UNIVERSIDAD CARLOS III DE MADRID working papers

UC3M WorkingPapers
StatisticsandEconometrics
15-18
ISSN 2387-0303
September 2015

DepartamentodeEstadística
UniversidadCarlosIIIdeMadrid
CalleMadrid,126
28903Getafe(Spain)
Fax(34)91624-98-48

Forecasting a large set of disaggregates with common trends and outliers

Guillermo Carlomagno[a], Antoni Espasa[b]

Abstract

This paper deals with macro variables that have a large number of components and our aim is to model and forecasts all of them. We adopt a basic statistical procedure for discovering common trends among a large set of series and propose some extensions to take into account data irregularities and small samples issues. The forecasting strategy consists on estimating single-equation models for all the components, including the restrictions derived from the existence of common trends. An application to the disaggregated US CPI shows the usefulness of the procedure in real data problems.

Keywords: Cointegration, pairwise testing, Disaggregation, Forecast model selection, Outliers treatment, Inflation

[a]Department of Statistics, Universidad Carlos III de Madrid.
[b]Department of Statistics, Instituto Flores de Lemus, Universidad Carlos III de Madrid.

This paper is a more elaborated version of a part of a previous paper distributed as "The pairwise approach a large set of disaggregates with common trends".

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

1 Introduction

The general practice of empirical macroeconomic analysis it to focus just on aggregated variables such as the CPI, the GDP, the Industrial Production Index (IPI), Imports, Exports, Unemployment, etc. As all these variables are constructed by aggregating their components, limiting the analysis just to the aggregate implies denying a great amount of information. The starting point of this paper is that this information should not be disregarded at the outset because it may be of interest for decision makers.

Consider for example the case of central banks whose monetary policy decisions are mainly based on current and expected inflation. An inflation forecast close to the target may trigger very different monetary policy decisions if all the components are expected to grow at the same rate, or if, say, core inflation is above the target and non-core components below it. For this reason, splitting the CPI in core and non-core components has became a widespread practice. Nonetheless, this level of disaggregation is still too limited. It could be the case that the forecasts of discordant inflationary components are limited just to a specific group of goods or services, and ignoring this information could not help to a proper assessment of inflation.

In addition to improving the information for monetary policy decisions, disaggregating the CPI and providing forecasts for the components could be very useful for decision makers in the different sectors of the economy. It also allows to study the interrelations between sectoral prices. Knowing how prices in one industry are related to those of other ones is relevant not only for making decisions at the firm level but also for an economic policy point of view. For example, taxation changes in one sector may affect the price level of that sector but may also have spillover effects on other ones. A disaggregated analysis that studies the interconnection between sectoral prices would be useful for anticipating those spillovers.

The benefits of disaggregating are not limited to the analysis of the components. On the one hand, it is useful to understand the dynamic properties of the aggregate generated by the statistical effects of the aggregation process (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), and Beck et al. (2011)). On the other hand, there is increasing agreement in the literature that the use of disaggregated information and disaggregated forecasts may lead to more accurate forecasts of the aggregate (see, Espasa et al. (2002), Giacomini and Granger (2004), Hendry and Hubrich (2005), Espasa and Albacete (2007), Castle and Hendry (2010), Hendry and Hubrich (2011)

and Espasa and Mayo-Burgos (2013)). Thus, even if we were interested only in the aggregate, disaggregation may also be important.

In this paper we expand the pairwise approach proposed by Espasa and Mayo-Burgos (2013) and extended by Carlomagno and Espasa (2014) (hereafter *CE*) — who analyzed the statistical properties of the procedure. Our application is devoted to model and forecast all the components of the US CPI. The procedure in *CE* consists of performing Johahnsen's cointegration tests between all the $N(N-1)/2$ pairs that exist among the N components of the CPI, and then, building subsets of all the components such that, in each subset, all its elements are pairwise cointegrated. Those subsets are denoted as *fully cointegrated* and have the property that all the components inside them share a unique common trend. *CE* show the good performance of the procedure to *discover* the true *fully cointegrated groups* when they exist. For forecasting the components *CE*, in line with Espasa and Mayo-Burgos (2013), propose to use single equation models for all of them including as potential regressors the cointegration relationships, if it is the case, own lags, and lags of the other components or some intermediate sub-aggregates. The selection among all these potential regressors can be done with the automatic selection algorithm *Autometrics* (see Doornik (2009)). This strategy is commented but not implemented by the mentioned authors.

As a global measure of the accuracy of the individual forecasts Espasa and Mayo-Burgos (2013) propose to asses the accuracy of the indirect forecast of the CPI obtained by adding up the disaggregated forecasts. Considering also common cycles restrictions, but using simpler approaches for cointegration tests and for specifying the single equation models than those in *CE*, Espasa and Mayo-Burgos (2013) show that the indirect forecasts for the CPI in US, UK and Euro Area outperforms both, direct procedures and other indirect competitors.

Our extensions to the proposal in *CE* are intended to robustify the procedure when applied to real data which may be subject to irregularities and/or to short samples problems. The robustification concerns both, the construction of the *fully cointegrated subsets* and the estimation of the single equation models.

Regarding the *fully cointegrated subsets*, *CE* show that their procedure's potency for discovering the subsets suffers an important deterioration in small samples. In this paper we propose a small samples correction and study its properties by Monte Carlo. Additionally, as it is shown by several authors (see e.g, Johansen et al. (2000), Saikkonen and Lutkepohl (2002) and Nielsen (2004)), cointegration tests are strongly affected by the presence of outliyng observa-

tions. For dealing with this issue we propose a solution that combines the feasible GLS approach in Saikkonen and Lütkepohl (2000) and Saikkonen and Lutkepohl (2002) with the Impulse Indicator Saturation (IIS) methodology in Santos et al. (2008), and study its performance by Monte Carlo.

Apart from robusfying the search of cointegration relationships against data irregularities and short samples issues, we extend the search by the inclusion of weakly exogenous variables and the admission of more general short run dynamics as suggested by Aron and Muellbauer (2013). Additionally, we add some 'quality' controls to the cointegration relationships to be accepted as 'good' ones.

For the single equation models that are used to forecast the components, our strategy is to follow the aforementioned proposal in *CE* extended with IIS.

The rest of the paper is organized as follows. In §2, we describe the outlier correction procedure and study its properties by Monte Carlo. §3 is devoted to the small samples correction procedure. In §4, we include the modifications of the cointegration relationships search, and in §5 we give a detailed description of the pairwise algorithm enlarged with our contributions. Finally, §6 and §7 include the application to the US CPI and the conclusions, respectively.

2 Data irregularities

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 (location shifts), 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)).

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 suffers location shifts but is stationary within regimes. Additionally, when trying to detect a location shift, most tests will reject the null of no break when the process has a unit root with constant parameters. Similarly, as noted by Quintos (1998) and stressed by Perron (2006), tests for location shifts 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 location shifts. Therefore, a circular problem exists when dealing with non-stationary series that may be subject to location shifts.

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 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 propose an empirical strategy for dealing with these issues in the framework of the pairwise approach.

2.1 Two strategies for dealing with outlying observations

2.1.1 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 in the constant or the deterministic trend, and the new distribution 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 a maximum of 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. Their starting point is the following unobserved components model for the N-dimensional vector X_t:

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

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 and $N-$dimensional unobserved vector, assumed to be at most I(1) and to follow a cointegrated VAR(p) process, whose VEqCM representation is: $Y_t = \alpha\beta' Y_{t-1} + \sum_{j=1}^{k-1} \Phi_j \Delta Y_{t-j} + \epsilon_t$; where α and β are $N \times r$ matrices, with $0 < r < N$; r is the number of cointegration relationships; and ϵ_t is a Gaussian white noise. Expressing (1) in terms of observable variables we get:

$$\begin{aligned} \Delta X_t &= \upsilon + \alpha(\beta' X_{t-1} - \vartheta(t-1) - \kappa DS_{t-1}) + \textstyle\sum_{j=1}^{k-1} \Phi_j \Delta X_{t-j} + \sum_{j=1}^{k-1} \Upsilon_j \Delta DS_{t-j} + \epsilon_t \\ &= \upsilon + \Pi^* X^*_{t-1} + \textstyle\sum_{j=1}^{k-1} \Phi_j \Delta X_{t-j} + \sum_{j=1}^{k-1} \Upsilon_j \Delta DS_{t-j} + \epsilon_t, \quad t = k+1, k+2, ..., \end{aligned} \tag{2}$$

where $\upsilon = -\alpha\beta'\mu_0 + \Psi\mu_1$ (with $\Psi = I_n - \Phi_1 - ... - \Phi_{k-1}$), $\vartheta = \beta'\theta$, $\Phi^* = \alpha[\beta' : \vartheta : \kappa]$, $X^{*'}_{t-1} = [X'_{t-1}, -(t-1), -DS_{t-1}]$, and Ω is the covariance matrix of ϵ. Finally:

$$\Upsilon_j = \begin{cases} \theta, \ j = 0 \\ -\Phi_j\theta, \ \ j = 1, ..., k-1 \end{cases} \tag{3}$$

Saikkonen and Lütkepohl (2000) propose to obtain initial estimators $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\Phi}_j$ and $\tilde{\Omega}$ from (2), without considering the restrictions in (3) and using the rank r_0 that is specified in the null hypothesis of the cointegration rank test. Then, defining $\tilde{A}(L) = I_n\Delta - \tilde{\alpha}\tilde{\beta}' L - \tilde{\Phi}_1 \Delta L - ... - \tilde{\Phi}_{k-1}\Delta L^{k-1}$, and $\tilde{Q}$ such that $\tilde{Q}\tilde{Q}' = \tilde{\Omega}^{-1}$, feasible GLS estimators of the deterministic components' coefficients in the original model (parameters μ_0, μ_1 and θ of expression (1)) are obtained pre-multiplying (1) by $\tilde{Q}'\tilde{A}(L)$:

$$\tilde{Q}'\tilde{A}(L)X_t = \tilde{H}_{0t}\mu_0 + \tilde{H}_{1t}\mu_1 + \tilde{K}_t\theta + \eta_t, \tag{4}$$

where $\tilde{H}_{it} = \tilde{Q}\tilde{A}(L)a_{it}$ $(i = 0, 1)$; $a_{0t} = 1$ for $t \geq 1$ and zero otherwise; and $a_{1t} = t$ for $t \geq 1$ zero otherwise. Deterministic components' coefficients can be estimated (under the null of $r = r_0$) by multivariate LS applied to (4). Once this is done the unobserved (and uncontaminated) vector Y_t can be estimated from (1).

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. 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.

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 using a VAR in the levels.

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.

2.1.2 Impulse Indicator Saturation

Recent developments on automatic model selection procedures applied to fully saturated regressions with impulse indicators seem to provide a general robust method to identify outlying observations. This method does not suffer from the drawbacks of traditional sequential searches. 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 φ are recorded. Next, the first $T/2$ indicators are dropped and those for the remaining observations are included. Finally, the significant indicators in each step are included altogether and those which are non-significant, dropped.

Johansen and Nielsen (2009) show that the efficiency loss due to testing the significance of T indicators is almost nonexistent for low nominal sizes of the tests (e.g., $\varphi \leq 1/T$). Indeed, in the case of no outliers and with $\varphi = 1/T$, 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 automatic model selection algorithm *Autometrics* which, starting from an initial General Unrestricted Model (GUM), reduces it up to a valid simpler expression (see Doornik (2009)). Castle et al.'s (2012) results show that non-relevant efficiency loses are caused by the inclusion of T impulse indicators when they are irrelevant. Additionally, the authors also study the performance of *Autometrics* with IIS under the presence of outliers and find a good performance in jointly selecting variables and detecting breaks.

2.2 Our empirical strategy for dealing with outlying observations in the pairwise approach

As aforementioned, the problem we are dealing with requires cointegration tests that consider the possibility of multiple outliers and breaks. In this section we propose an empirical strategy and check its properties by Monte Carlo.

Our proposal is to jointly select the dynamic structure and the interventions applying *Autometrics* with IIS in single equation models for the first differences of all the components (the *DGUM* in Castle et al. (2012))[1]. Next, after the interventions are identified, three alternative procedures may be applied for the pairwise cointegration tests: i) Estimate bivariate VEqCM models including the interventions found for both variables, drop the insignificant and test for cointegration with Johansen's test. This procedure may require simulating critical values for each test. ii) Test for cointegration in a single-equation framework using the *PcGive* approach (see Kiviet and Phillips (1992)). The outliers search could be done on these equations. This does not require the simulation of new critical values, but, because of being a single equation procedure, it requires exogenous variables. iii) Use the estimated dates and apply the GLS procedure proposed by Lütkepohl et al. (2004) and described in §2.1.1.

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

2.3 Simulation results for the Pairwise strategy with outlier correction

To analyze the performance of the strategy proposed in §2.2 we hash up the simulation exercise in Carlomagno and Espasa (2014) — *CE* — but applying the outlier correction strategy described above. In their *DGP 1*, the authors simulate a 100 dimensional VAR model (see expression (5) below), in which all the series are I(1) and a subset of n_1 series share a single stochastic trend. The notation n_1 will be used both, to denote the number of components in the *fully cointegrated*

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

subset and as its label.

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \Phi_1 \Delta Y_{t-1} + \epsilon_t, \tag{5}$$

where α and β are $N \times r$ matrices; $0 < r < N$; r is the number of cointegration relationships; and ϵ_t is a Gaussian white noise with covariance matrix equal to the identity. The specific forms of the matrices are:

$$\beta' = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & & & & & \\ -1 & 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}_{r \times N} \quad ; \quad \alpha = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\alpha_2 & 0 & 0 & \cdots & 0 \\ 0 & -\alpha_3 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & -\alpha_{n_1} \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{N \times r}$$

and the values α_i are taken from the uniform distribution with parameters $[0.15, 0.3]$ — these parameters were motivated by results in EM for CPI series. Matrix Φ is diagonal with its elements taken from the uniform distribution with parameters $[0.5, 0.8]$. See *CE* for a discussion about the properties of this DGP.

CE consider different four possibilities for n_1: in scenario 1 $n_1 = 10$, in scenario 2, $n_1 = 15$, in scenario 3, $n_1 = 25$, and in scenario 4, $n_1 = 40$. The authors also consider three different sample sizes: $T = 100$, $T = 200$ and $T = 400$. As will become clear later, the experiments are too complex to consider many different alternatives, thus, we do not reproduce the experiments for all scenarios and sample sizes. We focus just on their scenarios 1 and 3 and $T = 200$. In this section we maintain the name of the scenarios (1 and 3) just to keep track with *CE*. Furthermore, with the aim of not adding more complexity to the experiments, we apply IIS without Autometrics — what Castle et al. (2011) call the 1-cut approach — and check its properties by Monte Carlo.

2.3.1 Design of the experiments

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

follows: i) Apply 1-cut IIS with $\varphi = 1/T$ to the first differences of the 164 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 $[87\%, 65\%, 42\%]$, respectively. ii) Take the simulated series $(y_{i,t})$ of *CE* and contaminate them, according to previous percentages, as:

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

with Γ and LS_{t^*} being $(1 \times q)$ and $(q \times 1)$ vectors respectively; and q taking the values 1, 2 or 3. The vectors in LS_{t^*} take the value zero up to the date of the break and one from there on. The dates of the breaks (t^*) are randomly set in the interval $[20, 180]$, and this random position is maintained for all the 1000 replications. iii) The percentage of contaminated series is maintained for series inside and outside n_1.

With the contaminated series we replicate the simulations of *CE* in four different ways: a) *True Dates*: we use the true positions of the outliers to apply the pairwise strategy with outlier treatment. b) *Estim Dates*: we estimate the positions of the outliers using *1-cut* IIS in single equation models with three lags and then, we apply the pairwise procedure with outliers treatment. c) *No Outl:* we apply the pairwise procedure with outlier treatment (using estimated dates in b) to the uncontaminated series. d) *No Corr*: we apply the original pairwise procedure (with no outlier treatment) to the contaminated series.

The number of pairs among the 100 series in the DGP is 4950, and for each of the 1000 replications, we need to apply the GLS procedure described in §2.1.1 to all the pairs. Considering the 1000 replications and the three cases a) to c), the number of times that we need to apply the GLS procedure is $1000 \times 4950 \times 3 = 14.85$ *million*. To this figure, we need to add case d), in which the GLS procedure is not applied but still we need to perform $1000 \times 4950 = 4.95$ *million* cointegration tests. This complexity of the simulation experiments led us not to consider other possible DGPs or sample sizes.

2.3.2 Results

We asses the performance of the procedure using what Castle et al. (2011) call *potency* and *gauge* in the context of model selection. 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. In our context *gauge* measures the inclusion of wrong components in the estimated *fully cointegrated subset* and *potency* the

inclusion of correct components.

Table 1 includes the *gauge* and *potency* of the pairwise procedure with outlier treatment. Five main conclusions emerge from the table: i) Using the outlier correction treatment when it is not required (*Number of breaks: 0* block of the table), somewhat deteriorates the procedure's performance with respect to not correcting for outliers (the case of *CE*). 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 in *CE*). ii) 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 in *CE*). iii) Though *gauge* remains at very low levels, *potency* deteriorates with the number of outliers, even when using the *True Dates*. iv) When using the *Estim Dates* — instead of the *True Dates* — *potency* reductions are observed. But recall that, since we are using the *1-cut* approach, we are not selecting the lag length for the series in n_1 — which are the most relevant in determining the pairwise procedure's potency — so that potencies under *Estim Dates* columns 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. v) 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* slightly 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, for which the proportion of series in n_1 is larger.

To conclude, in this section we proposed an outlier treatment for the pairwise strategy and studied its properties by Monte Carlo. Though the proposal requires the identification of the outliers' dates by *Autometrics* with IIS, the use of *Autometrics* in our simulation setting can be very tedious. As an alternative we applied IIS with the simple 1-cut approach. The main disadvantage of this procedure is that it does not jointly select the dynamic structure and the impulses. In series with richer dynamics (those in n_1) this problem seems to be more important, as our results confirm. For this reason, we interpret potencies in Table 1 as lower bounds for the ones that will be obtanied with *Autometrics*. As a general conclusion from Table 1 we can say that the presence of outliers distorts the potency of the pairwise procedure and this distortion is larger, the larger the number of outliers. Some procedure is required for diminishing this

Table 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

- *Pot* = $\frac{100}{n_1 Nexp} \sum_{i=1}^{Nexp} Z_{1,i}$, with $Z_{1,i}$ being the number of correct series included in $\hat{n}_1$ (the estimated fully cointegrated subset) in experiment i.
- *Gauge* = $\frac{100}{(N-n_1)Nexp} \sum_{i=1}^{Nexp} Z_{2,i}$, with $Z_{2,i}$ being the number of incorrect series included in $\hat{n}_1$ in experiment i.
- *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.
- *No Corr*: the original pairwise procedure (with no outlier treatment) is applied to the contaminated series.
- Figures under *Estim Dates* columns 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.

problem when series are contaminated. Our results show that the proposal sketched in §2.2 can be regarded as an acceptable approach.

3 Small samples correction

As noted by *CE*, the potency of the pairwise procedure to discover the true *fully cointegrated subset(s)* deteriorates as T decreases. There are two reasons that explain this issue: the Johansen's trace test properties deteriorates in small samples, and the the 'almost equivalence' of the tests showed by *CE* — and summarized in the introduction of this paper — is valid only asymptotically, and also deteriorates in small samples.

To mitigate this problem, we propose a slight modification of the procedure: relax the '*full cointegration*' requirement to '*almost full cointegration*'. The relaxation consists of 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 φ nominal size, but it is found if test are

performed at φ^* ($\varphi^* > \varphi$) — a detailed description of the relaxation procedure is included in §5. This strategy will lead to increase the potency of the procedure but will also increase the risk of including wrong series.

Assume that we have three $I(1)$ 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. As we — wrongly — found the stochastic trend of S_3 to be that of S_1, which is also the same as that of S_2, 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 φ (the unconditional asymptotic probability). Call this conditional probability $\tilde{\varphi}$.

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 $\tilde{\varphi}$ is, the larger Z_2^λ would be. Similarly, for a given $\tilde{\varphi}$, the smaller $\hat{n}_1$ is, the larger Z_2^λ would be, hence, Z_2^λ is a decreasing function of the original $\hat{n}_1$.

Table 2 illustrates these arguments. It shows the mean number of potential candidates for each of the four scenarios considered by *CE* in their *DGP 1*, and relaxation parameter up to $\lambda = 3$ (see §2 for a brief description of the DGP, or *CE* for a more detailed discussion). We consider only the sample size $T = 100$ because for larger samples *CE* did not found relevant potency distortions. While (a) column of the table contains the mean number of series that have between 1 and λ holes in the current *(almost) fully cointegrated set*, (b) column includes the series of column (a) whose holes were filled after relaxing cointegration tests from 1% to 5% of significance. Columns (c) and (d) are analog to (a) and (b) but wrong candidates are excluded.

As the *Ratios* columns 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. The 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 ones. For $\lambda = 3$, while in scenario 1 only 40% of the candidates are correct series, in scenario 4, 97% of them are so.

Hence, results in Table 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 it is precisely for those scenarios that improving the original results is most needed. Recall

Table 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, with n_1 being the number of series that share a single common trend among the 100 series in the model.
- 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.

that the 'almost equivalence' property of cointegration tests studied in theorem 1 of *CE* is valid only asymptotically. Then, as T decreases, the asymptotic equivalence deteriorates and the probability of finding cointegration between all the correct pairs moves away from $(1 - \varphi)$ 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 Table 3 of *CE* 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.

Table 3 adds more evidence for the two arguments made above, namely, while the risk of relaxing the full cointegration requirement is decreasing in n_1, the potential benefit is increasing. Define Z_1 as the number of correct series included in $\hat{n}_1$, and Z_2 as the number of wrong series included in that subset. The table 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 (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 somewhat higher.

Table 3: Mean of the ratios Z_1/n_1 and Z_2/n_1 as a function of the relaxation parameter λ

	Mean Z_2/n_1				**Mean Z_1/n_1**			
	Sce1	**Sce2**	**Sce3**	**Sce4**	**Sce1**	**Sce2**	**Sce3**	**Sce4**
$\lambda = 0$	0.04	0.02	0.01	0.00	0.77	0.73	0.67	0.62
$\lambda = 1$	0.06	0.02	0.01	0.00	0.83	0.79	0.73	0.68
$\lambda = 2$	0.08	0.03	0.01	0.00	0.87	0.84	0.78	0.73
$\lambda = 3$	—	0.04	0.01	0.01	—	0.86	0.81	0.76
$\lambda = 4$	—	0.05	0.02	0.01	—	0.87	0.83	0.78
$\lambda = 5$	—	0.05	0.02	0.01	—	0.87	0.84	0.80
$\lambda = 6$	—	—	0.02	0.01	—	—	0.85	0.81

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

Therefore, 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).

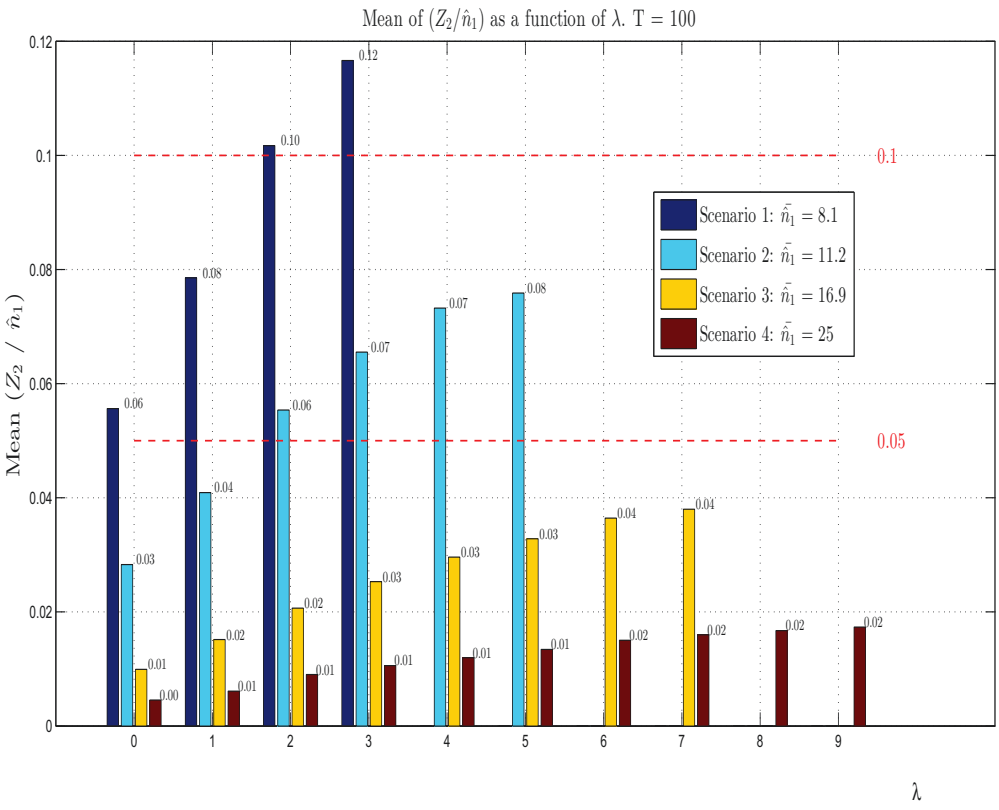
Using the same simulated series as *CE* we run the pairwise procedure but instead of requiring *full cointegration* we consider the relaxation to *almost full cointegration* using alternative relaxation parameters λ. The alternative values of λ considered were; $[1, 2, ..., 9]$.

To decide on the 'optimal' λ we consider the following criterion. 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[2]. Next, we average this ratio over all the 1000 experiments. Finally, a decision rule to choose the optimal λ has to be defined.

[2]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.

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 1 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 criterion 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 also Table 3). 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.



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 1: 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.

4 Generalization of the cointegration tests

This section has two objectives; propose some extensions to the pairwise cointegration tests that could be useful for empirical applications, and define some criteria to evaluate the 'quality' cointegration relationships found when working with real data.

4.1 Extensions of the pairwise tests

The original procedure is extended in three directions motivated by results in Aron and Muellbauer (2013). First, we consider the inclusion of a weak exogenous variable in all the bivariate models, second, we allow for a *parsimonious long lag* parametrization (*PLL*), and finally, we consider the possibility of a specific form of non-linearity.

Weakly exogenous variable

A third variable V_t in the originally bivariate VARs may help to find cointegration between two components of a macro variable. In models where V_t is significant in the long-run we assess its weak exogeneity. If weak exogeneity is rejected we do not consider it in the model[3].

Note that now we may have zero, one or two common trends between two components and V_t. When there is only one, we are in the regular case and blocks of prices can be constructed as summarized in section 5. Since we are considering only the cases where V_t is weakly exogenous, the common trend will be generated by the accumulation of the shocks to this variable.

For the case of two common trends, it can be shown that blocks can still be constructed and components inside them will share the two trends; one generated by the shocks of V_t and the other by a combination of the shocks in the two components (a proof of this result is available upon request).

Therefore, when including a third (weakly exogenous) variable the strategy for constructing the blocks of components is slightly changed. We first consider the models that have only one common trend and then those with two common trends.

Finally, note that the inclusion of V_t can be implemented in two alternative fashions. The simplest one is just adding V_t to all the pairs. Note however that this strategy could distort the results for the pairs that do not need this third variable. Hence, the second alternative is to proceed sequentially in two steps: first the block search is carried out not including V_t and then, using only the series not belonging to any (almost) fully cointegrated block, the procedure is repeated but including V_t.

[3] V_t is also considered for the outlier correction procedure.

Parsimonious Longer Lags (PLL)

As argued by Aron and Muellbauer (2013), a possible way of tackling the 'curse of dimensionality' present in VAR models with long lags is to impose 'parsimonious longer lag' (PLL) restrictions. We consider use a similar PLL form as Aron and Muellbauer (2013): full generality is permitted at short lag lengths (1 and 2) and for lag 12; lags 3 to 5 are replaced by $\Delta_2 X_{t-3}$; lags 6 to 11 are replaced by $\Delta_5 X_{t-6}$; and finally, lags 13 to 25 are replaced by $\Delta_{12} X_{t-13}$. Compared to unrestricted lags up to 25 months, 25 parameters are replaced by six.

This extension raises the problem highlighted by Nielsen and Nielsen (2008) about the appearance of large, albeit not significant, roots in the characteristic polynomial of the VAR. This issue is particularly relevant for our procedure since, as we propose in §4.2, we will disregard models with large second (third) root. We tackle this issue by computing the roots in the polynomial that remains after disregarding the PLL structure. This solution assumes that the polynomial associated with the PLL does not have unit roots.

Non-linearities

As argued by Aron and Muellbauer (2013), when dealing with prices, the possibility that producers adjust prices more frequently when cost changes are more volatile (see also Reis (2006)) may induce non-linearities in inflation; high recent inflation would induce disproportionately high future inflation. A simple way of considering this possibility is the one suggested by Aron and Muellbauer (2013); regress $(\Delta_6 log(P))^2$ on a constant and $\Delta_6 log(P)$, where P is a vector of two disaggregated prices. The residual of this regression and its 6-month lag may capture the non-linearities described above (see Aron and Muellbauer (2013) for further details) and can be included as a purely exogenous variable in the bivariate VARs.

4.2 Quality assessment of the cointegration tests' results

In order to obtain economically and statistically sensible cointegration restrictions we propose to consider only those that satisfy the following conditions: i) the 'long-run' relationship does not require a deterministic trend; ii) coefficients of both prices are statistically significant; iii) the second largest root of the bivariate VAR's characteristic polynomial is not close to one; and iv) the 'long-run' relationship is stable over time.

The first condition is important when dealing with macroeconomic variables as 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 (as it is the case of the application in

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

This strategy is not exactly the one suggested by Nielsen and Rahbek (2000). These authors find that cointegration rank tests are asymptotically similar with respect to the parameters of deterministic components. Thus, they proceed in two steps. First, they test the cointegration rank in a model that includes all the deterministic components (constant, trends and interventions) in the cointegrating relationships and its differences in the VAR. Second, once the cointegration rank is determined, hypothesis on deterministic parameters can be tested (see also Doornik et al. (1998) and Juselius (2006)). The reason for not not exactly following this procedure is that we are interested in testing cointegration only if the model does not require a trend in the cointegrating space.

With the second condition we want to exclude stationary variables from the *fully cointegrated sets*. Coefficients tests can be performed as in Johansen (1995) at, say, 10% of significance not to exclude too many pairs.

The third condition is relevant for the procedure does not exclude a priori series with an $I(2)$ behavior. Our proposal is to disregard models whose second root is larger than, say, 0.90.

For condition iv (stability of the 'long-run' relationships), we propose to consider both, the first eigenvalue and long-run coefficients' stability. Pairs that do not pass both tests should be disregarded. The reason for considering both tests is that constant eigenvalues do not imply constant long-run coefficients, and constancy of those coefficients does not imply that Johansen's test results are stable over time (see Hansen and Johansen (1999) and Juselius (2006)). Eigenvalues are assessed by the *fluctuation test* proposed by Hansen and Johansen (1999), and long-run coefficients in the evaluation period are compared with the full sample estimation (see Hansen and Johansen (1999) and Juselius (2006)). For the empirical application in §6, in both cases we consider forward recursive tests at 5% of significance based on the concentrated model and the evaluation period is the last five years of the sample (see Juselius (2006) for a discussion on the

pros and cons of using the concentrated model *versus* the full model).

5 Detailed algorithm of the Pairwise procedure

Since the original procedure has been modified in several ways to make it more useful for empirical applications, we devote this section for describing the final algorithm. Before doing so, a comment about seasonal unit roots is worthwhile.

The procedure studied in this paper does not deal with the possibility of seasonal unit roots. A proper treatment of this issue will highly increase our procedures' complexity, specially when testing for cointegration between series with different number of seasonal unit roots. Seasonally adjusted series may artificially generate common dynamics in the series and therefore distort estimations. Nonetheless, in order to avoid complex estimation problems, those kind of series are sometimes used by econometricians, Hendry and Hubrich (2011), Stock and Watson (2007) and Trenkler et al. (2007) are some relevant examples. Other — probably better — alternative could be to seasonally adjust the series using past filters only, but this is also complex, as statistical offices do not provide these series and the standard software do not allow to do that. For these reasons our approach is to use seasonal dummies.

The procedure involves ten steps:

i. Perform Johansen cointegration tests between all possible pairs of components, disregarding those which do not pass the four quality conditions in §4.2, and store the resulting p-values.

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

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

iv. As analyzed in §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 other elements of the *fully cointegrated subset* to enter the subset. Call this new set *almost fully cointegrated*. If the user does not want to consider this relaxation, in point iii, instead of finding just the the largest clique, all independent cliques should be found, and the procedure ends there. Otherwise, continue.

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

   (a) cointegration — at the original φ of confidence — is rejected with at most λ of the series already in the subset $\hat{n}_1$.

   (b) when the nominal size of the cointegration test is relaxed to φ^* the candidate is cointegrated with all the series already in the subset $\hat{n}_1$.

vi. Construct the set of candidates C_0. If all the candidates are pairwise cointegrated between each other (at the original φ), let all of them in and go to point ix (because there are not more potential candidates).

vii. If not, find the maximal clique (see point iii) 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.

viii. If there are not cointegrated candidates, try to include them 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.

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

x. Once the disaggregation map is obtained, the forecasting equations can be constructed.

Figure 2 summarizes the algorithm in five basic steps.

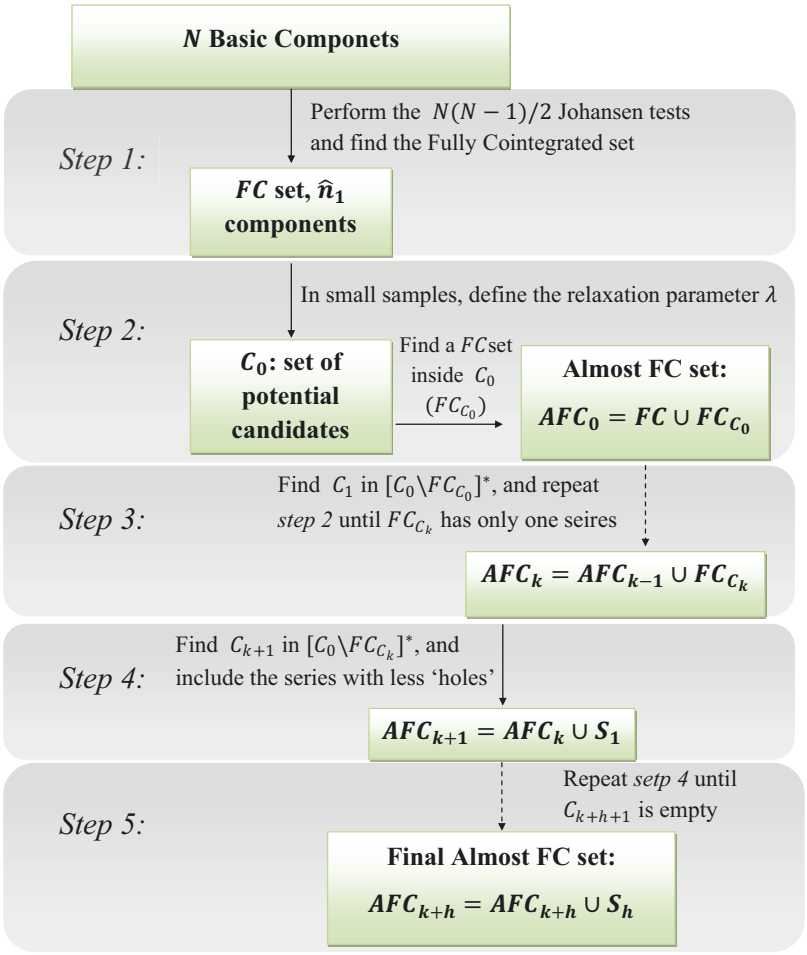


Figure 2: Sketch diagram for the algorithm to construct the set of components with a common trend

Notes:
- * The symbol '\' represents the set difference operator, so that $A \setminus B = \{x \in A : x \notin B\}$.
- In step 1 disregard pairs which do not pass the quality tests described in §4.2.

6 Empirical application: US CPI

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

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

6.1 Data

The CPI break down used in this analysis correspond to the maximum disaggregation level available to the public in the *Bureau of Labor Statistics* (seasonally un-adjusted CPI-U for all urban consumers) for the period $1999.1 - 2014.12$ (192 observations). The total number of components is 181. After dropping those with less than 162 valid observations we keep 172 basic components. From these series we exclude eight that evolve by steps (regulated prices) so that we end up with 164 series which, considering the weights of 2014, represent 91% of the CPI [5].

6.2 Outliers' analysis

As described in §2.2 the outliers search for all the 164 components is carried out in individual models for the differenced components using *Autometrics* with Impulse Indicator Saturation (IIS).

To make tables legible, basic components are grouped into six broad categories: non-energy industrial durable goods (MAN Dur), processed food (PF), services (SERV), non-processed food (NPF), non-energy industrial non-durable goods (MAN No Dur) and energy (ENE)[6].

Table 4 summarizes the results. Four main observations emerge from it: i) the average number of outliers per series is 4.7 (2.6% of the observations); ii) energy and services prices are the most contaminated with a mean proportion of 4.5% and 3.6% of outlying observations per component, respectively; iii) 37% of the outliers are large (larger than 4σ in absolute value); and iv) large outliers are more typical in services and energy prices, representing 52% and 45% of the total number of outliers, respectively.

Another point of interest regarding the outliers' analysis is its distribution by dates. Figure 3 shows the number of series with at least one outlier at each of the 192 months of the sample. As it shows, the distribution if far from uniform, with some months having 14 (8.5%) series with outliers and some others with none. Interestingly, there seems to be a concentration around years 2008-2009, the sub-prime crisis period (red box of the figure).

To confirm that there is a mean shift in the number of contaminated components during the sub-prime crisis, we estimate a model for the proportion of series with outliers including as potential regressors lags one to five, seasonal dummies and choosing the *Autometrics* option

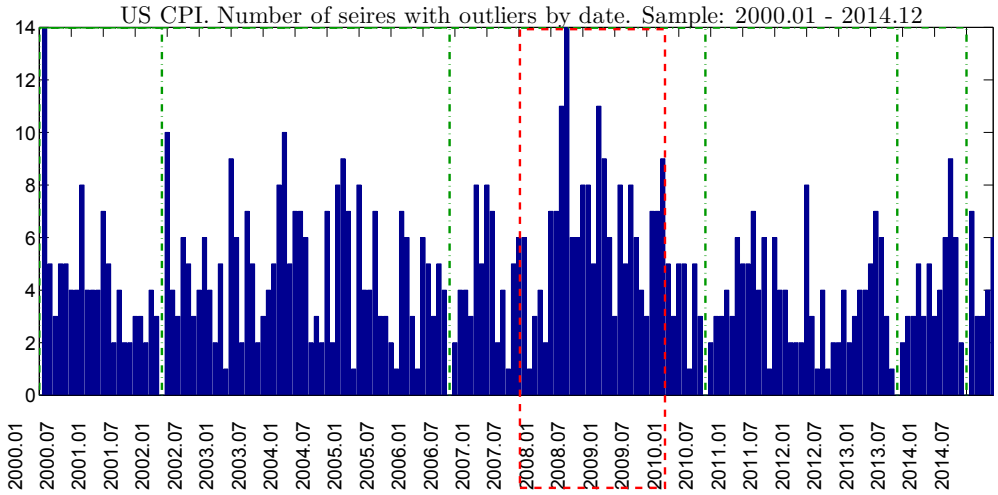
[5]The eight excluded series are: College tuition and fees, Elementary and high school tuition and fees, Child care and nursery school, Postage, Delivery services, Wireless telephone services, Food at employee sites and schools, and Housing at school excluding board.

[6]Note that this grouping is not perfect for a basic component could include prices belonging to two broad categories

Table 4: Mean number of outliers by size and category

	L+	S+	S-	L-	**Mean**	**Mean (% of T)**
NPF (25)	1.3	2.0	1.4	0.4	**5.2**	**2.9%**
ENE (6)	2.3	3.2	1.3	1.3	**8.2**	**4.5%**
PF (38)	0.6	1.9	0.9	0.2	**3.7**	**2.0%**
MAN_dur (51)	0.5	1.6	1.3	0.6	**3.9**	**2.2%**
MAN_NoDur (10)	0.3	1.5	1.1	0.6	**3.5**	**1.9%**
Serv (34)	2.3	2.3	0.8	1.1	**6.5**	**3.6%**
TOTAL(164)	7.4	12.4	6.9	4.2	**4.7**	**2.6%**
PROP.	**24%**	**40%**	**23%**	**13%**	**100%**	

Numbers in parenthesis after the category name indicate the number of series in the category.
L+: Large (larger than 4σ) and positive outliers.
S+: Small (smaller than or equal to 4σ) and positive outliers.
L-: Large and negative outliers.
S-: Small and negative outliers.



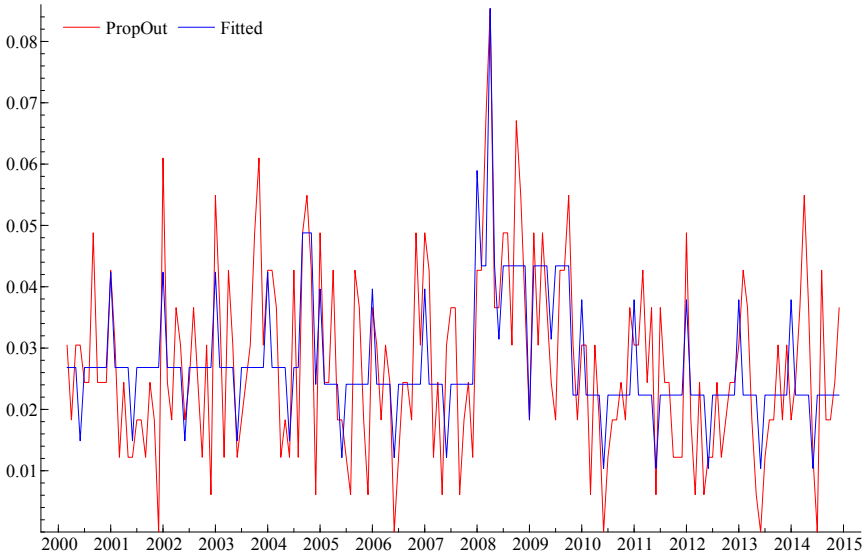
Note: Green doted lines indicate dates at which there are no contaminated series.

Figure 3: Number of series with at least one outlier

IIS+SIS[7] for impulses and steps detection with a target gauge equal to $1/T$. Results are summarized in Figure 4, from where five important conclusions can be drawn: i) There is seasonality in proportion of contaminated series. While in January there are, on average, more series with outliers, in June the proportion of series with outliers is reduced. ii) In 2004.11 there is a small reduction in the proportion of series with outliers with respect to previous years, which last until 2007.11. iii) In 2007.12 there is a significant and positive step up to 2009.10. iv) After 2009.10 the mean proportion of series with outliers is somewhat lower than before the crisis.

Finally it is noteworthy that the exhaustive outliers' search we made is not only relevant for

[7]SIS stands for *Step Indicator Saturation*. The option IIS+SIS saturates the regression not only with impulses but also with steps, see Doornik et al. (2013).



The estimation sample is: 2000(3) - 2014(12)					
	Coefficient	Std.Error	t-value	t-prob	Part.R^2
Constant	0.0226	0.0015	15.2000	0.0000	0.5773
CSeasonal	0.0155	0.0034	4.5700	0.0000	0.1101
CSeasonal_5	-0.0120	0.0032	-3.7600	0.0002	0.0771
I:2008(4)	0.0420	0.0120	3.4900	0.0006	0.0672
I:2009(1)	-0.0406	0.0124	-3.2700	0.0013	0.0594
S1:2004(8)	-0.0219	0.0070	-3.1500	0.0019	0.0555
S1:2004(11)	0.0247	0.0071	3.5000	0.0006	0.0676
S1:2007(12)	-0.0193	0.0033	-5.9300	0.0000	0.1721
S1:2009(10)	0.0211	0.0030	6.9800	0.0000	0.2239

Notes:

- Steps ($S1$:) take the value one from the first observation until the date indicated in the name of the step, and zero from then on.
- CSeasonal (Cseasonal_5) is the centered seasonal variable corresponding to January (June). It takes the value $1 - \frac{1}{12}$ in January (June) and $-\frac{1}{12}$ otherwise.

Figure 4: Changes in the mean proportion of series with at least one outlier

individual series, but also for modeling the CPI itself. This is so for outliers in the components are also outliers in the aggregate but, as we argue below, in many cases they can be estimated only in the components. In order to use the individual outliers in a model for the CPI, we construct the *aggregated outlier* series ($AggOut$) as the weighted sum of all individual outliers multiplied by their coefficients, and include this series in a model for the CPI. Since the individual outliers will enter the CPI weighted by the corresponding component's weight, we expect the coefficient of the *aggregated outlier* not to differ significantly from one.

For assessing the usefulness of this variable to model the CPI, we compare three simple models. Starting from the GUM; $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + \epsilon_t$, where S_{it} is a centered seasonal dummy, we consider three possibilities to be estimated with

Autometrics: (1) *Only IIS:* IIS is applied in previous GUM. (2) *Only AggOut:* The GUM is augmented with the series of $AggOut_t$ (IIS is not used). (3) *AggOut:* IIS is applied in the augmented GUM.

Table 5 includes model selection criteria for the three possibilities. As it shows, the two models including $AggOut_t$ outperform model (1). Interestingly, model (3) seems to be the best option. This last result suggests two conclusions: some components' outliers — which are also outliers of the CPI — are not identifiable in the model for the aggregate, and some CPI's outliers — which must be present in some component — are not identifiable in component's models, probably because these observations correspond to small outliers of the same sign in more than one component. Figure 5 includes the *aggregated outlier* series and the impulses retained in model (3). Blue peaks represent the outliers which are not identifiable in disaggregated series (2006.09, 2007.11 and 2008.06).

Table 5: Comparison of different models for the CPI

	AIC	**SIC**	$\boldsymbol{Adj.R^2}$
Only IIS	-9.14	-8.90	0.65
Only AggOutl	-9.22	-9.02	0.67
AggOutl + IIS	-9.33	-9.06	0.71

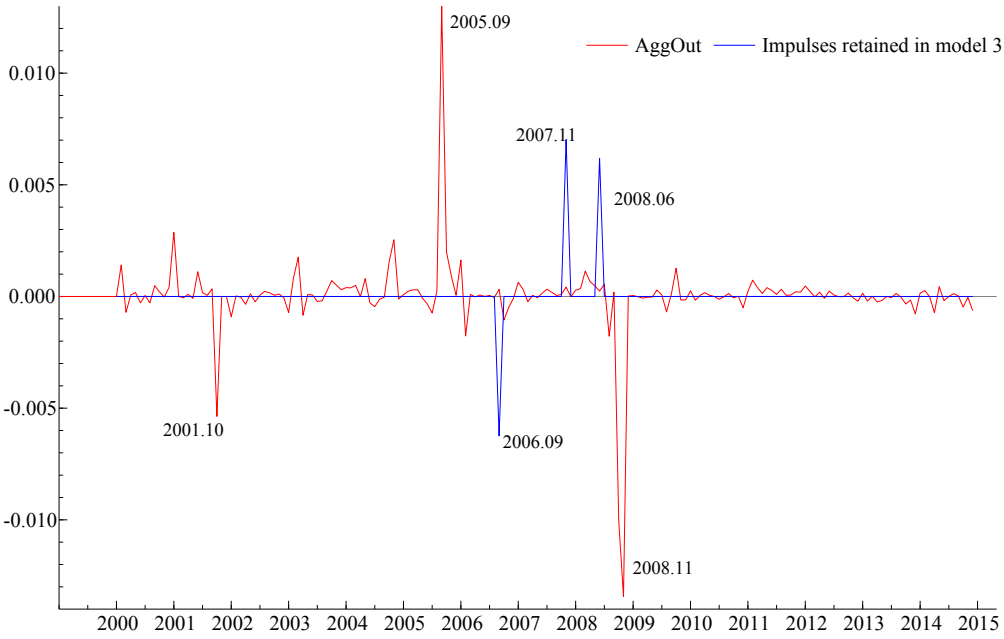
Basic GUM: $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + \epsilon_t$.
Only IIS: IIS is applied in previous GUM.
Only AggOutl: The GUM is augmented with the series of $AggOut_t$ (IIS is not used).
AggOutl: IIS is applied in the augmented GUM.

6.3 Pairwise tests' results

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

For the outlier corrected series (see §2), Johansen's tests are performed at the 5% of significance and the number of lags for each pair is determined with the AIC, in a model without trend in the 'long run' and one common unit root restriction. Centered seasonal dummies are included in all models.

Among the 164 components there are 13366 possible pairs. Without considering the last three quality conditions of §4.2, we find 2333 pairs with a common unit root (the first condition — no deterministic trend — is always applied). After disregarding those that do not satisfy the other



- *GUM*: $\Delta CPI_t = c + \sum_{i=1}^{4} \phi_i \Delta CPI_{t-i} + \phi_{12} \Delta CPI_{t-12} + \sum_{s=1}^{11} \rho_i S_{it} + AggOut_t + \epsilon_t$.
- IIS is applied in previous GUM and the resulting impulses are added up using their coefficients (blue line).

Figure 5: Aggregated and remaining outliers in ΔCPI

three conditions (significant coefficients, root not close to one and stability) we keep 853 pairs that are considered to satisfy stable and statistically sensible unit root restrictions.

For grouping the components by blocks we consider the strategy summarized in section 5 augmented by the *relaxation procedure* also described in that section. Blocks with less than four series are disregarded to avoid spurious grouping (see *CE*). In the *relaxation step*, a maximum of 2 *holes* is admitted[8] and the significance level augmented to 10%.

We also consider the three extensions mentioned in §4.1 (a weakly exogenous variable, parsimonious long lag, and a special form of non-linearity). As exogenous variable we use the Real Effective Exchange Rate (REER, constructed by the Bank for International Settlements[9], using consumer price indices for trading partners to deflate nominal effective exchange rate). The inclusion of non-linearities was the only one that leaded to lose common unit root restrictions so we finally did not use it.

Table 6 summarizes the results with and without the extensions of §4.1. As it shows, the extensions lead to discover 'long-run' relationships that were not found in the baseline case. Only by admitting Parsimonious Long Lag structures, the proportion of series included in some

[8] If the original set has less than 6 series only one hole is admitted.
[9] https://www.bis.org/statistics/eer/index.htm

block increases from 18 to almost 34. Regarding the *REER*, Table 6 suggests that including it when it is not needed may distort the results (compare the number of series in the second and third line of the table). This finding is in line with the simulations results in *CE* and those of Lütkepohl et al. (2003) and Johansen (1995) about that cointegration test's power decreases with the number of stochastic trends in the system. For this reason, we prefer the sequential procedure in which *REER* is included only in a second step after having constructed the blocks without considering this variable.

Table 6: Summary Pairwise procedure results: total proportion of series and weight of all the (almost) fully cointegrated subsets with outlier corrected series

Total proportion of series and weight of all the *(almost) fully cointegrated subsets*

	Num of Sets	Num of Sers	Weight(%)*	Weight top three(%)**
Without any extension	4	18	4.0	1.3
With PLL	8	34	7.8	1.8
With PLL and REER	7	30	32.4	27.6
With PLL and REER (sequential)[a]	10	42	11.2	3.4

(*) Represents the proportion of weight in the 164 series considered, which weight 89% of the CPI.
(**) Weight of the tree series with larger weights.
(a) In this strategy blocks are firstly formed without *REER* and then the procedure with *REER* is executed only for the series not included in any previous block.

Note, however, that the direct strategy for including the *REER* captures a significant larger weight of the CPI than the sequential procedure (compare the last two rows of the third column in Table 6 for). As the last column of the table shows, this difference is explained by the inclusion of few 'heavy' components in the direct procedure. In fact, the inclusion of *'Owners' equivalent rent of primary residence'* (which represent 25% of the weight of the series we are dealing with) explains all the difference. We therefore still prefer the sequential procedure.

The last remark regards the outliers treatment. Results commented up to now were for the outlier corrected series. To assess the impact of this correction, Table 7 replicates Table 6 but without outlier correction. A noteworthy difference between the two tables is that when the series are not corrected for outliers the number of them included in some block is systematically larger. This result may be due to two issues: co-breaking relationships may be 'confused' with common unit roots restrictions in not corrected series, and power problems of the *GLS* procedure may lead to incorrectly find no common unit roots restrictions in too many pairs.

Since the outlier correction strategy does not distort the *gauge* of the pairwise method (see simulation results of Table 1), the 42 series of Table 6 may be considered as lower bound of

the number of series that would be found without outlier treatment if the components were not contaminated. Hence, the number of series in Table 7 can be *correctly* larger only if power issues of the GLS method were the main source of the differences between the two tables. In this case, one should expect that series in Table 7 include a large proportion of those in Table 6 plus some other series. However, when comparing the individual blocks obtained with both procedures, coincidences are minor. For this reason, and because on a theoretical basis the appropriate approach is that with outlier correction, we select the results of Table 1 for our application.

Table 7: Summary Pairwise procedure results: total proportion of series and weight of all the (almost) fully cointegrated subsets with series *not* corrected for outliers

Total proportion of series and weight of all the *(almost) fully cointegrated subsets*

	Num of Sets	Num of Sers	Weight(%)*	Weight top three(%)**
Without any extension	6	34	22.1	12.2
With PLL	9	43	15.0	6.6
With PLL and REER	9	39	11.8	4.2
With PLL and REER (sequential)[a]	11	51	18.5	6.6

(*) Represents the proportion of weight in the 164 series considered, which weight 89% of the CPI.
(**) Weight of the tree series with larger weights.
(a) In this strategy blocks are firstly formed without $REER$ and then the procedure with $REER$ is executed only for the series not included in any previous block.

6.4 Some detailed results

Table 8 gives some details of the outcome for the procedure with PLL, *sequential* $REER$ and outliers correction. To make the table legible, we use the same six broad categories as in Table 4.

The main conclusion of the table is that blocks of series sharing one common unit root (or two for block 10 and 11 that have $REER$) cannot be assigned to a single broad category. However, in almost all the cases, more than 85% of the *Set*'s weight is explained by two broad categories. The exceptions are blocks 6 and 7 for which the two most important categories explain 70% of the block's weights.

This observation has two relevant implications: first, the ad-hoc method proposed by Boivin and Ng (2006) for extracting non-pervasive common factors — based on pre-grouping the series in broad categories — would not work for the US CPI; second, although a 'labeling' strategy that matches blocks with single broad categories is not possible, this could be done using just two categories.

Table 8: Detailed results of the Pairwise procedure with common unit roots: number of series and proportion of weight by broad categories and blocks

	MAN dur		PF		SERV		NPF		MAN No dur		ENE		Tot W
	Q	W	Q	W	Q	W	Q	W	Q	W	Q	W	
TOTAL	51	17.2	38	12.6	34	51.0	25	4.8	10	3.8	6	10.7	
Block 1	0	0.0	3	62.2	0	0.0	2	37.8	0	0.0	0	0.0	1.6
Block 2	1	11.8	2	43.5	0	0.0	2	44.7	0	0.0	0	0.0	0.7
Block 3	2	56.7	1	7.5	0	0.0	1	35.7	0	0.0	0	0.0	1.1
Block 4	1	7.0	0	0.0	0	0.0	2	60.1	0	0.0	1	32.9	0.8
Block 5	0	0.0	1	10.9	0	0.0	2	30.1	1	59.0	0	0.0	0.6
Block 6	1	16.0	1	50.9	1	21.4	1	11.6	0	0.0	0	0.0	0.8
Block 7	1	29.3	0	0.0	2	41.9	0	0.0	1	28.7	0	0.0	1.0
Block 8	2	14.1	0	0.0	1	52.7	0	0.0	1	33.2	0	0.0	1.2
Block 9	2	92.0	1	2.9	0	0.0	1	5.1	0	0.0	0	0.0	2.4
Block 10	0	0.0	2	48.0	0	0.0	1	14.4	1	37.6	0	0.0	1.0
TOTAL	10		11		4		12		4		1		11.2

Columns Q indicate the amount of series in each category and Set.
Columns W indicate the the total weight of each category in the CPI and the proportion of the weight of each category in each Set.
Last column contains the total weight of the blocks.

6.5 Forecasting the US CPI and all its components

After identifying the subsets of basic components in which the elements share just one common trend, we build single-equation models for all the basic components. When a component belongs to one of the fully cointegrated subsets, its econometric model include the corresponding cointegration relationships, so that, we are considering relevant restrictions in the long-run dynamics of the components. In each case, the process of building the econometric model is subject to a set of diagnostic tests included *Autometrics*, so we can conclude that they are reasonable for empirical applications. Additionally, since the basic components aggregate to the CPI, we can apply another test to the models for the disaggregates. It consist of comparing the forecast of the aggregate obtained by aggregating the forecasts of the components — a type of indirect forecast — with the forecasts from a scalar model for the aggregate — direct forecast. We denote this indirect approach by *I-PW* (*PW* stands for *pairwise*) and the direct one by *D*. The latter is our baseline model. We also compare the forecasting performance of *I-PW* with the indirect forecast for the aggregate using univariate models for each basic component, denoted as I-B, which is used as a baseline indirect procedure.

Hopefully, the pairwise strategy (*I-PW*) would not only provide models to analyze all the components, but it would also be an instrument to obtain more accurate forecasts of the aggre-

gate. This could be so because it incorporates more information than the corresponding direct forecast, and could palliate the curse of dimensionality in the number of variables by considering restrictions between them. Therefore, our approach to forecast the aggregate is an intermediate one between the direct approach and the vector-model approach (a full information method, that in our case of interest is not feasible).

With reference to the indirect disaggregated procedures, we initially consider four broad possibilities depending on the regressors to be included in the formulation of the initial General Unrestricted Model (GUM). Apart from own lags and seasonal dummies, we may include a) no other regressor, b) lags of the aggregated CPI, c) lags of the official subaggregates corresponding to a breakdown of the CPI in eight categories, or d) lags of all components.

For each of these four possibilities, series with no common unit roots can be modeled individually ,or all together with a scalar model for the sub-aggregate *rest*. Abusing notation we label this last possibility as *GP*, for Guerrero and Peña (2003).

In principle we have therefore eight different possibilities. Now, for each of these we could consider common unit roots restriction (*I-PW*) or not (*I-B*). The number of possibilities would now be 16, but note that when not considering common unit roots *GP* is not available, so we end up with end up with 13 different indirect possibilities. When not using common unit roots we add an additional possibility consisting of including dynamic factors estimated from all the disaggregates (I-DFM), what rises the number of options to 14.

Finally, as argued by *CE* different normalizations of the unit-root restrictions may lead to different forecasting accuracy (by changing the amount common unit roots restrictions relevant to each component), so we consider three alternative normalizations. In the first one, restrictions are expressed as deviations of all the variables with respect to the dependent variable in the corresponding equation (a different normalization is used in each equation). In the second normalization, we randomly select a variable and express all restrictions as deviations from the chosen variable. In the third one, restrictions are normalized with respect to the sub-aggregate formed by the series in the corresponding subset.

We end up with 20 indirect options that are summarized in Table 9 (only the case in which no lags of other components or sub-aggregates are included in the model -case *a* above- is considered in the table, the other three cases are simple extensions).

The equations in Table 9 represent the initial GUMs from where models are selected using *Autometrics* with Impulse Indicator Saturation. The selection is carried out in two steps. First

we use a target gauge of 0.5% to select variables, lags and impulses, and store the retained impulses. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.

Table 9: Summary of the indirect forecasting exercises

Model	Description
I-B	Baseline disaggregate. Individual univariate models for all the components $\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$
I -DFM-2	Indirect procedure with two DF $\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k_1} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} +$ $\sum_{k=1}^{K} \delta_{1,k} F_{1,t-k} + \sum_{k=1}^{K} \delta_{2,k} F_{2,t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$
	Pairwise
I - PW - Nj	Single-equation model for all the components including CT restrictions using normalization j **(i)** $\Delta x_{i,t} = c + \sum_{r=1}^{R} \alpha_{i,r} CR_{1_r,t-1} + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \phi_{12}\Delta x_{i,t-12} +$ $\phi_{24}\Delta x_{i,t-24} + \sum_{j=1}^{J} \theta_j \Delta SubAggCT_{i,t-j} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$ For series without common unit roots restrictions: **(ii)** $\Delta x_{i,t} = c + \sum_{k=1}^{K} \phi_k \Delta x_{i,t-k} + \phi_{12}\Delta x_{i,t-12} + \phi_{24}\Delta x_{i,t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_{i,t}$
I - PW - Nj - GP	Individual scalar model for all the components with CT (case i above) restrictions. For the other components, only its sub-aggregate is forecast

- All the equations represent the initial GUMs from where models are selected using *Autometrics* with Impulse indicator saturation. The selection is carried out in two steps. First we use a target size of 0.5% to select variables, lags and impulses. Retained impulses are stored. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.
- $K = J = 4$.
- In model I-DFM-2 the factors are forecast in a VAR model, where lags are selected with *Autometrics* with IIS. The same two step procedure applies in this case.
- $SubAggCT_i$ refers to the sub-aggregate formed by the components in the fully cointegrated subset i.

As explained in 6.1, the 164 components we are dealing with do not represent 100% of the CPI. After aggregating the components' forecasts we get a total weight around 90%, call this sub-aggregate CPI^*. To forecast the CPI we consider the following GUM:

$$\Delta CPI_t = c + \lambda_0 \Delta CPI_t^* + \sum_{k=1}^{K} \phi_k \Delta CPI_{t-k} + \sum_{k=1}^{K} \lambda_k \Delta CPI_{t-k}^* + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t, \qquad (6)$$

where S_{it} represent centered seasonal dummies which take the value $1 - 1/12$ in the i^{th} month and $-1/12$ otherwise. In this model, we select the regressors by *Autometrics* with Impulse Indicator Saturation and and apply the same two steps aforementioned procedure. We do not

select over CPI_t^*, so that it is always present in the model. The maximum number of lags, K, is 4.

Results of the indirect exercises are compared between each other and with five direct alternatives. The direct procedures differ between each other on the potential regressors considered. Models may include only the aforementioned eight sub-aggregates (*cat 1*), a more detailed disaggregation of 24 broad categories (*cat 2*), or no disaggregated prices at all. Besides, we when not including disaggregated prices, models may include the REER or not. Finally we also consider a direct alternative based on Dynamic Factor Models. Factors are extracted by principal components from the first difference of all the components and lags of this factor(s) included as regressors. In all cases we apply the two stpes procedure to select variables and impulses. The five direct exercises are summarized in Table 10.

Table 10: Summary of the direct forecasting exercises

	Model	**Description**
1	**D**	Direct baseline. Scalar model for the CPI $\Delta CPI_t = c + \sum_{k=1}^{K} \phi_k \Delta CPI_{t-k} + \phi_{12}\Delta CPI_{t-12} + \phi_{24}\Delta CPI_{t-24} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$
2	**D-DI-1**	Direct with disaggregated information 1. $\Delta CPI_t = c + \sum_{k=1}^{K} \phi_k \Delta CPI_{t-k} + \phi_{12}\Delta CPI_{t-12} + \phi_{24}\Delta CPI_{t-24} +$ $\sum_{i=1}^{8}\sum_{j=1}^{J} \theta_{i,j}\Delta SubAgg_{i,t-j} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$
3	**D-DI-2**	Direct with disaggregated information 2. $\Delta CPI_t = c + \sum_{k=1}^{K} \phi_k \Delta CPI_{t-k} + \phi_{12}\Delta CPI_{t-12} + \phi_{24}\Delta CPI_{t-24} +$ $\sum_{i=1}^{24}\sum_{j=1}^{J} \theta_{i,j}\Delta SubAgg_{i,t-j} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$
4	**D-REER**	Direct with REER. $\Delta CPI_t = c + \sum_{k=1}^{K} \phi_k \Delta CPI_{t-k} + \phi_{12}\Delta CPI_{t-12} + \phi_{24}\Delta CPI_{t-24} +$ $\sum_{q=1}^{Q} \delta_q \Delta REER_{t-q} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$
5	**D-DFM-1**	Direct with 1 Factor. $\Delta CPI_t = c + \sum_{k=1}^{K} \phi_k \Delta CPI_{t-k} + \phi_{12}\Delta CPI_{t-12} + \phi_{24}\Delta CPI_{t-24} +$ $\sum_{k=1}^{K} \delta_k F_{t-k} + \sum_{i=1}^{11} \gamma_i S_{i,t} + \epsilon_t$

- All the equations represent the initial GUMs from where models are selected using *Autometrics* with Impulse indicator saturation. The selection is carried out in two steps. First we use a target size of 0.5% to select variables, lags and impulses. Retained impulses are stored. In a second step we consider the same GUM augmented with the retained impulses and a target size of 5% with no IIS.
- $K = J = 4$ and $Q = 3$.
- In models 4 the factor is forecast in a scalar model, where lags are selected with *Autometrics* with IIS. The same two step procedure applies in this case.

Table 11 includes a summary of the forecasting results. The first row includes the root mean squared forecast error in the direct baseline for $\Delta_{12} log(CPI)$ for horizons $H = 1$ to $H = 12$. All

the other entries in the table are ratios with respect to the baseline. The evaluation period is $2010.1 - 2014.12$.

Forecasting results for direct approaches

Including disaggregated information (D-DI-1 and D-DI-2)

When using disaggregated information in a scalar model for the aggregate, as proposed by Hendry and Hubrich (2011), low disaggregation levels are preferred to higher ones[10]. For the lowest disaggregation level (8 sub-aggregates), while results are somewhat worse than in the baseline for short horizons, in longer ones (horizons 10 to 12) the use of disaggregated information improves over the baseline.

Scalar models enlarged with Dynamic Factors (D-DFM-1)

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

Forecasting results for indirect approaches

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

Univariate models for all the components (I-B)

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

Univariate models for all components enlarged with Dynamic Factors (I-DFM-2)

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

Pairwise procedures

In these procedures the basic components which do not belong to a subset sharing one common trend are forecast using univariate individual models. An alternative consists in forecasting just the sub-aggregate of those basic components ($I - PW - GP$). In short horizons the latter

[10]In Table 6 of Hendry and Hubrich (2011) there is another example of how the forecast accuracy deteriorates as the level of disaggregation increases.

alternative provides slightly better forecast for the aggregated CPI than the corresponding case in which basic components outside subsets with common features are forecast individually. For longer forecast the conclusion is the opposite.

The use of different normalizations has little effect in the RMSFE of the aggregate. The largest difference between the three possibilities is 3 points (at horizons 8 and 9).

The results using models which include common unit roots restrictions are better than in the basic indirect procedure ($I - B$). In particular the approach *I-CT-N1-GP* clearly dominates $I - B$ in the first four horizons[11]. In long horizons, when $I - B$ clearly dominates the baseline, the inclusion of common unit roots delivers some further improvements.

Interestingly enough we found that when not using IIS in the forecasting equations, the indirect procedures strongly deteriorate. This highlight the importance of the outliers and breaks treatment when dealing with disaggregates. In contrast, the direct procedure is not highly affected by the use of IIS.

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

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

Table 11: Relative RMSE $\Delta_{12}log(CPI)$. (First row: RMSE for the baseline. All the others are ratios with respect to the first)

	H=1	H=2	H=3	H=4	H=5	H=6	H=7	H=8	H=9	H=10	H=11	H=12
D	**0.22**	**0.39**	**0.50**	**0.52**	**0.56**	**0.59**	**0.67**	**0.74**	**0.80**	**0.83**	**0.88**	**0.96**
D-DI-1	1.03	1.02	1.05	1.09	1.10	1.04	1.02	1.04	1.01	0.97	0.92	0.91
D-DI-2	1.24	1.20	1.19	1.19	1.16	1.21	1.23	1.18	1.10	1.11	1.15	1.23
D-REER	1.00	0.99	0.99	0.98	0.97	0.98	1.00	0.99	0.99	1.00	1.00	1.00
D-DFM-1	0.98	0.97	0.96	0.97	0.98	0.99	1.01	1.00	1.00	0.98	0.96	0.97
I	1.07	1.05	1.08	1.12	1.10	1.05	0.96	0.89	0.78	0.66	0.62	0.61
I-CPI	1.15	1.13	1.12	1.13	1.11	1.06	0.99	0.91	0.80	0.69	0.63	0.59
I-DFM-2	1.43	1.30	1.24	1.26	1.23	1.22	1.18	1.18	1.16	1.11	1.06	1.00
PairWise												
I-CT-N1	1.07	1.05	1.08	1.12	1.08	1.04	0.95	0.87	0.75	0.65	0.61	0.59
I-CT-N2	1.06	1.05	1.07	1.11	1.09	1.05	0.96	0.90	0.78	0.67	0.63	0.62
I-CT-N3	1.08	1.06	1.08	1.11	1.08	1.03	0.95	0.88	0.77	0.66	0.62	0.60
I-CT-N1-GP	1.06	0.99	0.98	0.99	1.01	1.02	0.97	0.92	0.92	0.91	0.90	0.89
I-CT-N1-CPI	1.11	1.08	1.09	1.11	1.07	1.03	0.96	0.90	0.82	0.74	0.70	0.66
I-CT-N1-NoAgg	1.07	1.05	1.07	1.10	1.08	1.03	0.95	0.88	0.77	0.67	0.62	0.61
No IIS												
D-NoIIS	0.99	0.97	0.98	1.02	1.04	1.05	1.01	0.97	0.96	0.94	0.92	0.89
I-NoIIS	1.13	1.11	1.17	1.26	1.28	1.27	1.14	1.03	0.93	0.85	0.81	0.75

- See tables 9 and 10 for a description of each exercise.
- Dark red entrances highlight the loser procedure.
- Light red indicates procedures that are, at most, 5 points smaller than the worst one.
- Dark green indicates the best procedure.
- Light green indicates procedures that are, at most, 5 points larger than the best one.

7 Concluding Remarks

In this paper we proposed a robustification strategy for the pairwise approach to discover common trends proposed by Carlomagno and Espasa (2014), and applied it in a real data forecasting exercise. The robustification includes a strategy for dealing with data irregularities and with short samples issues as well as some extensions to the design of the pairwise cointegration tests.

The outliers' treatment combines the Impulse Indicator Saturation (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 new critical values. Outliers' dates are estimated by IIS and then these dates are used in the GLS procedure.

The outliers' correction strategy was analyzed in a simulation study. We found that using it 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 when outliers are not present (and no treatment is applied). When outliers' 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, in the simulation exercise we did not selected the lag structure in the IIS regressions, our results about *potency* must be considered as lower bounds for those that will be obtained when selecting the lags.

Comparing the results of correcting *vs.* not correcting for outliers in contaminated series we found that, although we have only lower bound potencies for the cases of estimated dates, estimating the dates is generally better than doing nothing.

Regarding small samples, we proposed a correction strategy and studied it by Monte Carlo. Results show that it provides significant *potency*[12] improvements, at the cost of a somewhat larger *gauge*[13]. This trade-off can be managed by what we called the *relaxation parameter* (λ), which determines the number of cointegration tests for which the nominal level is relaxed to a higher value. The larger λ is, the larger the *potency* improvements, and costs in terms of *gauge*. The user can select λ according to her preferences. In the application we used lambda equal two.

When applying the procedure to the US CPI we found that the groups generated by the pairwise procedure cannot be fully assigned to a single broad official category of prices but, in almost all the cases, more than 85% of the weight of the subset of fully cointegrated components is explained by two of them.

When forecasting the (aggregated) US CPI we found that while direct procedures dominate in short horizons (upto six periods ahead), indirect ones perform much better in longer horizons. From horizons 7 or larger the pairwise approach is the best performer.

Acknowledgments

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

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*.

Aron, J. and J. Muellbauer (2013). New methods for forecasting inflation, applied to the us*. *Oxford Bulletin of Economics and Statistics 75*(5), 637–661.

[12]The retention frequency of relevant variables.

[13]The retention frequency of irrelevant variables.

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.

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.

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

Carlomagno, G. and A. Espasa (2014). Discovering common trends in a large set of disaggregates: statistical procedures and their properties. *Working paper, Universidad Carlos III, Departamento de Estadística y Econometría 09*(13-14).

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

Castle, J. L., J. A. Doornik, and D. F. Hendry (2012). Model selection when there are multiple breaks. *Journal of Econometrics 169*(2), 239–246.

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

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

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

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

Doornik, J., D. Hendry, and F. Pretis (2013). Step-indicator saturation. *University of Oxford. Working Paper* (658).

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

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

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

Hansen, H. and S. Johansen (1999). Some tests for parameter constancy in cointegrated var models. *The Econometrics Journal 2*(2), 306–333.

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

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

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. (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), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, pp. 1–36. Oxford University Press.

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

Kiviet, J. F. and G. D. Phillips (1992). Exact similar tests for unit roots and cointegration. *Oxford Bulletin of Economics and Statistics 54*(3), 349–367.

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.

Lunnemann, P. and T. Mathä (2004). How persistent is disaggregate inflation? an analysis across eu 15 countries and hicp sub-indices. *Working paper, European Central Bank* (415).

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

Nielsen, B. and H. Nielsen (2008). Properties of estimated characteristic roots. *Univ. of Copenhagen Dept. of Economics Discussion Paper* (08-13).

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

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

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 & Economic Statistics 10*(3), 301–320.

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

Reis, R. (2006). Inattentive producers. *The Review of Economic Studies 73*(3), 793–821.

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

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.