
VOLUME 29, ISSUE 4
OCTOBER–DECEMBER 2013
ISSN 0169-2070
ELSEVIER
international journal of forecasting
Special Section: Flash Indicators
Guest Editors: Esther Ruiz, Miriam Scaglione and Dean Croushore
IIF
International Institute of Forecasters

International Journal of Forecasting 29 (2013) 733–735

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Some considerations about "Forecasting aggregates and disaggregates with common features"

Marcos Bujosa *, Alfredo García-Hiernaux [1]

Departamento de Fundamentos del Análisis Económico II. Facultad de Ciencias Econḿicas. Campus de Somosaguas, 28223 Madrid, Spain

ARTICLE INFO

Keywords:
Cointegration
Common trends
Multiple comparison procedures
Statistical power
Disaggregation map

ABSTRACT

Espasa and Mayo provide consistent forecasts for an aggregate economic indicator and its basic components as well as for useful sub-aggregates. To do so, they develop a procedure based on single-equation models that includes the restrictions arisen from the fact that some components share common features. The classification by common features provides a disaggregation map useful in several applications. We discuss their classification procedure and suggest some issues that should be taken into account when designing an algorithm to identify subsets of series that share one common trend.

1. Introduction

Espasa and Mayo (2013), hereafter EM, aim to provide coherent forecasts of both an aggregate economic indicator, such as a Consumer Price Index (CPI), and its basic components, as well as for useful sub-aggregates. To do this, they develop a procedure based on single-equation models that includes the restrictions which arise from the fact that some components share common features, typically common trends (Engle & Granger, 1987) and/or common serial correlations (Engle & Kozicki, 1993). This idea combines the feasibility and computational stability of single-equation models with the use of some of the additional information available as a result of the disaggregation.

The classification by common features provides a disaggregation map which will prove useful in several applications. For example, once the components with common features have been grouped, the authors build sub-aggregates from these components and use them to forecast the inflation in the Euro area, UK and USA. EM's procedure provides more precise forecasts than some other forecasts, whether indirect, based on basic components, or direct, based on the aggregate.

In their paper, Espasa and Mayo contribute to applied forecasting in several ways. From our point of view, their main contribution lies in the classification of the different components by common features. The authors demonstrate their procedure in Sections 4.1 and 4.2, and Figs. 1 and 2. Of particular interest is Fig. 2, which summarizes the classification algorithm.

2. Comments on EM's classification procedure

This section focuses on the procedure for identifying a subset of basic components with one common trend. EM propose four steps and a large number of cointegration tests using the methodology of Engle and Granger (1987). All of the aggregates are built using the official weights. The steps are as follows:

STEP 1 Identification of N1, the largest subset in which every element is cointegrated with all others; and the construction of its aggregate AN1.

STEP 2 All elements in N1 that are not cointegrated with AN1 *over a rolling window* are removed from the subset. The resulting set is called N2 and its aggregate AN2.

STEP 3 Any components outside N1 that are cointegrated with AN2 are incorporated into N2. The resulting subset and its aggregate are called N3 and AN3, respectively.

* Corresponding author. Tel.: +34 91 394 23 84; fax: +34 91 394 25 91.
E-mail addresses: marcos.bujosa@ccee.ucm.es (M. Bujosa), agarciah@ccee.ucm.es (A. García-Hiernaux).
[1] Tel.: +34 91 394 32 11; fax: +34 91 394 25 91.

0169-2070/$ – see front matter © 2012 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2012.10.001

STEP 4 All elements in N3 that are not cointegrated with AN3 *over a rolling window* are removed from the subset. The final set is called N, and its aggregate τ_{1t}.

Below, we discuss the main questions that arose when we analyzed this procedure.

1. *Should the significance level of the tests be adjusted because of the large number of tests*? As an example, the number of tests run for the USA is 25 440 (only in Step 1), as a result of testing 160 series for pairwise cointegration in both directions. Many of these tests are redundant, since N1 is the largest subset in which every element is cointegrated with all others and pairwise cointegration is a transitive property (see the Appendix).[2] Quoting Shaffer (1995): "*When many hypotheses are tested, and each test has a specified Type* I *error probability, the probability that at least some Type* I *errors are committed increases, often sharply, with the number of hypotheses. This may have serious consequences if the set of conclusions are evaluated as a whole*". Hence, EM's procedure is included in a large body of literature that is usually referred to as "multiple comparison procedures" or "simultaneous inference" (see, Rao & Swarupchand, 2009, for a detailed revision of the literature). A major part of this literature suggests methods for controlling the Type I error rate for any combination of true and false hypotheses. The most common method used in practice is the Bonferroni correction (see, e.g., Shaffer, 1986). An interesting question is whether the Bonferroni correction helps to improve EM's procedure. In what follows, we will explain why we do not think it does.

   In EM's procedure, the Bonferroni correction will reduce the significance level α for each individual test to $\alpha_f = \alpha/(k(k-1))$, where the denominator, $k(k-1)$, is the number of tests conducted. In EM's application, the numbers of series are $k = 79, 70, 160$ for the Euro area, the UK and the USA, respectively, and therefore α_f will be really small. Consequently, the tests will reject a smaller number of true cointegration relationships when using α_f than when using α, but will not reject a larger number of false cointegration relationships. This is the classical trade-off between Type I and Type II errors, which is aggravated here by the well-known low statistical power of the cointegration tests. However, EM's procedure requires a considerable level of statistical power, as the non-rejected series will be used to make up an aggregate for a comparison in the following steps. Therefore, it is extremely important that the series forming the aggregate are truly cointegrated, otherwise this aggregate will be a mixture of different common trends and the procedure will not work properly. Statistically speaking, in EM's procedure, Type II errors are much more harmful than Type I. The Bonferroni correction is therefore likely to lead to a wrong initial aggregate, which will spoil the results.

2. *Do we really need all of these steps and tests to get the final set N and the corresponding aggregate*, τ_{1t}? The answer to this question is uncertain, due to the low power of the cointegration tests. For example, Step 2 attempts to clear N1 of possible non-cointegrated series, i.e., to reduce the Type II error committed in Step 1. However, the user should be careful here, since Step 2 uses the aggregate computed in Step 1. As was mentioned above, it is crucial that the aggregate based on components of N1 be made up of truly cointegrated series. The user should probably be more loose with the individual significance level throughout the whole procedure in order not to expand the Type II errors. Accordingly, Steps 3 and 4 can be interpreted using the same statistical approach. Step 3 attempts to reduce the Type I error, which is certainly higher than the 5% level individually assumed by the authors, as they do not take the large number of tests into account (recall, again, the idea of the Bonferroni correction); while Step 4 is another attempt to reduce Type II errors. Hence, EM propose an iterative procedure that adds and removes series from a new set in each step, as a way of improving the low statistical power of the cointegration tests.
3. *Are all of these steps and tests enough to get the final set N? That is, does EM's procedure converge*? Convergence is a suitable property that ensures that the algorithm stops at some point. Unfortunately, we do not know whether EM's algorithm converges. The authors stop their procedure in Step 4, but they do not prove that this choice is optimal. As a matter of fact, it cannot be generalized that stopping in Step 4 is always going to produce better (or worse) results, as doing additional steps could lead to a different final set N (recall that some basic components enter and others exit at each step).
4. *Is the largest subset of basic components the best choice*? Two different issues arise in relation to this question. First, as the final aim is to forecast the aggregate, it would be more helpful to choose the subset that adds the most predictability to the aggregate (using some information criterion, load factors, etc.). To do this would mean that the classification procedure required all the groups to be identified and one (or several) to be chosen from among them. For that reason, an algorithm that finds several groups simultaneously would be preferable. Second, bearing in mind that Type II errors are much more harmful than Type I errors in the classification process, it should be noticed that the largest subset may also be heterogenous as a consequence of the low power of the cointegration tests.

3. Some considerations for improving EM's classification procedure

Based on the previous section and the procedure proposed by EM, we suggest some guidelines that could be helpful when designing an algorithm for identifying subsets of series with one common trend. (a) Although the method belongs to the literature on "multiple comparison procedures" or "simultaneous inference", the framework is fairly different. In this case, Type II errors are much more

[2] If x_t and y_t are cointegrated CI(1, 1), and z_t and x_t are CI(1, 1), then y_t and z_t are also CI(1, 1).

harmful than Type I errors, and therefore, unusually high significance levels should be applied. (b) Since pairwise cointegration is transitive, this property could be used to decide when the results of the cointegration tests are ambiguous. Nevertheless, practitioners should be aware of the risk of propagating false pairwise cointegration relationships. (c) The aggregates should be built using weights that ensure that they are *pairwise cointegrated* with all their components. The official weights of the CPIs do not assure this property. (d) The convergence of the procedure would be a suitable property, although probably one which would be hard to demand. In any case, both the stopping criterion of the algorithm and its consequences for the final subset deserve special attention. (e) If the final goal is forecasting, the subsets should be chosen using a predictive accuracy criterion, instead of a criterion related to the size. (f) Hence, the procedure should be able to provide several subsets simultaneously, allowing the level of predictability that each group adds to the aggregate to be determined.

4. Concluding remarks

Espasa and Mayo (2013) make a significant contribution to the literature on forecasting aggregates and disaggregates by taking the stable common features in the basic components into account. Especially appealing is the classification of basic components that share a common trend. Although the authors show that their classification procedure leads to better forecasts than other alternatives, this comment aims to provide some guidelines that could improve the classification procedure, and, as a consequence, increase the forecast accuracy. Espasa and Mayo's methodology for classifying series with common features is very useful for practitioners and can be applied in many different situations. This topic, which is open to future research, appears very promising.

Acknowledgments

The authors are indebted to Miguel Jerez and Antonio García-Ferrer for advice on the paper. The research was supported by grants ECO2009-10287 and ECO2011-23972 from the Spanish Ministerio de Ciencia e Innovación, and the Ramón Areces Foundation.

Appendix

Lemma 1. *Let y_{1t}, y_{2t} and x_t be integrated of order one,* I(1)*. If y_{1t} and x_t are cointegrated* CI(1, 1)*, and y_{2t} and x_t are* CI(1, 1)*, then y_{1t} and y_{2t} are also* CI(1, 1).

Proof of Lemma 1. Let y_{1t}, y_{2t} and x_t be integrated of order one, I(1). Also, let y_{1t} and x_t be CI(1, 1), and y_{2t} and x_t be CI(1, 1), as:

$$\begin{aligned} y_{1t} &= \alpha_0 + \alpha_1 x_t + \varepsilon_{1t}; \qquad \phi_1(B)\varepsilon_{1t} = \theta_1(B)a_{1t}, \\ &\quad \text{with } a_{1t} \sim \text{i.i.d.}N(0, \sigma_1^2), \qquad (1) \\ y_{2t} &= \beta_0 + \beta_1 x_t + \varepsilon_{2t}; \qquad \phi_2(B)\varepsilon_{2t} = \theta_2(B)a_{2t}, \\ &\quad \text{with } a_{2t} \sim \text{i.i.d.}N(0, \sigma_2^2), \qquad (2) \end{aligned}$$

where all of the roots of $\phi_i(B) = 0$, for $i = 1, 2$, are outside the unit circle.

Solving for x_t, Eqs. (1) and (2) can be written as:

$$x_t = \frac{1}{\alpha_1}\left(y_{1t} - \alpha_0 - \psi_1(B)a_{1t}\right) \qquad (3)$$

$$x_t = \frac{1}{\beta_1}\left(y_{2t} - \beta_0 - \psi_2(B)a_{2t}\right), \qquad (4)$$

where $\psi_i(B) = \theta_i(B)/\phi_i(B)$. Finally, solving Eqs. (3) and (4) for y_{1t}, we get:

$$y_{1t} = \gamma_0 + \gamma_1 y_{2t} + \eta_t \qquad (5)$$

where $\gamma_0 = \alpha_0 - \alpha_1\beta_0/\beta_1$, $\gamma_1 = \alpha_1/\beta_1$ and $\eta_t = \psi_1(B)a_{1t} - \gamma_1\psi_2 a_{2t}$. If all of the roots of $\phi_i(B)$ are outside the unit circle, then η_t is stationary and y_{1t} and y_{2t} are CI(1, 1). □

References

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica*, *55*, 251–276.
Engle, R. F., & Kozicki, S. (1993). Testing for common features. *Journal of Business and Economic Statistics*, *11*(4), 369–395.
Espasa, A., & Mayo, I. (2013). Forecasting aggregates and disaggregates with common features. *International Journal of Forecasting*, *29*(4), 718–732.
Rao, C. V., & Swarupchand, U. (2009). Multiple comparison procedures: a note and a bibliography. *Journal of Statistics*, *16*, 66–109.
Shaffer, J. P. (1986). Modified sequentially rejective multiple procedures. *Journal of the American Statistical Association*, *81*(395), 826–831.
Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, *46*, 561–584.

Marcos Bujosa is Associate Professor at Universidad Complutense de Madrid. He obtained his Ph.D. in Economics at Universidad Autonóma de Madrid in 2001. His research interests include modelling in the frequency domain and forecasting seasonal economic time series. On these topics, he has published in the *International Journal of Forecasting*, *Computational Statistics and Data Analysis*, and *Journal of Forecasting*.

Alfredo García-Hiernaux is Associate Professor at Universidad Complutense de Madrid. From 2006 to 2008 he was Assistant Professor at the Universidad Carlos III de Madrid. He obtained his Ph.D. in Economics at Universidad Complutense de Madrid in 2005. His research interests include state space models, subspace methods and forecasting economic time series. He has published in various journals, including the *Journal of Time Series Analysis*, *Computational Statistics*, the *Journal of Statistical Computation and Simulation*, and *Mathematics and Computers in Simulation*.