FACULTAD DE CIENCIAS ECONÓMICAS

Y EMPRESARIALES

IDENTIFYING SERIES WITH COMMON TRENDS TO IMPROVE FORECASTS OF THEIR AGGREGATE

Marcos Bujosa
Alfredo García-Hiernaux

Working Papers / Documentos de Trabajo. ISSN: 2255-5471
DT CCEE-1302 Abril 2013
http://eprints.ucm.es/20792

IDENTIFYING SERIES WITH COMMON TRENDS TO IMPROVE FORECASTS OF THEIR AGGREGATE

Abstract:
Espasa and Mayo-Burgos (2013) provide consistent forecasts for an aggregate economic indicator 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 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 procedure and suggest some issues that should be taken into account when designing an algorithm to identify subsets of series with one common trend. We also provide a naive algorithm following those suggestions.

Keywords: Cointegration, Common trends, Multiple Comparison Procedures, Statistical power, Disaggregation map.

IDENTIFICACIÓN DE SERIES CON TENDENCIAS COMUNES PARA MEJORAR LAS PREVISIONES DE AGREGADOS

Resumen:
Espasa y Mayo-Burgos (2013) proporcionan pronósticos consistentes de algunos indicadores económicos compuestos, así como de sus componentes básicos y sub-agregados de interés. Para ello desarrollan un procedimiento uniecuacional con restricciones que incorporan en el modelo la existencia de características comunes en algunos de los componentes básicos del agregado. La clasificación de los componentes básicos según rasgos comunes es de gran utilidad en algunas aplicaciones, por ejemplo en la previsión. En este documento discutimos el procedimiento de clasificación de Espasa y Mayo-Burgos, y sugerimos algunas recomendaciones que deberían tenerse en cuenta al diseñar algoritmos de identificación de conjuntos de series con tendencia común. Siguiendo dichas recomendaciones, proporcionamos un ingenuo algoritmo de clasificación.

Palabras clave: Cointegración, Tendencias Comunes, Test de Comparaciones Múltiples, Potencia estadística, Mapa de desagregación.

Materia: Métodos Matemáticos y Cuantitativos.
JEL: D24, R12, R40

Marcos Bujosa
Dpto. de Fundamentos del Análisis Económico II.
Facultad CC. Económicas y Empresariales.
Univ. Complutense de Madrid
Campus de Somosaguas, s/n
28223 POZUELO DE ALARCÓN (MADRID)
mbujosab@ucm.es

Alfredo García-Hiernaux
Dpto. de Fundamentos del Análisis Económico II.
Facultad CC. Económicas y Empresariales.
Univ. Complutense de Madrid
Campus de Somosaguas, s/n
28223 POZUELO DE ALARCÓN (MADRID)
agarciah@ucm.es

Marzo2013 (fecha de recepción)

Identifying series with common trends to improve forecasts of their aggregate

April 5, 2013

Abstract

Espasa and Mayo-Burgos (2013) provide consistent forecasts for an aggregate economic indicator 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 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 procedure and suggest some issues that should be taken into account when designing an algorithm to identify subsets of series with one common trend. We also provide a naïve algorithm following those suggestions.

1 Introduction

Espasa and Mayo-Burgos (2013), hereinafter EM, aim to provide coherent forecasts for 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 arisen from the fact that some components share common features, typically, common trends (Engle and Granger, 1987) and/or common serial correlations (Engle and Kozicki, 1993). This idea combines the feasibility and computational stability of single-equation models and the use of part of the information available by the disaggregation.

The classification by common features provides a disaggregation map 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 indirect forecasts, based in basic components, or direct forecasts, based on the aggregate.

In their paper, Espasa and Mayo provide several contributions for applied forecasters. From our point of view, the main one lies in the classification of the different components by common features.

2 Comments on EM's classification procedure

This section focuses on the procedure to identify a subset of basic components with one common trend. EM propose four steps and a large number of cointegration tests using Engle and Granger (1987) methodology. All 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 each other; and construction of its aggregate AN1.

STEP 2 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 Components outside N1 that are cointegrated with AN2 are incorporated to N2. The resulting subset and its aggregate are called N3 and AN3.

STEP 4 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 comment the main questions that arose when we analyzed the 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 25440 (only in Step 1), which results from testing for pairwise cointegration in both directions 160 series. Many of these tests are redundant, since N1 is the largest subset in which every element is cointegrated with each other and pairwise cointegration is a transitive property (see Appendix).[1] 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 be evaluated as a whole.*" Hence, EM's procedure is included in a large literature that is usually called "Multiple Comparison Procedures" or "Simultaneous Inference" (see, Rao and Swarupchand, 2009, for a detailed revision of the literature). A major part of this literature suggests methods to control 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 the following we will explain why we do not think so.

   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 number of series is $k = 79, 70, 160$ for the Euro Area, UK and USA, respectively, and therefore α_f will be really small. Consequently, the tests will reject a smaller number of true cointegration relations when using α_f than when using α, but will not reject a larger number of false cointegration relations. This is the classical trade-off between Type I and Type II errors, which is aggravated by the well-known low statistical power of the cointegration tests. However, EM's procedure requires a considerable statistical power, as the non-rejected series will be used to make up an aggregate to compare with in the following steps. Therefore, it is extremely important that the series forming the aggregate are truly cointegrated, otherwise their aggregate will be a mixture of different common trends and the procedure will not work properly. Statistically speaking, in EM's procedure Type II error is much more harmful than Type I. Therefore, the Bonferroni correction will probably lead to a wrong initial aggregate and will spoil the results.

2. *Do we really need all these steps and tests to get the final set N and the corresponding aggregate, τ_{1t}?* The answer 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 said above, it is crucial that the aggregate based on components of N1 is made up of truly cointegrated series. Probably the user should be more loose with the individual significance level in order not to expand the Type II errors throughout the whole procedure. Accordingly, Steps 3 and 4 can be interpreted with the same statistical approach. Step 3 attempts to reduce Type I error, which is certainly higher than the 5% level individually assumed by the authors, as they do not take into account the large number of tests (recall, again, the

[1] 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).

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 takes out series from a new set in each step, as a way to improve the low statistical power of the cointegration tests.

3. *Are all these steps and tests enough to get the final set N?, i.e., Does EM's procedure converge?* Convergence is a suitable property that assures 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 proof that their choice is optimal. As a matter of fact, it cannot be generalized that stopping in Step 4 is going to produce better (or worse) results, as some additional steps could lead to a different final set N (recall that some basic components go in and others go out in each step).

4. *Is the largest subset of basic components the best choice?* Two different issues arise regarding this question. First, as the final aim is forecasting the aggregate, it would be more helpful to choose the subset that adds more predictability to the aggregate (using some information criteria, load factors, etc). To do so, the classification procedure would require identifying all the groups and choose one (or several) among all of 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 Suggestions to improve the classification procedure

From the previous section and the procedure proposed by EM, we suggest some guidelines that could be helpful when designing an algorithm to identify subsets of series with one common trend: (a) Although the method belongs to the "Multiple Comparison Procedures" or "Simultaneous Inference", the framework is pretty different. In this case, Type II errors are much more harmful than Type I and, therefore, unusually high significance levels should be applied. (b) Since pairwise cointegration is transitive, this property could be used to make a decision 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 some weights that assure they are *pairwise cointegrated* with all their components. Official weights of the CPIs do not assure this property. (d) The convergence of the procedure would be a suitable property, although probably hard to demand. In any case, the stopping criterion of the algorithm and its consequences on the final subset should deserve a 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; and hence, (f) the procedure should be able to provide several subsets simultaneously, allowing to compare the predictability that each group add to the aggregate.

4 A naïve classification procedure

In this section we propose an alternative classification procedure based on the transitivity of *pairwise cointegration*. This property allows us to extend a reduced set of "highly likely" *pairwise cointegrated* series by completing with transitivity using graph theory. The main flaw of this approach is that type II errors may lead to bad results. For instance, if the test fails to reject a cointegration relationship, the extension using the transitivity property will expand this error and mix two different subsets into a new heterogeneous one. Therefore, using *pairwise cointegration* transitivity requires a stronger evidence of cointegration. Instead of usual significance levels, we suggest to start the tests by applying a extremely high significance levels, even close to the unity. Below, we briefly describe a proposal that it is far to be perfect, but it attempts to follow the guidelines given in Section 3. The algorithm consists of two parts: the computation of all *pairwise cointegration tests* and the *main loop*.

A.- Compute all *Pairwise Cointegration Tests*, and build the matrix $\mathbf{M_{pct}}$. The element (i, j) is the test statistic between x_i and x_j series. As EM, we build a symmetric matrix, $\mathbf{M_{pct}}$, such as the element (i, j) and the element (j, i) are equal to the maximum of the Engel-Granger test statistics computed in both directions. Each element on the diagonal must be a "$-\infty$". Additionally, we store the estimated slopes of all the auxiliary regressions of the Engle-Granger cointegration tests in the matrix $\mathbf{M}_\beta$. We use them in the aggregation process.

B.- Run the *main loop*: Fix an extremely low initial critical value, e.g., $k_0 < -9$, and gradually increase the critical value in each loop ($k_{j+1} = k_j + \epsilon_j$).[2] Then:

STEP 1 Build an *Adjacency Matrix* (**AM**): for each pair of series (or sub-aggregates), write 1 if the cointegration is not rejected and 0 otherwise (as in EM's procedure). Complete the matrix by transitivity.

STEP 2 Identify subsets of series in which all elements are *pairwise cointegrated* one with each other. Use the estimated slopes saved in $\mathbf{M}_\beta$ to build an aggregate that is *pairwise cointegrated* with each of its components.

STEP 3 Substitute each set obtained in STEP 2 by its aggregate and compute all the *pairwise cointegration* tests. GO TO STEP 1.

This algorithm fits guidelines 1 to 4 presented in Section 3. It also converges, but to a matrix full of ones, i.e., it places all the series in the same group. Hence, a stopping rule is required. In our Montecarlo experiments, we use a loop with 28 critical values between -9.0 and -2.6.[3] This is a heuristic decision and the search of an optimal stopping criterion is an interesting subject of future research. For the moment, our advise is to check the results every few loops and decide to continue or stop.

5 Three simulated examples

In this section we show the behavior of the algorithm with three simulation exercises. In the first example, the above heuristic criterion leads to a premature stop; in the second one, the rule works fine; while in the last one, the same criterion is not sufficient to stop the propagation of the type II errors. As a result, two large heterogeneous subsets are found.

In each example, we simulate five random walks with a sample size of 300 observations. Then, for each random walk we build ten *pairwise cointegrated* series. Series with indices from 1 to 10 are pairwise cointegrated with each other, series with indices from 11 to 20 are pairwise cointegrated with each other, the same for series with indices from 21 to 30 and so on.

Table 1 shows identified groups with four or more series. The left plots on Figure 1 show the five original random walks for each example, the plots on the center depict the fifty series, and the plots on the right show the aggregates of the identified subsets with four or more series.

In Example 1 the two first groups contain series with indices from 1 to 10 that should appear in only one group (see Table 1). On the other hand, the third group is a mixture of series with

[2]Notice that the critical value suggested here corresponds to a significance level much higher than 0.999 and a sample size of 10.

[3]$k = -2.6$ is the critical value for a sample size of 300 and $\alpha = 0.9$.

groups	Series in the group											
	Example 1: The heuristic rule has stopped the loop too soon											
1st	1	5	6	8								
2nd	2	3	9	10								
3rd	**4**	**7**	**34**	**37**								
4th	11	15	16	18								
5th	12	13	19	20								
6th	21	22	23	25	26	28	29	30				
7th	**24**	**27**	**44**	**47**								
8th	31	35	36	38								
9th	32	33	39	40								
10th	41	45	46	48								
11th	42	43	49	50								
	Example 2: Stopping on time											
1st	2	4	5	7	8	9	10					
2nd	11	12	14	15	17	19	20					
3rd	21	22	24	25	27	29	30					
4th	31	32	34	35	37	39	40					
5th	41	42	44	45	47	49	50					
	Example 3: The heuristic rule has stopped the loop too late											
1st	**2**	**9**	**11**	**13**	**14**	**16**	**17**	**19**	**20**			
2nd	**3**	**5**	**8**	**12**	**15**	**18**	**25**	**28**	**35**	**38**	**45**	**48**

Table 1: Subsets identified in the three examples. We only show the subsets with four or more series for simplicity's lack. The false pairwise cointegrated subsets are in bold.

different trends. However, the sixth group includes eight out of ten series that share the same trend. Likewise, Figure 1 shows that there are different groups with a common trend. Results in both, Table 1 and Figure 1 evidence that the algorithm stopped too soon.

Table 1 and Figure 1 indicate that in Example 2 the procedure behaves pretty well. Note that in this case there are no type II errors (see Table 1) and the aggregates almost replicate the corresponding left plot of Figure 1.

In Example 3 the two largest subsets are heterogeneous. Nevertheless, there are six subsets (not shown in the table neither the figure), with three series each, that are correctly identified; i.e., their series are truly pairwise cointegrated.

These results illustrate a general rule of the procedure: it is better to stop the loop sooner than later. Two advantages of this proposal are that all subsets are built simultaneously and it is easy to record the sequence in which they are formed. Hence, it is possible to analyze the sequence and decide when the loop should stop in each practical exercise. In any case, one should check if the final groups are homogeneous (studying the graphs of its components, using additional tests, etc). Note that this naïve algorithm is somehow blind, as it only works with the distance between series measured with a test statistic that has low power. The simulation exercise also highlights that a large subset is not always the best subset to use.[4]

6 Concluding remarks

Espasa and Mayo-Burgos (2013) provide a significant contribution to the literature on forecasting aggregates and disaggregates by taking into account the stable common features in the basic components. 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 with respect to other alternatives, this comment aims at giving some guidelines that could improve the classification procedure and, as a consequence, gain forecast accuracy. Espasa and Mayo's methodology to classify series with common features is very useful for practitioners and can be applied in many different situations. This topic, open to future research, seems to be very promising. We have made some suggestions and provided an alternative algorithm of classification.

[4]The code of the proposal and the simulation exercise is available from the authors upon request.

A 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). Let also y_{1t} and x_t be CI(1,1), and y_{2t} and x_t be CI(1,1), as:

$$y_{1t} = \alpha_0 + \alpha_1 x_t + \varepsilon_{1t}; \quad \phi_1(B)\varepsilon_{1t} = \theta_1(B)a_{1t}, \quad \text{with } a_{1t} \sim iidN(0, \sigma_1^2) \tag{1}$$

$$y_{2t} = \beta_0 + \beta_1 x_t + \varepsilon_{2t}; \quad \phi_2(B)\varepsilon_{2t} = \theta_2(B)a_{2t}, \quad \text{with } a_{2t} \sim iidN(0, \sigma_2^2) \tag{2}$$

where all the roots of $\phi_i(B) = 0$, for $i = 1, 2$, are outside the unit circle.
Solving for x_t, equations (1) and (2) can be written:

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

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

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

$$y_{1t} = \gamma_0 + \gamma_1 y_{2t} + \eta_t \tag{5}$$

where $\gamma_0 = \alpha_0 - \alpha_1\beta_0/\beta_1$, $\gamma_1 = \alpha_1/\beta_1$ and $\eta_t = \psi_1(B)a_{1t} - \gamma_1\psi_2 a_{2t}$. All 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). $\square$

References

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, (55):251–276.

Engle, R. F. and Kozicki, S. (1993). Testing for common features. *Journal of Business and Economic Statistics*, 11(4):369–395.

Espasa, A. and Mayo-Burgos. I. (2013) Forecasting aggregates and disaggregates with common features. *International Journal of Forecasting*, http://dx.doi.org/10.1016/j.ijforecast.2012.10.004

Rao, C. V. and 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.

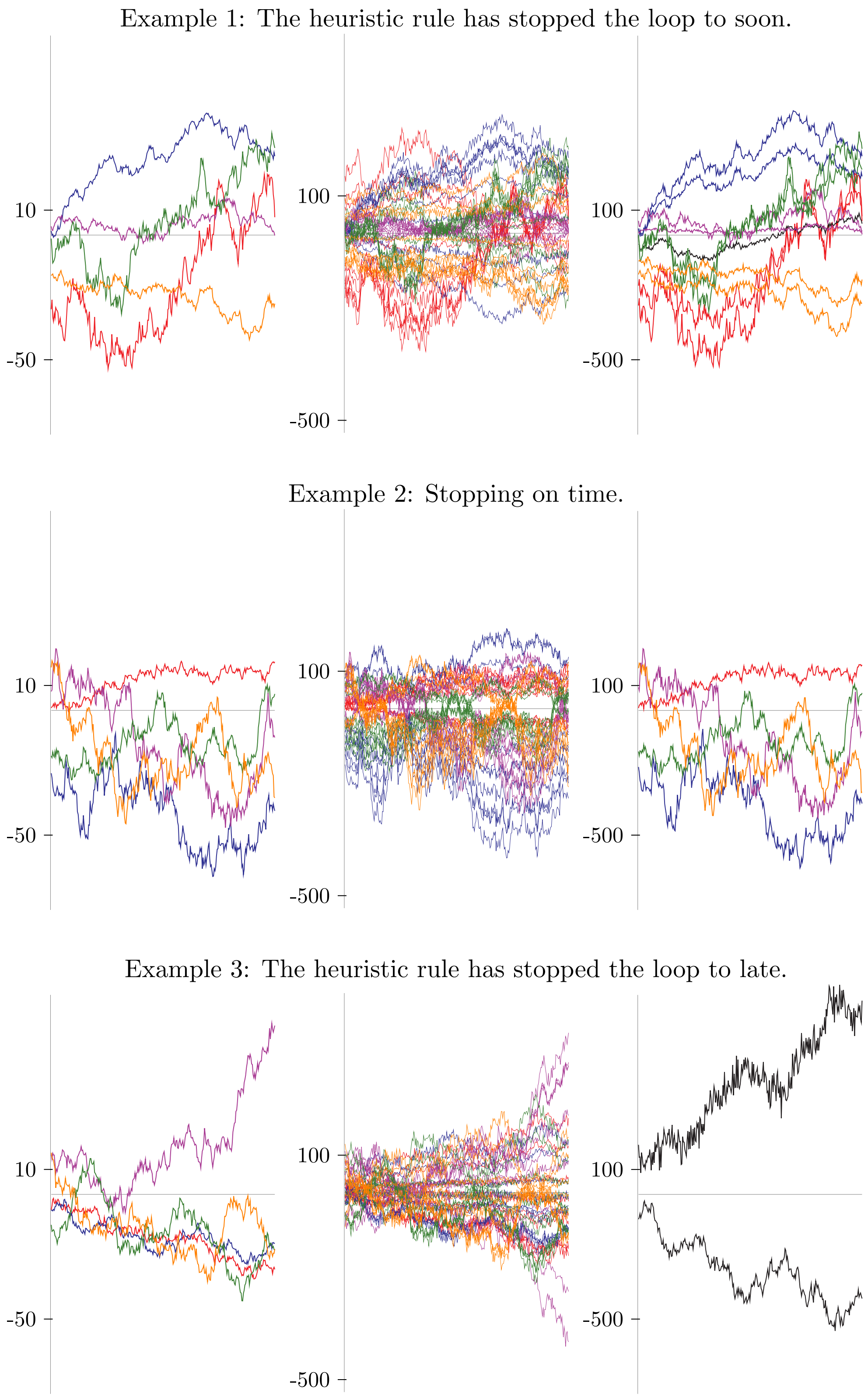

Figure 1: Trends, series and final aggregates for the three examples. We only show the aggregates of subsets with four or more series. The column on the left shows the five original random walks for each example, the column on the center depicts the 50 series, and the column on the right shows the aggregates of the subsets.