = 25$ (see section 2.1.1) and not belonging to the fully cointegrated set. Note that the univariate AR model for such series has only one lag, and we use this information.

iii. DGP3: $\Delta y_t = \epsilon_t$, with $\epsilon_t \sim iid, N(0, 1)$. Here we also use the information that the true model has no lags.

The sample size is $T = 200$ and the five level shifts distributions are: $[LS_{102}]$, $[LS_{77}, LS_{177}]$, $[LS_{52}, LS_{102}, LS_{152}]$, $[LS_{52}, LS_{53}]$, and $[LS_{52}, LS_{53}, LS_{54}]$. Additionally we consider a case with no breaks.

For analyzing the strategy of consecutive outliers correction, we compute the mean number of consecutive outliers retained that are not *'original'* ones when applying and when not applying the consecutive correction.

Table D.1 of the Appendix includes the results. As it shows, the consecutive correction procedure is useful for separating the *true* outliers form those that correspond to the model's dynamics regardless the DGP and the outliers' distribution.

Note that for the outliers distributions 4 and 5 the consecutive outliers correction may reduce IIS's potency since it may eliminate *'true'* consecutive outliers. Table D.2 of the Appendix shows that this is not the case; the potency of the procedure is almost not affected after applying the correction for consecutive outliers.

From table D.2 one can also analyze the properties of *1-cut* IIS for alternative DGP's, and outlier's distributions and sizes. Results for $DGP3$ and outliers distribution 1 can be compared with those in table 8 of Castle et al. (2012) for DGP IUc. There, somewhat larger potency and gauge is obtained. These small differences are due to the application of *Autometrics*.

Finally a comparison of the results in table D.2 for $DGP1$ and $DGP2$ gives an idea of the consequences of wrongly specifying the dynamics in equation for IIS[16]. Potency differences are significant in all cases, and as will be discussed in section 5.2.2, this specially affects the performance of the pairwise strategy since we are making the largest errors in the most important series. The use of *Autometrics*, that jointly selects the interventions and the dynamic structure, will reduce these gaps between $DGP1$ and $DGP2$ and improve the performance of our proposal.

5.2 The behavior of the pairwise strategy with outlier correction

In this subsection we replicate the simulations of section 3.2 but considering the strategy for dealing with outliers presented in section 4.2.

5.2.1 Design of the experiments

- For time reasons, we do not reproduce the experiments for all scenarios and sample sizes; we focus just on scenarios 1 and 3 ($n_1 = 10$ and $n_1 = 25$ respectively) and $T = 200$.

- To study the behavior of the pairwise strategy under the presence of outliers we contaminate the series used in section 3 with 1, 2 or 3 level shifts of size $\gamma = 4\sigma$ or $\gamma = 5\sigma$, with σ being the standard error for the univariate model of the uncontaminated series. For doing this we proceed as follows:

  - Apply IIS with $\alpha = 1/T$ to the first differences of the 130 basic components of the US CPI and store the percentage of series that have at least 1, 2 or 3 level breaks. The results were $[87\%, 65\%, 42\%]$ for 1 ,2 and 3 outliers respectively.
  - Take the simulated series $(x_{i,t})$ of section 3 and contaminate the corresponding percentage as:

$$y_{i,t} = x_{i,t} + \Gamma LS_{t^*},$$

[16] Recall that for $DGP1$ the lag length is unknown (and we fix it at 3) whereas for $DGP1$ the true lag length is 1, and we use this information.

with Γ and LS_{t^*} being $(1 \times q)$ and $(q \times 1)$ vectors respectively, where q takes the values 1, 2 or 3. The dates of the breaks (t^*) are randomly set in the interval $[20, 180]$, but this random position is maintained for all the 1000 replicas.

  - The percentage of contaminated series is maintained for series inside and outside n_1.

- With the contaminated series we replicate the simulations of section 3 in three different ways:

  - *True Dates*: the true outliers' position is used to apply the pairwise strategy with outlier treatment.
  - *Estim Dates*: outliers' positions are estimated using *1-cut* IIS with the consecutive correction.
  - *No Corr*: the original pairwise procedure (with no outlier treatment) is applied to the contaminated series.

- Finally, we also apply the pairwise procedure with outlier treatment to the uncontaminated series.

5.2.2 Results

Table 5.1 includes the *gauge* and *potency* of the pairwise procedure with outlier treatment. Four main conclusions emerge form the table:

- Using the outlier correction treatment when it is not required (block *Number of breaks: 0* of the table), somewhat deteriorates the procedure's performance with respect to not correcting for outliers (the case of section 3). While when correcting for outliers we include on average $0.917 \times 10 = 9.2$ and $0.885 \times 25 = 22.1$ correct series in scenarios 1 and 3 respectively, if no correction is used these figures increase to $0.969 \times 10 = 9.7$ and $0.956 \times 25 = 23.9$ (see table 3.4).

- When using the *True Dates* with only one break per series the procedure's performance is similar to the the case with no breaks (compare with table 3.4).

- Though *gauge* remains at very low levels, *potency* deteriorates with the number of outliers, even when using the *True Dates*.

- When using the *Estim Dates* -instead of the *True Dates*- significant *potency* reductions are observed. But recall that since we are not selecting the lag length for the series in n_1 -which are the most relevant in determining our procedure's potency- potencies under columns *Estim Dates* are just lower bounds for the ones that will be obtained when using *Autometrics*. This is more relevant for scenario 3, as the proportion of series in n_1 is larger. Table D.3 of the Appendix -which describes the performance in the *Estimating Dates* stage- confirms this observation; the average potency (over the 100 series) of *1-cut* IIS is systematically lower in scenario 3 than in scenario 1.

- The comparison between columns under *Estim Dates* and those under *No Corr* shows that -although we have only lower bound potencies for *Estim Dates*- estimating the dates is better than nothing. This conclusion is not true when the number of breaks is low (see block *Number of breaks: 1*). Additionally, we find that *No Corr* outperforms *Estim Dates* for two breaks and $\gamma = 4$ in scenario 3. This last result reinforces the argument that not selecting the dynamic structure is more damaging for scenario 3, where the proportion of series in n_1 is larger.

Table 5.1: Gauge and Potency of the pairwise procedure with outlier treatment

	Scenario 1, ($n_1 = 10$)						**Scenario 3**, ($n_1 = 25$)					
	True dates		**Estim dates**		**No corr**		**True dates**		**Estim dates**		**No corr**	
	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot
Number of breaks: 0												
			0.3	91.7					0.2	88.5		
Number of breaks: 1												
γ =4	0.3	95.4	0.3	85.6	0.3	92.4	0.2	93.2	0.2	78.5	0.2	88.1
γ =5	0.3	95.3	0.3	85.0	0.3	88.3	0.2	93.1	0.2	78.3	0.2	82.3
Number of breaks: 2												
γ =4	0.3	91.2	0.3	77.0	0.3	75.6	0.2	88.6	0.2	70.2	0.2	72.9
γ =5	0.3	90.4	0.4	77.2	0.4	64.5	0.2	87.8	0.2	70.6	0.2	63.0
Number of breaks: 3												
γ =4	0.3	87.5	0.4	74.5	0.4	72.3	0.2	82.7	0.2	69.1	0.2	68.5
γ =5	0.3	86.2	0.3	78.6	0.4	66.7	0.2	81.4	0.2	73.1	0.2	64.2

- *True Dates*: the true outliers' position is used to apply the pairwise strategy with outlier treatment.
- *Estim Dates*: outliers' positions are estimated using IIS with the consecutive correction.
- *No Corr*: the original pairwise procedure (with no outlier treatment) is applied to the contaminated series.
- Figures under columns *Estim Dates* have to be interpreted as a worst possible case because we are not selecting the dynamic structure for series in n_1, what deteriorates IIS's potency. Note that series in n_1 are in fact the most important series to determine the potency of the pairwise strategy.

6 Detailed algorithm of the Pairwise procedure

The procedure involves eleven steps:

i. Recall that we are assuming that the series are at most I(1), the first step is checking that series are not I(2). Individual unit root tests can be used in this step.

ii. Perform Johansen cointegration tests between all possible pairs of components and store the resulting p-values. When dealing with macroeconomic variables the most general case is that they show systematic growth, so some procedure for dealing with deterministic terms should be considered. When dealing with prices, the inclusion of a linear trend in the cointegration relationship should be considered with extreme caution since forecasts will show a price systematically increasing over the other. Unless there are strong theoretical foundations for such a forecast, our suggestion is not to consider cointegration relationships including linear trends. Then, we proceed as follows:

- Estimate all pairwise VEqCM models (under the null of $r = 1$) including a trend in the cointegrating space.
- Test the significance of the trend and disregard that pair as being cointegrated if the trend is required.
- For the pairs that do not require a trend test for cointegration not including the trend.

Note that this strategy is not exactly the one suggested by Juselius (2006). Based on the results in Nielsen and Rahbek (2000) about asymptotic similarity in cointegration tests, Juselius (2006)

proposes a procedure in which, in a first step the cointegration rank is determined by allowing all the deterministic components included in the model (constant, trends and interventions) to be in the cointegrating relations and its differences in the VAR. Once the cointegration rank is determined, hypothesis over deterministic parameters can be tested.Nielsen and Rahbek (2000) show that this way of proceeding produces asymptotic similarity with respect to deterministic components in the rank tests (see also Doornik et al. (1998)). The reason for this slight is that we are interested in testing cointegration only if the model does not require a trend in the cointegrating space, cases with trend in the cointegration relationship would not be of interest.

iii. Construct a $N \times N$ boolean adjacency matrix, A, that contains a 1 if the corresponding pair is cointegrated and zero otherwise.

iv. Find the *maximal clique* on A using, for example, the Bron-Kerbosch algorithm (see Bron and Kerbosch (1973)). The *maximal clique* is defined as the largest sub-graph in which all nodes are pairwise connected (see also, Bollobás and Erdös (1976)). We rename these maximal cliques as *fully cointegrated subsets*

v. As analyzed in section 3.3, in relatively short samples it may be desirable to relax the requirement of *full cointegration* and let components that are cointegrated with *almost all* the others to enter the subset. Call this new set *almost fully cointegrated*. If the user does not want to consider this relaxation, in point iv instead of finding just the the largest clique, all independent cliques should be found, and the procedure ends here. Otherwise, continue up to point ix.

vi. Define the relaxation parameter ($1 \leqslant \lambda < \hat{n}_1$) to identify the candidates to enter the *almost fully cointegrated* set. A series outside the original set is a candidate if it satisfies two conditions:

   (a) cointegration - at the original 1% of confidence - is rejected with at most λ of the series already in the set.

   (b) when the nominal size of the cointegration test is relaxed to 5% the candidate is cointegrated with all the series already in the set.

vii. Construct the set of candidates C_0. If all the candidates are pairwise cointegrated (at the original 1%), let all of them in and go to point x (because there are not more potential candidates).

viii. If not, find the maximal clique (see point iv) inside C_0 and let in all the series in the maximal clique. Note that after including these series there could still remain some potential candidates, so check for this possibility, construct a new set of candidates C_1, and go to previous point.

ix. If there are not cointegrated candidates, let them in sequentially starting with the one which is cointegrated with most components of series already in the set. In case of conflict (there are candidates that are cointegrated with the same number of variables already in the set), use the p-values stored in step i to decide. An adhoc criteria could be, for example to include the series whose sum of p-values for the null $r = 0$ in cointegration tests with the series already in (or with the ones for which cointegration was rejected) is the maximum. Other adhoc possibility could be to include the series whose sum of p-values for the null $r = 1$ in cointegration tests with the series already in (or with the ones for which cointegration was rejected) is the minimum.

x. Repeat steps iv to ix but excluding the series already included in some almost fully cointegrated set.

xi. Once the disaggregation map is obtained, the forecasting equations can be constructed (see section 2.1.3).

Figure 6.1 summarizes the algorithm in five basic steps.

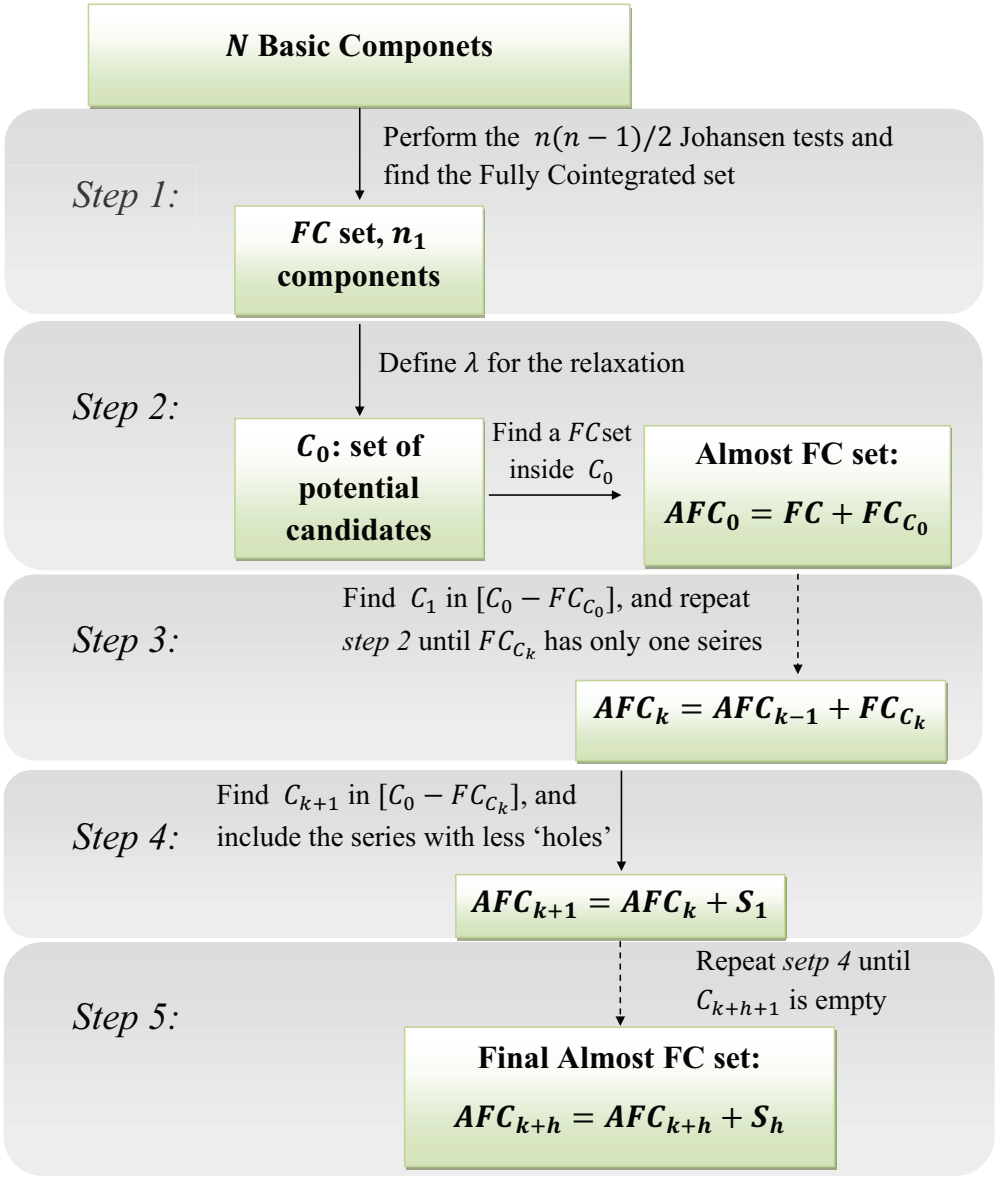


Figure 6.1: Sketch diagram for the algorithm to construct the set of components with a common trend

7 Concluding Remarks

In this paper we studied the properties of a pairwise procedure for testing cointegration between all possible pairs of an aggregate's components. This procedure allows to discover blocks of series that share a unique common trend. 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 these pairs of components.

For the components outside that block, we provided upper bounds for the expected number of series to be -wrongly - included in the estimated sets of series that share the trend. These bounds imply already tolerable expected errors.

In a Monte Carlo experiment we confirmed the asymptotic results, studied the small sample properties of the procedure and compared its performance with a DFM alternative.

We observed a relevant deterioration of the pairwise approach as T decreases, hence a small samples correction was proposed. We studied this correction by simulation and found that it provides a significant *potency*[17] improvement, at the cost of a somewhat larger *gauge*[18]. This trade-off can be managed by what we called the *relaxation parameter* (λ). The larger λ, the larger the *potency* improvements and cost in terms of *gauge*. The user can select λ according to her preferences.

The comparison with DFM showed that the pairwise procedure dominates in situations where the number of series that share the trend (n_1) is relatively small; the DFM alternative fails in those situations. For relatively large n_1 and T the DFM alternative may be preferred, though the pairwise approach also performs very good in those scenarios.

Other relevant results are those related to the specification of the bi-variate sub-models in which cointegration tests are performed. First, we argued that changing the normalization of the cointegrating matrix may lead to an improvement in the forecasting accuracy of the individual models for the components. This potential improvement is derived form a possible reduction in the number of regressors -which does not cause informational looses- and hence in the estimation uncertainty of those models. Second, we established the conditions under which the forecasting strategy in Espasa and Mayo-Burgos (2013) is correct and proposed a generalization of that strategy that does not relay on those stringent conditions. Our proposal is to consider all the $n_1 - 1$ cointegration relationships as potential regressors for all the forecasting equations and select the significant ones applying a selection algorithm. 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. As showed in table 2.1 these improvements are remarkable in many situations.

A strategy for dealing with outliers in the context of the pairwise procedure was also proposed. Briefly, it combines the IIS methodology (see Santos et al. (2008)) with the feasible GLS procedure proposed by Saikkonen and Lütkepohl (2000) to test cointegration in multivariate systems without the need of simulating critical values. Outliers' dates are estimated by IIS and then these dates are used in the GLS procedure. This combination requires picking out 'genuine' outliers from those that belong to the model's dynamics, and we developed a strategy for doing this.

The outliers' correction strategy is analyzed in a simulation study. We found that using the treatment for outliers when it is not required deteriorates the procedure's performance but not dramatically. Additionally, when outliers' dates are known, the pairwise approach behaves similarly to the the case with no outliers (and no treatment).

When dates have to be estimated relevant *potency* reductions are observed. In relation with this issue we highlighted the importance of correctly specifying the dynamic structure of the models in which the outliers' dates are estimated. Since, for simplicity, we did not selected the lag structure in the IIS regressions, these results about *potency* were considered as lower bounds for the ones that will be obtained when selecting the lags. This is particularly relevant in scenarios with large n_1 since, as we argued, the lag structure of the univariate models for these series is more complex.

Comparing the results of correcting *vs.* not correcting for outliers in contaminated series we found

[17] The retention frequency of relevant variables.
[18] The retention frequency of irrelevant variables.

that, although we have only lower bound potencies for the cases of estimated dates, estimating the dates is generally better than doing nothing, except when there is only one outlier.

In current work we are extending the results in four directions. First, all Monte Carlo experiments are to be extended to consider forecasting performance. Second, an empirical application will be carried out. Third, adaptations of the pairwise approach for dealing with relative magnitudes (e.g. relative prices) are to be designed. Finally, other types of common features are to be considered.

References

Altissimo, F., B. Mojon, and P. Zaffaroni (2007). Fast micro and slow macro: Can aggregation explain the persistence of inflation? *Federal Reserve Bank of Chicago WP 2007-02*.

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

Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business and Economic Statistics 11*(1), 81–92.

Banerjee, A. and M. Marcellino (2009). Factor-augmented error correction models. *In Castle, and Shephard (2009), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Oxford University Press*.

Beck, G., K. Hubrich, and M. Marcellino (2011). On the importance of sectoral and regional shocks for price-setting. Working Paper Series 1334, European Central Bank.

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)*, 289–300.

Bils, M. and P. J. Klenow (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy 112*(5).

Boivin, J., M. P. Giannoni, and I. Mihov (2009). Sticky prices and monetary policy: Evidence from disaggregated us data. *The American Economic Review 99*(1), 350–384.

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.

Bosco, B., L. Parisio, M. Pelagatti, and F. Baldi (2007). A robust multivariate long run analysis of european electricity prices. *Nota di Lavoro 103*.

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

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.

Chen, C. and L.-M. Liu (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association 88*(421), 284–297.

Clark, T. (2000). Forecasting an aggregate of cointegrated disaggregates. *Journal of Forecasting 19*(1), 1–21.

Clark, T. E. (2006). Disaggregate evidence on the persistence of consumer price inflation. *Journal of Applied Econometrics 21*(5), 563–587.

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

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

Espasa, A. and R. Albacete (2007). Econometric modelling for short term inflation forecasting in the euro area. *Journal of Forecasting 26*(5), 303–316.

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

Espasa, A., E. Senra, and R. Albacete (2002). Forecasting inflation in the european monetary union: A disaggregated approach by countries and by sectors. *The European Journal of Finance 8*(4), 402–421.

Franses, P., T. Kloek, and A. Lucas (1998). Outlier robust analysis of long-run marketing effects for weekly scanning data. *Journal of Econometrics 89*(1), 293–315.

Franses, P. and A. Lucas (1998). Outlier detection in cointegration analysis. *Journal of Business and Economic Statistics 16*(4), 459–468.

Franses, P. H. and N. Haldrup (1994). The effects of additive outliers on tests for unit roots and cointegration. *Journal of Business and Economic Statistics 12*(4), 471–478.

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

Gómez, V. and A. Maravall (1997). Programs tramo and seats: Instructions for the user (beta version: June 1997) working paper 97001, ministerio de economía y hacienda, dirreccion general de análisis y programacíon presupuestaria, madrid.

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

Hendry, D. and K. Hubrich (2005). Forecasting aggregates by disaggregates. *Computing in Economics and Finance 2005 270*.

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

Hendry, D. F. (1995). On the interactions of unit roots and exogeneity. *Econometric reviews 14*(4), 383–419.

Imbs, J., H. Mumtaz, M. O. Ravn, and H. Rey (2005). Ppp strikes back: Aggregation and the real exchange rate. *The Quarterly Journal of Economics 120*(1), 1–43.

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.

Johansen, S., R. Mosconi, and B. Nielsen (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *The Econometrics Journal 3*(2), 216–249.

Johansen, S. and B. Nielsen (2009). An analysis of the indicator saturation estimator as a robust regression estimator. *In Castle, and Shephard (2009)* (08-03).

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

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics 18*(3), 337–349.

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.

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.

Lucas, A. (1997). Cointegration testing using pseudolikelihood ratio tests. *Econometric theory*, 149–169.

Lucas, A. (1998). Inference on cointegrating ranks using lr and lm tests based on pseudo-likelihoods. *Econometric reviews 17*(2), 185–214.

Lunnemann, P. and T. Mathä (2004). How persistent is disaggregate inflation? an analysis across eu 15 countries and hicp sub-indices.

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

Nielsen, B. (2004). On the distribution of likelihood ratio test statistics for cointegration rank. *Econometric reviews 23*(1), 1–23.

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

Nielsen, H. (2008). Influential observations in cointegrated var models: Danish money demand 1973–2003. *The Econometrics Journal 11*(1), 39–57.

Peña, D., G. Tiao, and R. Tsay (2001). *A course in time series analysis*. Wiley New York.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: Journal of the Econometric Society*, 1361–1401.

Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics 1*, 278–352.

Perron, P. and G. Rodríguez (2003). Searching for additive outliers in nonstationary time series*. *Journal of Time Series Analysis 24*(2), 193–220.

Perron, P. and T. J. Vogelsang (1992). Nonstationarity and level shifts with an application to purchasing power parity. *Journal of Business and Economic Statistics 10*(3), 301–320.

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

Quintos, C. E. (1998). Stability tests in error correction models. *Journal of Econometrics 82*(2), 289–315.

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

Saikkonen, P. and H. Lutkepohl (2002). Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process. *Journal of Time Series Analysis 21*(4), 435–456.

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

Stock, J. and M. Watson (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and Banking 39*, 3–33.

Trenkler, C., P. Saikkonen, and H. Lütkepohl (2007). Testing for the cointegrating rank of a var process with level shift and trend break. *Journal of Time Series Analysis 29*(2), 331–358.

Tsay, R. S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association 81*(393), 132–141.

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

Vogelsang, T. J. (1999). Two simple procedures for testing for a unit root when there are additive outliers. *Journal of Time Series Analysis 20*(2), 237–252.

Appendix A Detailed algebra for the two DGPs

A.1 Proof that all components in n_1 are cointegrated with the sub-aggregate

Assume that all the components are I(1) and share a unique common trend. As showed in the main text cointegration vectors can always be normalized as $\beta'_{r\times N} = [\tilde{\beta}'_{r\times(N-r)}, I_{r\times r}]$. Then, without loss of generality, the $r = N-1$ cointegration relationships can be written as,

$$\begin{bmatrix} \tilde{\beta}_2 & 1 & 0 & 0 & \dots & 0 \\ \tilde{\beta}_3 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ \tilde{\beta_N} & 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{N,t} \end{bmatrix} = \begin{bmatrix} e_{2,t} \\ e_{3,t} \\ \vdots \\ e_{N,t} \end{bmatrix}, \tag{A.1}$$

where e_{it} are stationary processes.

Assuming that the weights in the aggregate are constant (this seems quite a strong assumption since weights of macro-variables are not exactly constant), the aggregate A_t is

$$A_t = \sum_{i=1}^{N} w_i X_{it}, \tag{A.2}$$

with w_i being the components' weights [19]. Then we have that,

$$A_t = w_1 X_{1t} + \textstyle\sum_{i=2}^{N} w_i X_{it} = w_1 X_{1t} + \sum_{i=2}^{N} w_i(-\tilde{\beta}_i X_{1t} + e_{it})$$

$$A_t = X_{1t}(w_1 - \textstyle\sum_{i=2}^{N} w_i \tilde{\beta}_i) + \sum_{i=2}^{N} w_i e_{it} = X_{1t}\beta_1^* + \sum_{i=2}^{N} w_i e_{it}$$

then,

$$X_{1t} = \frac{1}{\beta_1^*} A_t - \frac{1}{\beta_1^*} \sum_{i=2}^{N} w_i e_{it},$$

so that,

$$X_{1t} = \frac{1}{\beta_1^*} A_t + e_{1t}^* \tag{A.3}$$

where $\beta_1^* = w_1 - \sum_{i=2}^{N} w_i \tilde{\beta}_i$ and $e_{1t}^* = -\frac{1}{\beta_1^*} \sum_{i=2}^{N} w_i e_{it}$.

Since all the e_{it} are stationary, e_{1t}^* is also stationary and X_1 is cointegrated with the aggregate.

We now show that all the other components are also cointegrated with the aggregate. Using (A.1) and A.3 we can write:

$$\begin{aligned} X_{2t} &= -\tilde{\beta}_2 X_{1t} + e_{2t} = -\tfrac{\tilde{\beta}_2}{\beta_1^*} A_t - \tilde{\beta}_2 e_{1t}^* + e_{2t} = \beta_2^* A_t + e_{2t}^* \\ X_{3t} &= -\tilde{\beta}_3 X_{1t} + e_{3t} = -\tfrac{\tilde{\beta}_3}{\beta_1^*} A_t - \tilde{\beta}_3 e_{1t}^* + e_{3t} = \beta_3^* A_t + e_{3t}^* \quad , \\ &\vdots \\ X_{Nt} &= -\tilde{\beta_N} X_{1t} + e_{Nt} = -\tfrac{\tilde{\beta_N}}{\beta_1^*} A_t - \tilde{\beta_N} e_{1t}^* + e_{Nt} = \beta_n^* A_t + e_{Nt}^* \end{aligned} \tag{A.4}$$

[19] Note that we are calling *aggregate* to the series that results of aggregating all the components in n_1

where $e^*_{it} = -\tilde{\beta}_i e^*_{1t} + e_{it}$, and $\beta^*_i = -\frac{\tilde{\beta}_i}{\beta^*_1}$. Since e^*_{1t} and e_{it} are stationary, e^*_{it} is also stationary and X_{it} $(\forall i)$ is cointegrated with the aggregate.

In summary we showed that, provided that (A.2) holds, cointegration system (A.4) is equivalent to the original one. Note that we have assumed that the aggregate is non-stationary, what implies assuming that the vector of weights is not a cointegration vector.

A.2 Derivation of the aggregate's dynamics in model *b*

In this Appendix we aggregate's dynamics when the components follow the model b described in point ii of section 2.1.1. Recall from that section that in model b we assume that the equilibrium correction reactions of each variable are due to deviations of their *own* cointegration relationships with respect to the aggregate. Since the aggregate is not necessarily weakly exogenous in this model, we derive both its equilibrium correction and its short run dynamics.

We first note that,

$$\Delta A_t = w_1 \Delta X_{1t} + W_{nf} \Delta X_{nf,t},$$

where W_nf is a row vector containing all weights except the first one, and $\Delta X_{nf,t}$ is a column vector containing the first differences of all variables except the first one.

Let α^*_i denote component's i^{th} adjustment speed for $i = 1, ..., N$ (note that these N coefficients are to be set as parameters in the DGP simulation). Then:

$$\Delta A_t = W_{nf} \boldsymbol{\alpha}^*_{nf} e^*_{nf,t-1} + w_1 \alpha^*_1 e^*_{1,t-1} + W_{nf} \Phi_{nf} \Delta X_t + w_1 \Phi_1 \Delta X_t + W_{nf} \epsilon_{nf,t} + w_1 \epsilon_{1,t}$$

Call:

$$\begin{aligned} \Delta^E A_t &= W_{nf} \boldsymbol{\alpha}^*_{nf} e^*_{nf,t-1} + w_1 \alpha^*_1 e^*_{1,t-1}, \quad and \\ \Delta^{sr} A_t &= W \Phi \Delta X_{t-1} + W \epsilon_t, \end{aligned} \tag{A.5}$$

so that $\Delta A_t = \Delta^E A_t + \Delta^{sr} A_t$.

Given an election of Φ, $\Delta^{sr} A_t$ is directly given by (A.5). It remains to find $\Delta^E A_t$.

Recall from (A.3) that $e^*_{1,t-1} = X_{1,t-1} - \frac{1}{\beta^*_1} A_{t-1}$, and note that $w_1 X_{1,t-1} = A_{t-1} - W_{nf} X_{nf,t-1}$. Then, substituting in $\Delta^E A_t$:

$$\Delta^E A_t = W_{nf} \boldsymbol{\alpha}^*_{nf} e^*_{nf,t-1} + \alpha^*_1 (A_{t-1} - W_{nf} X_{nf,t-1} - \frac{w_1}{\beta^*_1} A_{t-1})$$

$$\Delta^E A_t = W_{nf} \boldsymbol{\alpha}^*_{nf} e^*_{nf,t-1} + \alpha^*_1 [A_{t-1}(1 - \frac{w_1}{\beta^*_1}) - W_{nf} X_{nf,t-1}] \tag{A.6}$$

In Appendix A.1 we stated:

$$\begin{cases} \beta^*_1 = w_1 - W_{nf} \tilde{\beta}_i \quad and \\ \beta^*_i = -\dfrac{\tilde{\beta}_i}{\beta^*_1} \end{cases}$$

Hence,

$$A_{t-1}(1-\frac{w_1}{\beta_1^*})-W_{nf}X_{nf,t-1}=A_{t-1}(\frac{w_1-W_{nf}\tilde{\beta}-w_1}{\beta_1^*})-W_{nf}X_{nf,t-1}$$

$$=A_{t-1}(\frac{-W_{nf}\tilde{\beta}}{\beta_1^*})-W_{nf}X_{nf,t-1}$$

$$=A_{t-1}W_{nf}\beta_{nf}^*-W_{nf}X_{nf,t-1}$$

$$=-W_{nf}(X_{nf,t-1}-\beta_{nf}^*A_{t-1})$$

$$=-W_{nf}e_{nf}^*$$

Plugging this expression in (A.6) we get:

$$\Delta^E A_t=W_{nf}\boldsymbol{\alpha}_{nf}^*e_{nf,t-1}^*-W_{nf}\alpha_1^*e_{nf,t-1}^*=W_{nf}(\boldsymbol{\alpha}_{nf}^*-\alpha_1^*I_r)e_{nf,t-1}^* \tag{A.7}$$

If we want to set the same speed adjustment for all the components, we need $\alpha_1^*=\alpha_2^*=...=\alpha_n^*$. Then, from (A.7), we get:

$$\Delta^E A_t=0 \tag{A.8}$$

Appendix B Proof of theorem 2.1

Proof:

The first step in Johansen's procedure is concentrating the model with respect to $\alpha\beta'$, what is done by regressing ΔX_t and X_{t-1} on $(\Delta X_{t-1},...,\Delta X_{t-k+1})$. These 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 of the problem:

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

whose eigenvalues are the solution of:

$$|\lambda S_{11}-S_{10}S_{00}^{-1}S_{01}|=0, \tag{B.2}$$

Let $\boldsymbol{X}_t$ be the vector containing the series in n_1, from the Granger Representation Theorem, disregarding deterministic terms, the cointegrated VAR can be written as;

$$\boldsymbol{X}_t=\boldsymbol{X}_0+C(1)\sum_{i=1}^{t}\epsilon_t+C^*(L)\epsilon_t,$$

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 the individual series in n_1 can be written as:

$$X_{mt} = \delta_m CT_t + w_{mt} \tag{B.3}$$

where CT_t is a $I(1)$ process and w_{it} are stationary ones. From (B.3) any series in n_1 can be expressed as $X_{mt} = \gamma_{mr} X_{rt} + \eta_{mr,t}$, with $\eta_{it} \sim I(0)$.

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

First write $X_{it} = \gamma_{i1} X_{1t} + \eta_{i1,t}$, and $X_{jt} = \gamma_{j2} X_{2t} + \eta_{j2,t}$ and call $X_t^* = [X_{it}, X_{jt}]'$. The auxiliary regressions are now:

$$\begin{array}{c} \Delta X_t^* = B_0^* \Delta \tilde{X}_{t-1}^* + R_{0t}^* \\ X_{t-1}^* = B_1^* \Delta \tilde{X}_{t-1}^* + R_{1t}^*, \end{array} \tag{B.4}$$

where $\Delta \tilde{X}_{t-1}^* = [\Delta X_{t-1}^{*\prime}, \Delta X_{t-2}^{*\prime}, ..., \Delta X_{t-p^*}^{*\prime}]$, and B_i^* is a $2 \times p^*$ matrix of coefficients.

Note that (B.4) can be written as

$$\begin{array}{c} \Gamma \Delta X_t + \Delta \eta_t = B_0^* (\Gamma \Delta \tilde{X}_{t-1} + \Delta \tilde{\eta}_{t-1}) + R_{0t}^* \\ \Gamma X_{t-1} + \eta_{t-1} = B_1^* (\Gamma \Delta \tilde{X}_{t-1} + \Delta \tilde{\eta}_{t-1}) + R_{1t}^*, \end{array} \tag{B.5}$$

where $\Gamma = \begin{bmatrix} \gamma_{i1} & 0 \\ 0 & \gamma_{j2} \end{bmatrix}$, and $\Delta \tilde{\eta}_{t-1} = [\Delta \eta'_{t-1}, \Delta \eta'_{t-2}, ..., \Delta \eta'_{t-p^*}]$.

Let $B_i^* = \Gamma B_i \Gamma^{-1} + b_i$, with b_i an adequate $2 \times p^*$ matrix. To save notation, we assume $p^* = p$, but the argument only requires $p^* \geq p$. Equations in (B.5) become:

$$\begin{array}{c} \Gamma \Delta X_t + \Delta \eta_t = \Gamma B_0 \Delta \tilde{X}_{t-1} + b_0 \Gamma \Delta \tilde{X}_{t-1} + B_0^* \Delta \tilde{\eta}_{t-1} + R_{0t}^* \\ \Gamma X_{t-1} + \eta_t = \Gamma B_1 \Delta \tilde{X}_{t-1} + b_1 \Gamma \Delta \tilde{X}_{t-1} + B_1^* \Delta \tilde{\eta}_{t-1} + R_{1t}^* \end{array} \tag{B.6}$$

Recall that the original R_{it} are $R_{0t} = \Delta X_t - B_0 \Delta \tilde{X}_{t-1}$ and $R_{1t} = X_{t-1} - B_1 \Delta \tilde{X}_{t-1}$. Then, solving (B.6) for R_{it}^*:

$$\begin{array}{c} R_{0t}^* = \Gamma R_{0t} + \epsilon_{0t} \\ R_{1t}^* = \Gamma R_{1t} + \epsilon_{1t} \end{array}, \tag{B.7}$$

where, ϵ_{0t} and ϵ_{1t} are the following stationary processes:

$$\begin{array}{c} e_{0t} = -[b_0 \Gamma \Delta \tilde{X}_{t-1} + B_0^* \Delta \tilde{\eta}_{t-1} - \Delta \eta_t] \\ e_{1t} = -[b_1 \Gamma \Delta \tilde{X}_{t-1} + B_1^* \Delta \tilde{\eta}_{t-1} - \eta_{t-1}] \end{array}, \tag{B.8}$$

Hence, the new (2×2) matrices S_{ij}^* are:

$$S_{ij}^* = T^{-1} (\Gamma R_i + \epsilon_i)(\Gamma R_j + \epsilon_j)' \tag{B.9}$$

Then;

$$S_{11}^{*} = T^{-1}[\Gamma R_1 R_1'\Gamma' + \Gamma R_1 \epsilon_1' + \epsilon_1 R_1'\Gamma' + \epsilon_1\epsilon_1'] \tag{B.10}$$

In (B.10) all terms in the right hand side 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; $S_{11} \to \Gamma R_1 R_1'\Gamma'$ as $T \to \infty$. The remaining S_{ij}^{*} are $Op(1)$ and can be written as

$$S_{ij}^{*} = \Gamma S_{ij}\Gamma + \Omega_{ij}, \text{ for } (i,j) \neq (1,1), \tag{B.11}$$

where Ω_{ij} is $Op(1)$.

The new eigenvalue problem is:

$$(S_{10}^{*} S_{00}^{*-1} S_{01}^{*})v^{*} = \lambda^{*} S_{11}^{*} v^{*} \tag{B.12}$$

Using (B.10) and (B.11) we get:

$$[(\Gamma S_{10}\Gamma' + \Omega_{10})(\Gamma S_{00}\Gamma' + \Omega_{00})^{-1}(\Gamma S_{01}\Gamma' + \Omega_{01})]v^{*} = \lambda^{*}(\Gamma S_{11}\Gamma')v^{*} \tag{B.13}$$

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

$$(\Gamma S_{00}\Gamma' + \Omega_{00})^{-1} = c(\Gamma S_{00}\Gamma')^{-1} + \tilde{\Omega}_{00} = c\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \tilde{\Omega}_{00} \tag{B.14}$$

where $c = \dfrac{|\Gamma S_{00}\Gamma'|}{|\Gamma S_{00}\Gamma' + \Omega_{00}|} \leq 1$, and $\tilde{\Omega}_{00} = \dfrac{1}{|\Gamma S_{00}\Gamma' + \Omega_{00}|} Adj[\Omega_{00}]$.

Hence, plugging (B.14) into (B.13) we get:

$$[(\Gamma S_{10}\Gamma' + \Omega_{10})(c\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \tilde{\Omega}_{00})(\Gamma S_{01}\Gamma' + \Omega_{01})]v^{*} = \lambda^{*}(\Gamma S_{11}\Gamma')v^{*}, \tag{B.15}$$

from where:

$$[c\Gamma S_{10}S_{00}^{-1}S_{01}\Gamma' + (\Gamma S_{01}\Gamma'\tilde{\Omega}_{00} + c\Omega_{10}\Gamma^{-1}S_{00}^{-1}\Gamma^{-1} + \Omega_{10}\tilde{\Omega}_{00})(\Gamma S_{01}\Gamma' + \Omega_{01})]v^{*} = \lambda^{*}(\Gamma S_{11}\Gamma')v^{*}, \tag{B.16}$$

and

$$[c\Gamma S_{10}S_{00}^{-1}S_{01}\Gamma' + \Psi]v^{*} = \lambda^{*}(\Gamma S_{11}\Gamma')v^{*}, \tag{B.17}$$

where,

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

Left multiplying (B.17) by Γ^{-1} we obtain:

$$[cS_{10}S_{00}^{-1}S_{01}\Gamma' + \Gamma^{-1}\Psi]v^{*} = \lambda^{*}(S_{11}\Gamma')v^{*}, \tag{B.18}$$

Let now $\underset{\sim}{\Psi} = \Gamma'\Psi\Gamma^{-1}$, so that (B.18) becomes:

$$[cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi}]\Gamma' v^* = \lambda^* S_{11}\Gamma' v^*, \tag{B.19}$$

Compare (B.19) with (B.1):

- If $X_m \equiv X_1$ and $X_r \equiv X_2$ we get $\Psi = 0$, $c = 1$ and $\Gamma = I$ so we recover the original problem.
- In the extremely unlikely case that $\Omega_{ij} = 0$ we get $\Psi = 0$ and $c = 1$, so that the eigenvalue problem would be

  $$[S_{10}S_{00}^{-1}S_{01}\Gamma']v^* = \lambda^*(S_{11}\Gamma')v^*,$$

  which solution is $\lambda^* = \lambda$ and $v^* = \Gamma' v$. Hence, even in small samples, the cointegration test statistic is exactly the same as the one for the pair (X_1, X_2).
- 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 (B.18) are the solutions of the second order polynomial in λ^*

  $$|\lambda^* S_{11} - (cS_{10}S_{00}^{-1}S_{01} + \underset{\sim}{\Psi})| = 0 \tag{B.20}$$

  Given that S_{11} is $Op(T)$ and the other matrices are $Op(1)$, it can be shown that for $T \to \infty$ $\hat{\lambda}_2 \to 0 \Leftrightarrow \hat{\lambda}_2^* \to 0$. 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 are the roots of the polynomial

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

  where s_{ij} and υ_{ij} are the elements of the matrices S_{11} and Υ respectively.

  Since $B < 0$

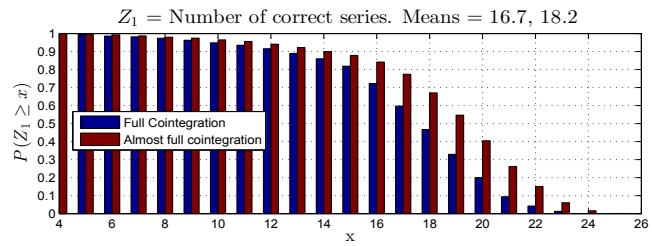
  $$\lambda_2 = \frac{-B - \sqrt{B^2 - 4|S_{11}|C}}{2|S_{11}|} = \frac{G}{2|S_{11}|},$$

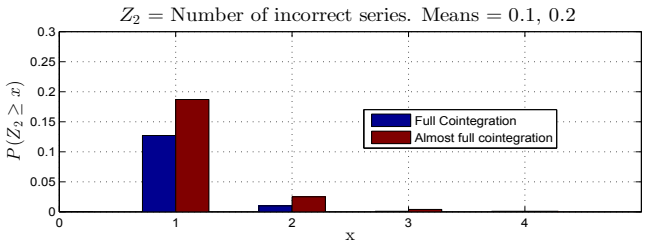
  hence $\lambda_2 \to 0$ if and only if G grows slower than $|S_{11}|$.

  Now, replace θ_{ij} by θ_{ij}^* in (B.21) to get B^*, C^* and G^*. Since θ_{ij} and θ_{ij}^* are $Op(1)$, the asymptotic behavior of G^* is the same as the one of G, thus $|S_{11}|$ grows faster than G if and only if it also grows faster than G^*.

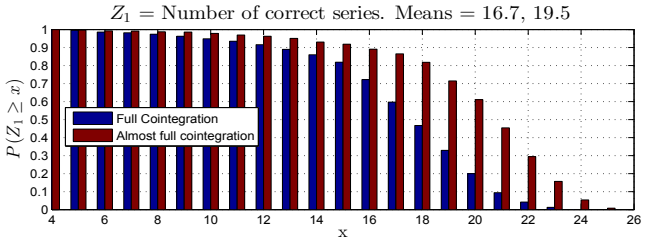
  Then, asymptotically, once we reject two cointegration relationships in one pair, we will reject it in all pairs, *and the proof is complete* ■

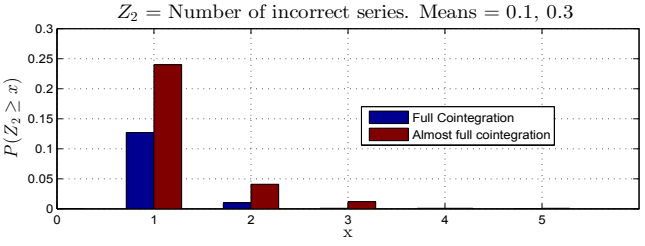
Appendix C Detailed probability functions for the relaxation procedure for scenario 3



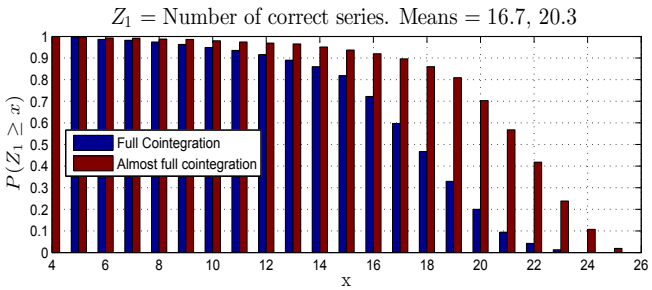


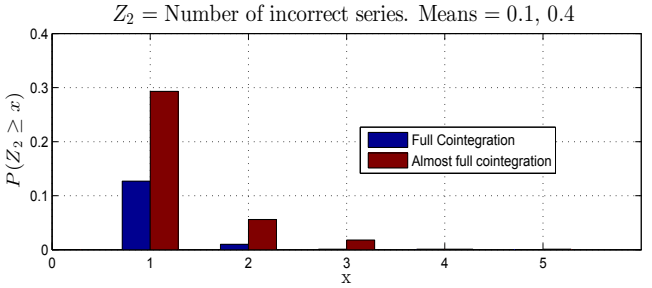
$\lambda = 1$



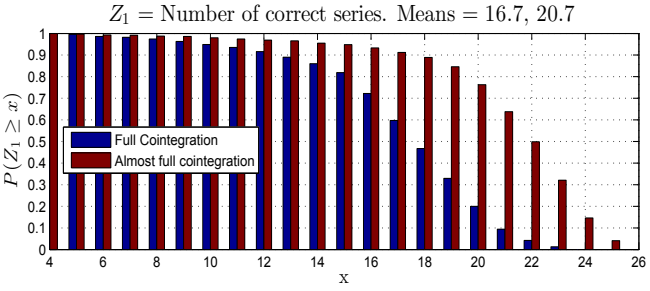


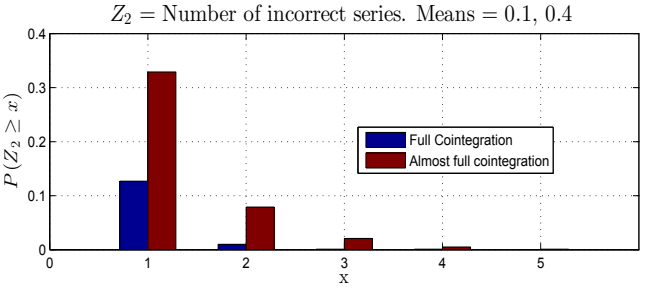
$\lambda = 2$





$\lambda = 3$





$\lambda = 4$

Figure C.1: $P(Z \geq x)$. $T = 100$. Scenario 3 ($n_1 = 25$)

Appendix D Additional tables for section 5

Table D.1: Mean number of retained consecutive outliers that are not *'original'* ones

	$\alpha = 0.005$		$\alpha = 0.01$	
	No Consec Corr	With Consec Corr	No Consec Corr	With Consec Corr
Outliers distribution1 ($LS_{(102)}$)				
DGP1				
γ =4	0.20	0.01	0.36	0.02
γ =5	0.46	0.01	0.63	0.03
DGP2				
γ =4	0.25	0.01	0.38	0.02
γ =5	0.47	0.01	0.58	0.02
Outliers distribution2 ($LS_{(77,177)}$)				
DGP1				
γ =4	0.22	0.01	0.40	0.01
γ =5	0.42	0.01	0.63	0.02
DGP2				
γ =4	0.20	0.01	0.33	0.01
γ =5	0.33	0.00	0.49	0.01
Outliers distribution3 ($LS_{(52,102,152)}$)				
DGP1				
γ =4	0.16	0.01	0.33	0.02
γ =5	0.34	0.01	0.57	0.02
DGP2				
γ =4	0.19	0.01	0.34	0.02
γ =5	0.34	0.00	0.51	0.02
Outliers distribution4 ($LS_{(52,53)}$)				
DGP1				
γ =4	0.03	0.00	0.08	0.01
γ =5	0.04	0.00	0.10	0.01
DGP2				
γ =4	0.01	0.00	0.04	0.01
γ =5	0.01	0.01	0.04	0.02
Outliers distribution5 ($LS_{(52,53,54)}$)				
DGP1				
γ =4	0.12	0.04	0.24	0.08
γ =5	0.30	0.11	0.45	0.19
DGP2				
γ =4	0.14	0.10	0.26	0.15
γ =5	0.27	0.17	0.44	0.27

DGP 1: Series that $\in n_1$ in scenario 3, *DGP 2:* Series not in n_1 for in scenario 3, *DGP 3:* No lags.
γ is the size (in number of the model's standard deviations) of the outliers.
α is the nominal size used in the IIS.
Number of experiments = 1000.
Figures for *DGP*3 are excluded because they are always 0 due to the absence of dynamic structure.

Table D.2: Gauge and Potency of *1-cut* IIS

	No consecutive Correction				**With consecutive correction**			
	$\alpha = 0.005$		$\alpha = 0.01$		$\alpha = 0.005$		$\alpha = 0.01$	
	Gauge	Pot	Gauge	Pot	Gauge	Pot	Gauge	Pot
	Outliers distribution0 *No breaks*							
DGP1								
γ =4	0.39	—	0.85	—	0.38	—	0.83	—
γ =5	0.39	—	0.85	—	0.38	—	0.83	—
DGP2								
γ =4	0.43	—	0.88	—	0.43	—	0.87	—
γ =5	0.43	—	0.88	—	0.43	—	0.87	—
DGP3								
γ =4	0.44	—	0.93	—	0.44	—	0.93	—
γ =5	0.44	—	0.93	—	0.44	—	0.93	—
	Outliers distribution1 ($LS_{(102)}$)							
DGP1								
γ =4	0.45	57.1	0.87	67.9	0.35	56.9	0.70	67.5
γ =5	0.51	83.4	0.89	89.7	0.29	83.4	0.59	89.6
DGP2								
γ =4	0.44	82.2	0.84	87.8	0.32	82.2	0.67	87.8
γ =5	0.49	97.4	0.86	98.4	0.26	97.3	0.58	98.3
DGP3								
γ =4	0.36	85.7	0.79	90.9	0.36	85.7	0.79	90.9
γ =5	0.33	98.6	0.70	99.3	0.33	98.6	0.70	99.3
	Outliers distribution2 ($LS_{(77,177)}$)							
DGP1								
γ =4	0.42	51.1	0.77	61.3	0.31	51.0	0.58	61.1
γ =5	0.45	73.4	0.74	80.8	0.24	73.4	0.43	80.8
DGP2								
γ =4	0.32	76.0	0.66	83.5	0.23	76.0	0.50	83.4
γ =5	0.30	93.4	0.56	96.4	0.14	93.4	0.32	96.4
DGP3								
γ =4	0.24	80.8	0.54	86.8	0.24	80.8	0.54	86.8
γ =5	0.16	95.3	0.39	97.3	0.16	95.3	0.39	97.3
	Outliers distribution3 ($LS_{(52,102,152)}$)							
DGP1								
γ =4	0.43	48.2	0.78	58.9	0.32	48.1	0.57	58.8
γ =5	0.47	70.6	0.79	79.3	0.24	70.5	0.41	79.2
DGP2								
γ =4	0.32	73.8	0.62	81.9	0.20	73.8	0.41	81.8
γ =5	0.31	91.7	0.54	95.5	0.11	91.7	0.24	95.5
DGP3								
γ =4	0.17	78.3	0.43	85.0	0.17	78.3	0.43	85.0
γ =5	0.11	93.9	0.27	96.6	0.11	93.9	0.27	96.6
	Outliers distribution4 ($LS_{(52,53)}$)							
DGP1								
γ =4	0.41	30.3	0.81	35.9	0.34	30.0	0.70	35.3
γ =5	0.47	43.7	0.85	47.0	0.34	43.3	0.68	46.2
DGP2								
γ =4	0.41	45.3	0.79	49.3	0.33	45.1	0.68	48.5
γ =5	0.47	54.9	0.82	57.9	0.32	54.8	0.66	57.3
DGP3								
γ =4	0.29	85.5	0.65	90.8	0.29	85.5	0.65	90.8
γ =5	0.26	97.8	0.57	98.6	0.26	97.8	0.57	98.6
	Outliers distribution5 ($LS_{(52,53,54)}$)							
DGP1								
γ =4	0.37	20.8	0.77	24.6	0.33	20.4	0.70	23.7
γ =5	0.40	30.0	0.78	33.1	0.32	29.0	0.67	31.6
DGP2								
γ =4	0.43	32.4	0.79	36.2	0.42	31.8	0.76	34.9
γ =5	0.49	39.4	0.82	42.7	0.46	38.0	0.76	41.1
DGP3								
γ =4	0.24	85.0	0.55	90.4	0.24	85.0	0.55	90.4
γ =5	0.22	97.8	0.49	98.8	0.22	97.8	0.49	98.8

DGP 1: Series that $\in n_1$ in scenario 3, *DGP 2:* Series not in n_1 for in scenario 3, *DGP 3:* No lags.
γ is the size (in number of the model's standard deviations) of the outliers.
α is the nominal size used in the IIS.
Number of experiments = 1000

Table D.3: Average Gauge and Potency of *1-cut* IIS applied to the 100 series in each scenario

	Scenario 1, ($n_1 = 10$)		**Scenario 3**, ($n_1 = 25$)	
	Gauge	Pot	Gauge	Pot
Number of breaks: 1				
γ =4	0.38	83.1	0.39	79.1
γ =5	0.38	96.6	0.38	94.4
Number of breaks: 2				
γ =4	0.39	78.6	0.40	74.7
γ =5	0.40	93.9	0.40	91.6
Number of breaks: 3				
γ =4	0.40	73.6	0.39	68.8
γ =5	0.32	90.4	0.32	86.6

- This table describes the performance of *1-cut* IIS when applied to the series used in section 5.2.2.
- Figures are averages across series and experiments.