

**2009/20**



What do we know about comparing aggregate  
and disaggregate forecasts?

Giacomo Sbrana and Andrea Silvestrini

**CORE**

Voie du Roman Pays 34

B-1348 Louvain-la-Neuve, Belgium.

Tel (32 10) 47 43 04

Fax (32 10) 47 43 01

E-mail: [corestat-library@uclouvain.be](mailto:corestat-library@uclouvain.be)

<http://www.uclouvain.be/en-44508.html>

CORE DISCUSSION PAPER  
2009/20

**What do we know about comparing aggregate  
and disaggregate forecasts?**

Giacomo SBRANA<sup>1</sup> and Andrea SILVESTRINI<sup>2</sup>

March 2009

**Abstract**

This paper compares the performance of “aggregate” and “disaggregate” predictors in forecasting contemporaneously aggregated vector ARMA processes. An aggregate predictor is built by forecasting directly the aggregate process, as it results from contemporaneous aggregation of the data generating vector process. A disaggregate predictor is obtained by aggregating univariate forecasts for the individual components of the data generating vector process. The necessary and sufficient condition for the equality of mean squared errors associated with the two competing methods is provided in the bivariate VMA(1) case. Furthermore, it is argued that the condition of equality of predictors as stated in Lütkepohl (1984b, 1987, 2004) is only sufficient (not necessary) for the equality of mean squared errors. Finally, it is shown that the equality of forecasting accuracy for the two predictors can be achieved using specific assumptions on the parameters of the VMA(1) structure. Monte Carlo simulations are in line with the analytical results. An empirical application that involves the problem of forecasting the Italian monetary aggregate M1 in the pre-EMU period is presented to illustrate the main findings.

**Keywords:** contemporaneous aggregation, forecasting.

**JEL Classification:** C10, C32, C43, C52

---

<sup>1</sup> United Nations, Department of Economic and Social Affairs, New York, NY 10017, USA.

<sup>2</sup> Bank of Italy, Economics, Research and International Relations, Roma, Italy; Université catholique de Louvain, CORE, B-1348 Louvain-la-Neuve, Belgium.

The authors would like to thank Luc Bauwens, Christian M. Hafner, Marco Lippi, Helmut Lütkepohl, Fatemeh Shadman Mehta and David Veredas for useful comments and suggestions. The views expressed in this paper are those of the authors and do not necessarily reflect the positions or policies of the Institutions in which they work.

This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.



## 1 Introduction

This paper focuses on the issue of forecasting contemporaneous time series aggregates. Our study is motivated by the practical problem of predicting aggregate economic variables like private consumption in national accounting, which is the sum of individual private consumptions over all households. Another interesting example is presented by Lütkepohl (2007): the gross domestic product (GDP) in one year is the sum of private consumption, gross private investment, government purchases, and net exports for that year. Assume that we are interested in predicting at the “macro” level. The main questions we try to answer may be summarized as follows: should we directly forecast the GDP or should we project its subcomponents and sum the forecasts? Under what conditions do these two prediction methods deliver equal accuracy?

There are many other relevant cases in virtually every field of economics. Consider Euro-zone inflation forecasting. As is well known, Euro area inflation (overall index) is a contemporaneous aggregation along the countries and along the sectors. Therefore, as underlined by Espasa, Albacete and Senra (2002) and Benalal *et al.* (2004), several methods can be used to predict this aggregate. For instance, it is possible to compare forecasts based on the aggregate overall harmonized index of consumer prices (HICP) and on the main HICP subindexes (energy, unprocessed food, etc.) at the area-wide level. Alternately, it makes sense to compare predictions based on the aggregate HICP overall index and on the HICP overall index for each of the Euro area countries.

The problem of forecasting aggregate variables using competitive predictors has been extensively discussed in econometrics. As a consequence, there is abundant literature on the effects of contemporaneous aggregation on forecasting. This literature has basically followed two related strands of research: an empirical one and a theoretically oriented one.

On the empirical side, Fagan and Henry (1998) and Dedola, Gaiotti and Silipo (2001) focus on the informational content of national contributions to model and estimate a money demand equation for the Euro area. Both papers find that an area-wide equation has superior properties than equations estimated at the national level. In a different context, Bodo, Golinelli and Parigi (2000) and Zizza (2002) compare several disaggregate and aggregate predictors of the industrial production index for the Euro area. Zellner and Tobias (2000) examine the effects of aggregation in forecasting the median growth rate of eighteen industrialized countries. Espasa, Albacete and Senra (2002) evaluate a disaggregate approach (by countries and by sectors) in forecasting Euro area inflation and show that further improvements in forecasting the aggregate can be obtained by working at the disaggregate level.

More recently, Marcellino, Stock and Watson (2003) consider the problem of forecasting four Euro area variables (inflation, real GDP, industrial production and unemployment) pooling country-specific forecasts and directly forecasting the aggregate variables. Baffigi, Golinelli and Parigi (2004) study the choice of the level of model aggregation in forecasting the Euro area GDP. Hubrich (2005) compares the precision of forecasting directly the aggregate inflation as opposed to aggregating forecasts for inflation subindexes. Hsiao, Shen and Fujiki (2005) raise the issue of whether Japan has a stable money demand function using both aggregate and disaggregate

data. Sbrana (2008a) focuses on the forecasting accuracy of Euro area and specific national models in forecasting aggregate money demand. No attempt is made in this introduction to give an exhaustive survey of all the contributors to the aggregation debate.

Turning to the theoretical papers, the first studies of the aggregation problem date back to Theil (1954) and Grunfeld and Griliches (1960). Theil shows that using the structural information at the disaggregate level, it is possible to improve the model specification of the aggregate variable. On the contrary, Grunfeld and Griliches argue that contemporaneous aggregation is not necessarily bad if the equations at the disaggregate level are not correctly specified or if the micro data are subject to large errors. In the context of ARMA and vector ARMA models, other relevant papers are Rose (1977), Tiao and Guttman (1980), Wei and Abraham (1981), Kohn (1982), Lütkepohl (1984a, 1984b, 1984c, 1987, 2004), Pino, Moretton and Mentz (1987). Most of the theoretical results are collected in Chapter 4 of Lütkepohl (1987) and in Section 2.4 of Lütkepohl (2004). Again, no attempt is made to survey this theoretical literature. We refer to Granger (1990) for a comprehensive review of several aggregation schemes and special topics relevant in the time series context. The interested reader can go back to Lütkepohl (1984b, 1987, 2004) and Giacomini and Granger (2004) for more formal work on contemporaneous aggregation and forecasting.

Our paper is related to this stream of theoretical literature. To gain some insight into our key focus, let us refer to Wei and Abraham (1981), among others. According to these authors, to forecast a contemporaneous aggregate of disaggregate variables, predictors can be built from a) the whole multivariate process, b) the aggregate process or c) the individual components of the multivariate process. If the data generating process (DGP) is known and no estimation uncertainty is faced, it has been established in the literature that the approach in a) is optimal, in the sense that it delivers the smallest forecast mean squared error (MSE) with respect to the methods in b) and c).<sup>1</sup> Intuitively, this is not surprising, since the predictor in a) uses the largest information set (Lütkepohl, 2004). This result is formally proven by Wei and Abraham (1981) and Lütkepohl (1984b, 1987).

Notice that most of the papers rank the forecasting approaches in a) and b). Rose (1977), p. 377, is the first to give a necessary and sufficient condition for equal forecasting efficiency of the two predictors built following a) and b). This condition is provided assuming that the interest is in forecasting an aggregate of independent ARIMA models and that no problems of identification or estimation have to be faced. Tiao and Guttman (1980) assume a stationary vector moving average DGP and state (Theorem 1 on p. 223) a necessary and sufficient condition for equal efficiency of predictors based on a) and b). The proof of the theorem makes use of linear algebra, and a nice geometric interpretation of the result is provided. Kohn (1982) specifies a condition under which forecasts of the aggregate variable drawn from methods in a) and b) are equal in terms of MSE (see Theorem 1, p. 339), assuming that the DGP is a Gaussian second-order stationary process. This result is extended to ARMA processes and to the ARIMA case. A procedure to test the equivalence condition is also illustrated.

However, as noted by Lütkepohl (1987), a multivariate model is more difficult to estimate than

---

<sup>1</sup>As noted by Lütkepohl (2007), this result does not necessarily carry over to estimated processes.

a univariate one, due to the curse of dimensionality. Therefore, forecasting procedures b) and c) are often used as an alternative to a). That is, in practice, we have two choices about how to proceed:

1. Forecasting univariately each disaggregate subcomponent (individual component) and contemporaneously aggregating these forecasts (*disaggregate approach*);
2. Forecasting the contemporaneously aggregated variable directly (*aggregate approach*).

In general, it is unclear which of these two methods outperform the other in forecasting. For instance, Wei and Abraham (1981) construct an example in which the predictor based on the contemporaneously aggregated process has lower MSE than the one based on the subcomponents (individual components) series aggregated into a single prediction for the aggregate variable (see Example 3 in their paper). Another case is proposed in which the opposite holds true (Example 2). Two exceptions are the already cited Theorem 2 in Lütkepohl (1984b) and Proposition 4.1 - Corollary 4.1.1 in Lütkepohl (1987), where a necessary and sufficient condition for the equality of h-step ahead predictors is presented.<sup>2</sup> However, in this latter monograph, it is also argued that whether or not one predictor outperforms the other in terms of MSE “Depends on the structure of the disaggregate process and the aggregation matrix” (p. 104). To our knowledge, no further guidelines are given in the literature.

Our primary aim is to compare these two predictive approaches and to assess their relative accuracy, assuming that the DGP is known and falls in the class of vector ARMA processes. Clearly, the ranking has to be made on the basis of some metric. In general, the accuracy of a forecasting model is measured by a forecast error loss function.<sup>3</sup> Usually, two competing forecasts are compared in terms of MSEs. The same criterion is used in the paper: aggregate versus disaggregate predictors are classified by sorting the corresponding MSEs.

More precisely, this paper aims to reconsider and extend the issue of comparing forecasting accuracy in ranking aggregate and disaggregate (i.e. based on subcomponents) predictors. First, we present the necessary and sufficient condition for the equality of MSEs of the aggregate and disaggregate processes, whenever the data generation process can be expressed as a vector moving average of order one (VMA(1)). Second, we show that the condition of equality of predictors in Lütkepohl (1984b, 1987, 2004) is only sufficient but not necessary for the equality<sup>4</sup> of MSEs. Third, we argue that the equality of forecasting accuracy can be achieved by using specific assumptions on the parameters of the VMA(1) structure. Finally, a Monte Carlo experiment and an empirical application are used to illustrate the main issues and our findings.

The remainder of the paper is structured as follows. Section 2 briefly defines the econometric framework, which is broadly based on Lütkepohl (1987). This latter contribution was extremely

---

<sup>2</sup>This condition is also reported in Lütkepohl (2004).

<sup>3</sup>Some popular loss functions are, for instance, the squared error loss function and the absolute error loss function.

<sup>4</sup>The equality of MSEs between two or more competing methods is the key requirement for equal forecasting accuracy.

important in advancing our understanding of issues related to aggregation and forecasting and inspired much of our work. Section 3 compares in terms of MSE the predictors built on the aggregate and disaggregate models for a bivariate VMA(1) process. Conditions on the micro parameters are given for the equality of MSEs. Proofs are gathered in the Appendix. Section 4 presents some simulation results based on a Monte Carlo experiment and gives an assessment of the forecasting accuracy of the competing predictors. Section 5 contains an empirical application dealing with the problem of forecasting the Italian monetary aggregate M1 (1948-1998). Section 6 concludes.

## 2 Two competing predictors to forecast contemporaneously aggregated vector ARMA processes

In this Section, we introduce the notation that will be used in the rest of the paper and present the forecasting methods under scrutiny.

Let us assume that the DGP is a  $k$ -dimensional stationary stochastic vector  $\mathbf{x}_t$  that can be expressed in Wold MA form as

$$\mathbf{x}_t = \sum_{i=0}^{\infty} \boldsymbol{\phi}_i \boldsymbol{\varepsilon}_{t-i} = \boldsymbol{\Phi}(L) \boldsymbol{\varepsilon}_t, \quad t \in \mathbb{Z}, \quad (1)$$

where  $|\boldsymbol{\phi}(z)| \neq 0$  for  $z \in \mathbb{C}$ ,  $|z| \leq 1$ , and  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{kt})'$  is a vector white noise innovations sequence with non-singular covariance matrix  $\boldsymbol{\Sigma}_\varepsilon$ , that is,  $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ ,  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \boldsymbol{\Sigma}_\varepsilon$  and  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \mathbf{0}$  for  $s \neq t$ . As usual  $\boldsymbol{\phi}_0 = \mathbf{I}_k$  and  $L$  is the backward shift operator, such that  $L\boldsymbol{\varepsilon}_t = \boldsymbol{\varepsilon}_{t-1}$ .

We focus on contemporaneous linear transformations of  $\mathbf{x}_t$  of the form

$$\mathbf{y}_t = \mathbf{F} \mathbf{x}_t, \quad (2)$$

where  $\mathbf{F}$  is an  $m \times k$  aggregation matrix of full rank  $m$ . Notice that if  $m = 1$ , the aggregate variable is a scalar. On the other hand, for  $m > 1$ , the aggregate variable is an  $m \times 1$  vector. In general, it has been shown that linear transformations of vector ARMA processes are again vector ARMA processes (Lütkepohl, 1984a).

Throughout the analysis, we suppose all the processes (including the covariances) are completely known, whereas in general the orders are unknown and the parameters have to be estimated.

The goal is to forecast the contemporaneously aggregated variable  $\mathbf{y}_t$  in (2). In the previous section we have explained that this is usually done by forecasting *ex-ante* the subcomponents (equation by equation) and summing *ex-post* the univariate predictions or by predicting directly the aggregate variable. Formally, the two competing approaches can be built:

1. By making univariate predictions of the individual components of  $\mathbf{x}_t$  in (1) and then by aggregating them into a single forecast for  $\mathbf{y}_t$  in (2). This is what we call *disaggregate approach*, that is to make univariate forecasts first, then aggregate them.



Denote the univariate Wold representations for the  $j = 1, \dots, k$  individual components of  $\mathbf{x}_t$  in (1) as

$$x_{jt} = \sum_{i=0}^{\infty} \theta_{ji} w_{j,t-i} = \theta_j(L) w_{jt}, \quad \theta_{j0} = 1 \quad (\forall j), \quad j = 1, \dots, k. \quad (3)$$

Also in this case, the parameters of the models for the individual components of  $\mathbf{x}_t$  are a function of the DGP's parameters in (1). This will become clearer in the next section. We can express the whole vector process in matrix form as

$$\begin{bmatrix} x_{1t} \\ \vdots \\ x_{kt} \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{\infty} \theta_{1i} L^i & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \sum_{i=0}^{\infty} \theta_{ki} L^i \end{bmatrix} \begin{bmatrix} w_{1t} \\ \vdots \\ w_{kt} \end{bmatrix} = \mathbf{\Theta}(L) \mathbf{w}_t,$$

where  $\mathbf{\Theta}(L) := \text{diag}[\theta_1(L), \dots, \theta_k(L)]$  and  $\mathbf{w}_t$  is a  $k \times 1$  vector. Each univariate component is forecast by using the optimal  $h$ -step ahead predictor  $x_{kt}^u(h)$ , i.e.

$$x_{jt}^u(h) = \sum_{i=0}^{\infty} \theta_{j,h+i} w_{j,t-i}, \quad j = 1, \dots, k, \quad (4)$$

which is the best linear predictor of  $x_{j,t+h}$  with information up to time  $t$  (see for instance Hamilton, 1994, p. 77). Stacking these  $j = 1, \dots, k$  forecasts in a vector, we can re-write (4) as

$$\mathbf{x}_t^u(h) = \sum_{i=0}^{\infty} \boldsymbol{\theta}_{h+i} \mathbf{w}_{t-i}. \quad (5)$$

Based on the univariate forecasts in (5), an  $h$ -step ahead predictor for  $\mathbf{y}_t$  in (2) may be obtained as

$$\mathbf{y}_t^u(h) = \mathbf{F} \mathbf{x}_t^u(h), \quad (6)$$

that is taking their sum or another kind of linear transformation induced by the  $\mathbf{F}$  aggregation matrix. We define  $\boldsymbol{\Sigma}_y^u(h)$  the forecast MSE matrix of  $\mathbf{y}_t^u(h)$  in (6),

$$\boldsymbol{\Sigma}_y^u(h) = \text{E} [(\mathbf{y}_{t+h} - \mathbf{y}_t^u(h))(\mathbf{y}_{t+h} - \mathbf{y}_t^u(h))'], \quad (7)$$

that is, the covariance matrix of the forecast error vector. See Lütkepohl (1987), p. 103, or Pino, Morettin and Mentz (1987), p. 304, for more details.

2. By forecasting the aggregate variable using directly the contemporaneously aggregated process in (2). This is the *aggregate approach* indicated earlier,<sup>5</sup> which corresponds to aggregate

---

<sup>5</sup>See Man (2004) for a similar approach in the context of temporal aggregation and forecasting.

first, then forecast on the basis of

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \boldsymbol{\psi}_i \mathbf{v}_{t-i} = \boldsymbol{\Psi}(L) \mathbf{v}_t, \quad \boldsymbol{\psi}_0 = \mathbf{I}_m, \quad (8)$$

where (8) is a multivariate Wold MA representation and  $\{\mathbf{v}_t\} \sim WN(\mathbf{0}, \boldsymbol{\Sigma}_v)$ , being  $\boldsymbol{\Sigma}_v$  a non-singular covariance matrix.

We know that the process in (8) is again a vector MA process of dimension  $m \times 1$  after aggregation. The VMA class is in fact closed with respect to linear transformation (Lütkepohl, 2007, Proposition 11.1, p. 435). Moreover, the parameters of the aggregate process  $\mathbf{y}_t$  can be recovered as a function of the DGP's parameters in (1).

This prediction approach requires to aggregate *ex-ante* the  $k$  individual components to form the aggregate process  $\mathbf{y}_t$ . This latter is forecast directly using the optimal  $h$ -step ahead predictor  $\mathbf{y}_t(h)$

$$\mathbf{y}_t(h) = \sum_{i=0}^{\infty} \boldsymbol{\psi}_{h+i} \mathbf{v}_{t-i}, \quad (9)$$

whose MSE is

$$\boldsymbol{\Sigma}_y(h) = \sum_{i=0}^{h-1} \boldsymbol{\psi}_i \boldsymbol{\Sigma}_v \boldsymbol{\psi}_i'. \quad (10)$$

See for instance Lütkepohl (2007), p. 434.

Heuristically, it is believed that the prediction built aggregating the forecasts of the individual components (*disaggregate approach*) is more efficient than the prediction obtained by forecasting directly the aggregate process (*aggregate approach*). For instance, quoting Marcellino, Stock and Watson (2003), p. 9, “From a theoretical perspective, pooling the country-specific forecasts should produce lower mean squared forecast errors than directly forecasting the aggregates, provided that the country-specific models are time invariant, that they are correctly specified, that the model parameters differ across countries, that there are no data irregularities, and that there are enough observations, etc. Whether these assumptions are useful approximations in practice, and thus whether pooled country forecasts or direct forecasts of the aggregates actually work better, is an empirical question.”

However, as pointed out by Wei and Abraham (1981) and by Lütkepohl (1987), there is no general unconditional inequality between the forecasts obtained using the two competing approaches. According to Lütkepohl (1987, 2004), in fact, the ranking of the predictors in (6) and (9) depends on the structure of the underlying DGP and on the aggregation matrix. The paper aims to cast light on this statement.

In the next section we explore and compare the forecasting performance of the aggregate and the disaggregate (based on the individual components) predictors. To keep things as simple

as possible we let  $m = 1$  in (2), that is we assume that the aggregate variable is a scalar. We remark that our paper focuses on the bivariate VMA(1) process since this represents the benchmark framework extensively employed by the whole of the aggregation literature (see Wei and Abraham, 1981, Lütkepohl, 1984b, 1984c, 1987, 2007). For a bivariate VMA(1), we provide the necessary and sufficient condition for the equality of one-step ahead MSEs. Furthermore, we show that assumptions on the parameters of the DGP can be made to guarantee that the condition of equal forecasting performance is satisfied.

### 3 Disaggregate and aggregate predictors for a bivariate vector MA process

We focus on the bivariate framework of a VMA(1). Thus, the DGP in (1) is

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1 + \phi_{11}L & \phi_{12}L \\ \phi_{21}L & 1 + \phi_{22}L \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \quad (11)$$

where  $\boldsymbol{\varepsilon}_t$  is a vector white noise innovation sequence, that is,  $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ ,  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$  and  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \mathbf{0}$  for  $s \neq t$ .

In what follows, we are going to express the parameters of the competing predictors as a function of the parameters of the DGP in (11). We remark that there is no loss of generality to assume that the variance is one across the innovations. This is done for the sake of simplicity. All the results in the sequel can be extended to the case of different variances (see Section 5 for further details).

#### 3.1 The parameters of the disaggregate predictor

Let us start to derive the parameters needed to implement the *disaggregate approach*, that is to model and forecast each subcomponent of  $\mathbf{x}_t$  separately, then to form the aggregate forecast. If we are interested in making univariate predictions for each of the two individual components of (11), we can re-parameterize the process in (11) as

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1 + \theta_1 L & 0 \\ 0 & 1 + \theta_2 L \end{bmatrix} \begin{bmatrix} w_{1t} \\ w_{2t} \end{bmatrix}, \quad (12)$$

where  $\mathbf{w}_t$  is a zero mean vector process with covariance matrix  $\boldsymbol{\Omega}$ . The  $\theta_1$  and  $\theta_2$  parameters in (12) can be recovered solving the system of equations:

$$\begin{aligned} \frac{1 + \phi_{11}^2 + \phi_{12}^2 + 2\rho\phi_{11}\phi_{12}}{2(\phi_{11} + \rho\phi_{12})} &= \frac{1 + \theta_1^2}{2\theta_1} \\ \frac{1 + \phi_{22}^2 + \phi_{21}^2 + 2\rho\phi_{22}\phi_{21}}{2(\phi_{22} + \rho\phi_{21})} &= \frac{1 + \theta_2^2}{2\theta_2}. \end{aligned} \quad (13)$$

Letting  $c_1 = \frac{1+\phi_{11}^2+\phi_{12}^2+2\rho\phi_{11}\phi_{12}}{2(\phi_{11}+\rho\phi_{12})}$  and  $c_2 = \frac{1+\phi_{22}^2+\phi_{21}^2+2\rho\phi_{22}\phi_{21}}{2(\phi_{22}+\rho\phi_{21})}$ , we can re-write (13) as two second degree equations:

$$\begin{aligned}\theta_1^2 - 2c_1\theta_1 + 1 &= 0 \\ \theta_2^2 - 2c_2\theta_2 + 1 &= 0\end{aligned}$$

Therefore we obtain  $\theta_1$  and  $\theta_2$

$$\begin{aligned}\theta_1 &= \frac{(1+\phi_{11}^2+\phi_{12}^2+2\rho\phi_{11}\phi_{12})-\sqrt{(1+\phi_{11}^4+\phi_{12}^4+4\rho^2\phi_{11}^2\phi_{12}^2-2\phi_{11}^2+2\phi_{12}^2+2\phi_{11}^2\phi_{12}^2-4\rho\phi_{11}\phi_{12}+4\rho\phi_{11}^3\phi_{12}+4\rho\phi_{11}\phi_{12}^3-4\rho^2\phi_{12}^2)}}{2(\phi_{11}+\rho\phi_{12})} \\ \theta_2 &= \frac{(1+\phi_{22}^2+\phi_{21}^2+2\rho\phi_{22}\phi_{21})-\sqrt{(1+\phi_{22}^4+\phi_{21}^4+4\rho^2\phi_{22}^2\phi_{21}^2-2\phi_{22}^2+2\phi_{21}^2+2\phi_{22}^2\phi_{21}^2-4\rho\phi_{22}\phi_{21}+4\rho\phi_{22}^3\phi_{21}+4\rho\phi_{22}\phi_{21}^3-4\rho^2\phi_{21}^2)}}{2(\phi_{22}+\rho\phi_{21})},\end{aligned}\tag{14}$$

picking the invertible solution with  $|\theta_1| < 1$  and  $|\theta_2| < 1$ .

Moreover, since  $E(w_{1t}^2) = \frac{\phi_{11}+\rho\phi_{12}}{\theta_1}$  and  $E(w_{2t}^2) = \frac{\phi_{22}+\rho\phi_{21}}{\theta_2}$ , the covariance between  $w_{1t}$  and  $w_{2t}$  may be easily calculated from

$$\begin{aligned}\text{cov}(w_{1t}, w_{2t}) &= E(w_{1t}w_{2t}) \\ &= \rho + \phi_{11}\phi_{21} + \rho\phi_{11}\phi_{22} - \rho\theta_2\phi_{11} + \rho\phi_{12}\phi_{21} + \phi_{12}\phi_{22} - \theta_2\phi_{12} - \theta_1\phi_{21} - \rho\theta_1\phi_{22} + \theta_1\theta_2 E(w_{1,t-1}w_{2,t-1}),\end{aligned}$$

and is equal to

$$\text{cov}(w_{1t}, w_{2t}) = \frac{(\phi_{11} - \theta_1)\phi_{21} + (\phi_{22} - \theta_2)\phi_{12} + (1 - \theta_2\phi_{11} - \theta_1\phi_{22} + \phi_{12}\phi_{21} + \phi_{11}\phi_{22})\rho}{1 - \theta_1\theta_2}.\tag{15}$$

As a consequence the  $\mathbf{\Omega}$  matrix is:

$$\left[ \begin{array}{cc} \frac{\phi_{11}+\rho\phi_{12}}{\theta_1} & \frac{(\phi_{11}-\theta_1)\phi_{21}+(\phi_{22}-\theta_2)\phi_{12}+(1-\theta_2\phi_{11}-\theta_1\phi_{22}+\phi_{12}\phi_{21}+\phi_{11}\phi_{22})\rho}{1-\theta_1\theta_2} \\ \frac{(\phi_{11}-\theta_1)\phi_{21}+(\phi_{22}-\theta_2)\phi_{12}+(1-\theta_2\phi_{11}-\theta_1\phi_{22}+\phi_{12}\phi_{21}+\phi_{11}\phi_{22})\rho}{1-\theta_1\theta_2} & \frac{\phi_{22}+\rho\phi_{21}}{\theta_2} \end{array} \right]$$

Similar results for the case of a multivariate VMA(1) process with diagonal covariance matrix  $\mathbf{\Sigma}_\varepsilon$  can be found in Sbrana (2008b).

Within this framework, suppose that  $\mathbf{F}$  is a  $(1 \times 2)$  vector of ones.<sup>6</sup> To forecast a linear transformation of the  $\mathbf{x}_t$  process in (12), i.e.

$$y_t = \mathbf{F}\mathbf{x}_t$$

a one-step ahead predictor of  $y_t$  based on univariate components of  $\mathbf{x}_t$  has MSE in (7) equal to

$$\begin{aligned}\Sigma_y^u(1) &= \frac{\phi_{11} + \rho\phi_{12}}{\theta_1} + \frac{\phi_{22} + \rho\phi_{21}}{\theta_2} \\ &+ 2 \left( \frac{(\phi_{11} - \theta_1)\phi_{21} + (\phi_{22} - \theta_2)\phi_{12} + (1 - \theta_2\phi_{11} - \theta_1\phi_{22} + \phi_{12}\phi_{21} + \phi_{11}\phi_{22})\rho}{1 - \theta_1\theta_2} \right),\end{aligned}\tag{16}$$

where  $\theta_1$  and  $\theta_2$  are given in (14).

---

<sup>6</sup>This assumption will be relaxed in Section 5, where results relative to the weighted aggregation, rather than the simple sum, will be provided.

Finally, for multi-step ahead forecasts the MSE is

$$\begin{aligned}\Sigma_y^u(\forall h > 1) &= (1 + \theta_1^2)E(w_{1t}^2) + (1 + \theta_2^2)E(w_{2t}^2) + 2(1 + \theta_1\theta_2)E(w_{1t}w_{2t}) \\ &+ 2\theta_1E(w_{2t}w_{1,t-1}) + 2\theta_2E(w_{1t}w_{2,t-1}).\end{aligned}\quad (17)$$

We already provide the expressions for  $\theta_1$ ,  $\theta_2$ ,  $E(w_{1t}w_{2t})$  in (14) and (15). Therefore, only  $E(w_{1t}w_{2,t-1})$  and  $E(w_{2t}w_{1,t-1})$  are needed to calculate (17). These two cross-moments are not null and are equal to

$$\begin{aligned}E(w_{1t}w_{2,t-1}) &= \frac{(1 - \rho\theta_1\phi_{21} - \theta_1\phi_{22})\phi_{12} - (\phi_{11} - \theta_1)\theta_1\phi_{21} + (\phi_{11} - \theta_1 - \theta_1\phi_{22}\phi_{11} + \theta_1^2\phi_{22})\rho}{1 - \theta_1\theta_2} \\ E(w_{2t}w_{1,t-1}) &= \frac{(1 - \rho\theta_2\phi_{12} - \theta_2\phi_{11})\phi_{21} - (\phi_{22} - \theta_2)\theta_2\phi_{12} + (\phi_{22} - \theta_2 - \theta_2\phi_{11}\phi_{22} + \theta_2^2\phi_{11})\rho}{1 - \theta_1\theta_2}.\end{aligned}$$

As a consequence, as pointed out by Lütkepohl (1987), p. 102,  $\mathbf{w}_t$  is not a two-dimensional vector white noise process.

### 3.2 The parameters of the aggregate predictor

Let us now consider the parameters of the aggregate predictor, which are needed for the *aggregate approach*, that is to form the contemporaneous aggregation, then to model and forecast the aggregate process. It may be proven that summing up across  $j = 1, 2, \dots, k$  moving average processes of order  $q_j$  leads to an MA( $q^*$ ) model, where  $q^* \leq \max(q_j)$ . Therefore, the contemporaneous aggregation of  $j = 1, 2$  moving average processes of order one is again a moving average of order one<sup>7</sup> as

$$y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = (1 + \psi L)v_t, \quad (18)$$

with moving average parameter  $\psi$  and innovations  $v_t \sim WN(0, \sigma_v^2)$ .

It is possible to obtain the parameters of the contemporaneously aggregated model as a direct function of the parameters of the VMA(1) in (11). More precisely, the  $\psi$  in (18) “implied” by the model in (11) can be recovered solving the system

$$\begin{aligned}E(y_t^2) &= (1 + \psi^2)\sigma_v^2 \\ E(y_t y_{t-1}) &= \psi\sigma_v^2,\end{aligned}\quad (19)$$

which reduces to the second degree equation  $\psi^2 - 2\delta\psi + 1 = 0$ , with  $\delta := \frac{E(y_t^2)}{2E(y_t y_{t-1})}$ . After some algebra, this yields

$$\psi = \frac{2 + \alpha_1^2 + \alpha_2^2 + 2(1 + \alpha_1\alpha_2)\rho - \sqrt{((\alpha_1 - 1)^2 + (\alpha_2 - 1)^2 + 2(\alpha_1 - 1)(\alpha_2 - 1)\rho)((\alpha_1 + 1)^2 + (\alpha_2 + 1)^2 + 2(\alpha_1 + 1)(\alpha_2 + 1)\rho)}}{2(\alpha_1 + \alpha_2 + \rho(\alpha_1 + \alpha_2))}, \quad (20)$$

<sup>7</sup>We refer among others to Hamilton (1994), pp. 106-107, for proof of this result.

where  $\alpha_1 := (\phi_{11} + \phi_{21})$  and  $\alpha_2 := (\phi_{12} + \phi_{22})$ .

Moreover, the variance of the innovations in (18) is

$$\sigma_v^2 = \frac{\alpha_1 + \alpha_2 + \rho(\alpha_1 + \alpha_2)}{\psi}, \quad (21)$$

where  $\psi$  is given in (20). In this specific case<sup>8</sup> the  $\Sigma_y(1)$  matrix in (10) reduces to scalar and coincides with  $\sigma_v^2$ . Therefore a one-step ahead predictor of  $y_t$  based on the aggregate process in (18) has MSE equal to  $\sigma_v^2$ , whereas as well known  $\Sigma_y(\forall h > 1) = (1 + \psi^2)\sigma_v^2$ .

It can also be seen that whenever  $\alpha_1 = -\alpha_2$  the aggregate process collapses to a white noise with variance equal to  $\sigma_v^2 = 2(1 + \rho) + 2(1 - \rho)\alpha_2^2$ .

Finally, to show that  $v_t$  innovations sequence in (18) is white noise, it is enough to note that

$$E(v_t, v_{t-1}) = (\alpha_1 + \alpha_2) + \rho(\alpha_1 + \alpha_2) - \psi E(v_{t-1}^2) = 0,$$

on the basis of (20) and (21). In addition,  $E(v_t, v_{t-j}) = 0, \forall j > 1$ .

### 3.3 A comparison of the predictors via forecast mean squared errors

We can now try to rank the competing predictors. Notice that so far we have expressed the corresponding MSEs as a function of the parameters of the DGP in (11). The structure and the parameter values of the latter determine the accuracy of the forecasting procedures built on individual components and on the contemporaneously aggregated process.

In particular, when  $\phi_{21} = \phi_{12} = 0$ , it is known that  $\Sigma_y(1) = \sigma_v^2 \geq \Sigma_y^u(1)$ . In this case, the approach based on individual components is the best forecasting procedure, and the corresponding predictor is equal to the optimal predictor built employing the whole multivariate DGP, as shown by Lütkepohl (1987) in Corollary 4.1.1, case iii, p. 107. This, of course, makes sense: intuitively, if the basic univariate time series are independent, it is preferable to forecast the individual components separately and then form a forecast for the aggregate variable (Pino, Morettin and Mentz, 1987). On the contrary, the more dependent the individual components, the worse the ability of the disaggregate predictor to forecast the aggregate process.

Moreover, as already found by previous literature, when  $(\phi_{11} + \phi_{21}) = (\phi_{12} + \phi_{22})$ , the aggregate approach outperforms the disaggregate one, since it becomes the optimal forecasting procedure. That is, the corresponding predictor is equal to the predictor based on the whole multivariate DGP (see Corollary 4.1.1, case i, p. 107 in Lütkepohl, 1987), which is the optimal one. As a result, the more  $(\phi_{11} + \phi_{21}) \neq (\phi_{12} + \phi_{22})$ , the worse the forecasting ability of the aggregate predictor.

This is evident in Figure A1, which displays the contour and three-dimensional plots of  $\sigma_v^2$  in (21) as a function of the parameters of the bivariate VMA(1) in (11):  $\alpha_1 = (\phi_{11} + \phi_{21})$  and  $\alpha_2 = (\phi_{12} + \phi_{22})$ . In both panels, the extra-diagonal element of the covariance matrix of the innovations  $\rho$  is set equal to 0.3.

---

<sup>8</sup>Note that (20) and (21) hold under the assumption that  $\sigma_{\varepsilon_1}^2 = \sigma_{\varepsilon_2}^2 = 1$  in (11).

[FIGURE A1 ABOUT HERE]

The three-dimensional plot of  $\sigma_v^2$  is clearly symmetric across the 45 degree line on the  $(\alpha_1, \alpha_2)$  cartesian plane, and has its minima where  $\alpha_1 = \alpha_2$ , on the same plane. As we move away from the 45 degree line,  $\sigma_v^2$  increases. This is expected and broadly consistent with previous theoretical results.

Note that the analytical expression of  $\sigma_v^2$  in (21), and its graphical representation, represent an improvement with respect to previous contributions in the aggregation literature. Intuitively, it was possible to figure out that  $\sigma_v^2$  increases as the difference between the  $\alpha$  parameters widens. However, an exact expression had never been formalized (and visualized) before.

To summarize, in the bivariate framework of a VMA(1), the forecasting performance of the disaggregate process is based on the level of dependence of the individual components, while that of the aggregate depends on the distance between  $(\phi_{11} + \phi_{21})$  and  $(\phi_{12} + \phi_{22})$ .

So far, we have provided some intuition behind the performance of a forecasting approach relative to the other. Based on (21) and (16), it is obvious how to rank the MSEs of the competitive predictors. In what follows, we provide a condition that guarantees equal predictive efficiency.

**Proposition 1** *Given the vector process in (11) and the MSEs (21) and (16), the necessary and sufficient condition for the equality of MSEs of the aggregate process (aggregate approach) and of the individual components (disaggregate approach) is*

$$\frac{\alpha_1 + \alpha_2 + \rho(\alpha_1 + \alpha_2)}{2 + \alpha_1^2 + \alpha_2^2 + 2(1 + \alpha_1 \alpha_2)\rho - \sqrt{((\alpha_1 - 1)^2 + (\alpha_2 - 1)^2 + 2(\alpha_1 - 1)(\alpha_2 - 1)\rho)((\alpha_1 + 1)^2 + (\alpha_2 + 1)^2 + 2(\alpha_1 + 1)(\alpha_2 + 1)\rho}} = \frac{\phi_{11} + \rho\phi_{12}}{\theta_1} + \frac{\phi_{22} + \rho\phi_{21}}{\theta_2} + 2 \left( \frac{(\phi_{11} - \theta_1)\phi_{21} + (\phi_{22} - \theta_2)\phi_{12} + (1 - \theta_2\phi_{11} - \theta_1\phi_{22} + \phi_{12}\phi_{21} + \phi_{11}\phi_{22})\rho}{1 - \theta_1\theta_2} \right), \quad (22)$$

with  $\alpha_1 := (\phi_{11} + \phi_{21})$ ,  $\alpha_2 := (\phi_{12} + \phi_{22})$  and  $\theta_1, \theta_2$  as in (14).

We stress the relevance of Proposition 1, which gives the necessary and sufficient condition for equal forecasting accuracy delivered by the aggregate and disaggregate predictors.<sup>9</sup> First, (22) is not only a theoretical condition but also a computationally easy rule to rank the predictors under scrutiny, simply on the basis of the structure of the underlying DGP in (11). Second, and related to this, all the quantities in (22) are expressed as a function of the DGP's parameters. This is important to underline since (22) provides a clear link between the DGP, the aggregate and disaggregate predictors and their precision. To the best of our knowledge nowhere in the

<sup>9</sup>In the case under study, it is well known that for multi-step ahead predictions ( $h > 1$ ) the two procedures deliver the same forecasts.

literature is there a similarly straightforward way to determine the accuracy of the aggregate approach relative to the disaggregate one.

Furthermore, we remark that in (22) the forecast MSE is employed as a measure of predictive efficiency, as done in most forecasting applications. As already explained by Pino, Morettin and Mentz (1987), the efficiency of the aggregate predictor relative to the disaggregate can be measured either by comparing the corresponding forecast errors or by comparing their forecast MSEs. Yet, it is worth pointing out that the evaluation of forecasting accuracy has almost exclusively been conducted under the assumption of mean squared error loss functions.<sup>10</sup>

To interpret Proposition 1, which is a highly nonlinear function of five parameters, it can be useful to fix some of them, moving the others, and drawing the equality condition. This latter can be viewed as an implicit function of the form:  $\sigma_v^2 - \Sigma_y^u(1) = 0$ .

[FIGURE A2 ABOUT HERE]

Figure A2 displays three-dimensional and contour plots of  $\sigma_v^2 - \Sigma_y^u(1)$ , as a function of the parameters of the bivariate VMA(1) in (11), i.e.,  $\phi_{11}, \phi_{21}, \phi_{12}, \phi_{22}$  and  $\rho$ . In the plots, two of the parameters (i.e.,  $\phi_{11}, \phi_{22}$ ) vary while the other three parameters (i.e.,  $\rho, \phi_{12}, \phi_{21}$ ) are kept fixed and set equal to the values reported below each panel.

Some interesting conclusions may be easily drawn from Figure A2. First, from panels (a) and (b), we remark that as  $\phi_{12}$  and  $\phi_{21}$  move toward the point (0, 0) in absolute value, the function values increase, i.e., the performance of the disaggregate predictor is improved. This tendency is visible by looking at the top panels from the right to the left. This is also evident by looking at panels (c) and (d), where the level curves are depicted. Second, from a careful look at the contours sketched in panels (c), we note that for the chosen combination  $\rho = 0.3, \phi_{12} = 0, \phi_{21} = 0$ , the disaggregate outperforms the aggregate predictor across all the displayed region. In particular, the difference between the MSEs is almost zero close to the  $\alpha_1 = \alpha_2$  line, and increases steadily as we move away from the 45 degree line. In panel (d), for the chosen combination  $\rho = 0.3, \phi_{12} = 0.8, \phi_{21} = 0.8$ , the aggregate predictor outperforms the disaggregate in a rather wide region, close to the  $\alpha_1 = \alpha_2$  line and on the top left of the graph, where the level curves are negative. As we move away from the 45 degree line toward the bottom right of the graph, the contours become positive, and the ranking changes, i.e., the disaggregate outperforms the aggregate.

### 3.4 An insight on the conditions for equal forecasting efficiency

What has already been given in the aggregation literature (Lütkepohl 1984b, 1987, 2004) is a necessary and sufficient condition for the equality of the two predictors in (6) and (9). This author states the following *necessary and sufficient* condition for the *equality of h-step ahead predictors*

---

<sup>10</sup>If this function of the forecast error is employed, the predictor with the smallest MSE has the best forecasting record and is the one preferred. For more explanations, see Granger and Newbold (1986).



based on the individual components and on the aggregate process (Corollary 4.1.1, case ii, p. 107 in Lütkepohl, 1987):

$$\mathbf{y}_t^u(h) = \mathbf{y}_t(h) \iff \mathbf{F}\Theta(L) = \Psi(L)\mathbf{F}. \quad (23)$$

The proof of (23) is provided by Lütkepohl (1987), p. 106.

It is worth discussing the meaning of (23), which is implicitly a condition on the equality of forecast errors of the prediction methods in (6) and (9). In general, the forecast errors are the same whenever the linear aggregation of the parameters of the individual components is equal to the parameter(s) of the aggregate process. In our bivariate framework, for instance, the condition in (23) reads as:

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 + \theta_1 L & 0 \\ 0 & 1 + \theta_2 L \end{bmatrix} = \begin{bmatrix} 1 + \psi L \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix}. \quad (24)$$

It is clear that (24) is in general difficult to meet, since in our bivariate framework it is verified if and only if  $\theta_1 = \theta_2 = \psi$ .

To get further insight into (24), it is possible to reaffirm this condition as a function of the DGP's parameters, that is,  $\phi_{11}$ ,  $\phi_{12}$ ,  $\phi_{21}$ ,  $\phi_{22}$  and  $\rho$ . This is done in the following proposition.

**Proposition 2** *For the VMA(1) system in (11), for any  $\phi_{11}$ ,  $\phi_{22}$  and assuming  $\rho \neq 0$ , the conditions*

$$\begin{aligned} \phi_{21} &= (\phi_{11} - \phi_{22}) \left( \frac{1}{2\rho} \right) \\ \phi_{12} &= -\phi_{21} \end{aligned} \quad (25)$$

*are sufficient for the equality of MSEs of the aggregate process (aggregate approach) and of the individual components (disaggregate approach).*

**Proof.** If (25) holds, it can be shown that

$$\theta_1 = \theta_2 = \psi = \frac{(\phi_{11} - \phi_{22})^2 + \rho^2 \left( 4(1 + \phi_{11}\phi_{22}) - \sqrt{\frac{((\phi_{11} - \phi_{22})^2 + 4\rho^2(\phi_{11} - 1)(\phi_{22} - 1))((\phi_{11} - \phi_{22})^2 + 4\rho^2(\phi_{11} + 1)(\phi_{22} + 1))}{\rho^4}} \right)}{4\rho^2(\phi_{11} + \phi_{22})}.$$

Since  $\theta_1 = \theta_2 = \psi$ , the equality of the MSEs of the methods based on the aggregate and disaggregate process is guaranteed.  $\square$

However, although (23) is necessary<sup>11</sup> and sufficient for the equality of the predictors, *it is sufficient but not necessary for the equality of the corresponding MSEs*. This can be easily shown with a counter-example in the bivariate framework of a VMA(1). Consider the following parameter values in (11):  $\phi_{11} = .6$ ,  $\phi_{12} = -.3$ ,  $\phi_{21} = .2$ ,  $\phi_{22} = .4$ ,  $\rho = .3335$ . As already illustrated, we can derive  $\theta_1 = .4531$ ,  $\theta_2 = .4466$  and  $\psi = .4184$ . Hence  $\theta_1 \neq \theta_2 \neq \psi$  and (23) is not satisfied. On the other hand,  $\Sigma_y^u(1) = \Sigma_y(1) = 2.8681$ , and therefore the two competing methods have the same forecasting accuracy. The bottom line is that the equality of MSEs can be achieved under several circumstances, regardless of the values of  $\theta_1$ ,  $\theta_2$  and  $\psi$ .

Here below we provide a sufficient condition for the equality of MSEs that holds in the bivariate framework of the VMA process in (11). However, as it will become clear shortly, this condition does not satisfy (23). This is a key result of our work and may be considered a novelty.

**Proposition 3** *Let us focus on the VMA(1) in (11). For any  $\phi_{11}$ ,  $\phi_{22}$  and assuming  $\rho \neq 0$ , the following*

$$\begin{aligned}\phi_{21} &= (\phi_{11} - \phi_{22}) \left( \frac{1}{2} + \rho \right) \\ \phi_{12} &= \frac{\phi_{11} - \phi_{22}}{2}\end{aligned}\tag{26}$$

*are sufficient conditions for the equality of MSEs of the aggregate process (aggregate approach) and of the individual components (disaggregate approach).*

We defer the proof of Proposition 3 to the Appendix. The reader can check that when (26) holds condition (23) is not met, since  $\theta_1 \neq \theta_2$ .

It is interesting to note that all the illustrations and numerical examples proposed by the aggregation literature focus on  $\rho$  equal to zero (e.g., Wei and Abraham, 1981, Lütkepohl, 1984c, 1987, 2007, Hendry and Hubrich, 2007). To our knowledge, nowhere in the literature is the case  $\rho \neq 0$  discussed and analyzed. Yet, this latter case deserves particular attention due to its great practical importance in empirical analysis since, very often, the individual components series are correlated.

Finally, assuming  $\rho = 0$  and hence focusing on a framework in which the covariance matrix of the innovations is diagonal, we introduce a further condition for equal forecasting efficiency that does not satisfy (23). This means that, when the errors are uncorrelated, it is possible to find a sufficient condition for the equality of MSEs which does not respect (23), since also in this simplified framework  $\theta_1 \neq \theta_2$  (we refer to the proof in the Appendix for further details).

---

<sup>11</sup>In what follows we briefly summarize the steps of the proof in Lütkepohl (1987) to show necessity of (23). Let us focus on the bivariate framework of the VMA in (11). To show that (24) is a necessary condition for the equality of one-step ahead predictors, assume that  $y_t^u(1) = y_t(1)$  holds. Remind that  $y_{t+1} - y_t(1) = v_{t+1}$  and  $y_{t+1} - y_t^u(1) = \mathbf{F}\mathbf{w}_{t+1}$  by construction. Hence  $y_t^u(1) = y_t(1) \Rightarrow \mathbf{F}\mathbf{w}_{t+1} = v_{t+1}$ . Therefore

$$\mathbf{F}\Theta(L)\mathbf{w}_t := y_t := \psi(L)v_t = \psi(L)\mathbf{F}\mathbf{w}_t$$

and thereby  $\mathbf{F}\Theta(L) = \psi(L)\mathbf{F}$ , which reads as  $\theta_1 = \theta_2 = \psi$  in our framework.

**Proposition 4** Consider the VMA(1) in (11) and assume that  $\rho = 0$ . If we impose  $\phi_{12} = \phi_{21}$ , then

$$\phi_{12} = \pm \frac{\phi_{22} - \phi_{11}}{2} \quad (27)$$

is a sufficient condition for the equality of MSEs of the aggregate process (aggregate approach) and of the individual components (disaggregate approach).

The linear combination in (27) guarantees the equality of forecasting performance of the competitive predictors in the framework of a bivariate VMA(1) process, despite that  $\theta_1 \neq \theta_2$ . The proof of Proposition 4 is given in the Appendix.

It is worth focusing on the relevance of these results for the applied research. It is well known that empirical forecasting accuracy is mainly based on the comparison of mean squared errors of competitive models and not on the equality of predictors (being a particularly strong condition). Therefore, our analysis has some direct consequences on the empirical debate on the use of aggregate and disaggregate forecasts. In fact, it is not possible to establish *a priori* which is the best forecasting model as claimed by some authors, since both the aggregate and the disaggregate (individual components) predictors are sub-optimal procedures if compared with the optimal procedure, i.e. aggregating the forecasts based on the original DGP in (1). In addition, Propositions 3 and 4 provide conditions for the equality of MSEs. They both reinforce the fact that (23) is sufficient but not necessary for equal forecasting efficiency. Lastly, as we have shown, their respective mean squared errors depend on the structure of the DGP. Therefore, in general, the two forecasting procedures need to be evaluated in each specific empirical case.

## 4 A simulation study

This section focuses on the main results from a Monte Carlo simulation. We adopt the framework already suggested by Lütkepohl (1984c, 1987) and more recently by Hendry and Hubrich (2007), taking into account the potential problems of model misspecification and estimation uncertainty linked to small sample size.

More specifically, a bivariate VMA(1) is generated as follows

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 + \phi_{11}L & \phi_{12}L \\ \phi_{21}L & 1 + \phi_{22}L \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \quad t = 1, 2, \dots, T \quad (28)$$

with

$$\varepsilon_t \sim i.i.d N \left( \mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).$$

Note that in (28), we introduce a positive contemporaneous covariance between the innovations, that is,  $\rho \neq 0$ . We remark that this is a novelty with respect to the Monte Carlo simulations presented in Lütkepohl (1984c, 1987) and Hendry and Hubrich (2007).

For each exercise, the number of replications is 10,000. Moreover, we assume that the DGP is unknown. Only autoregressive processes are used to fit and forecast. In this way, we take into account possible model misspecification. As in Lütkepohl (1987c), we employ AR(p) processes with  $p = 1, 2, \dots, 6$ . The standard information criteria are applied for model selection (in particular, the Akaike Information Criterion, AIC, and the Schwartz Information Criterion, BIC).

The idea underlying the design of the experiment is to compare the out-of sample MSEs of two different competitive models: the disaggregate process (individual components) and the aggregate process. The MSE is used as a metric for forecast accuracy. The structure of the parameters is the only feature that makes our analysis differ from the previously mentioned Monte Carlo simulations. That is, we assume four different structures of parameters:

- DGP 1 :  $\phi_{11} = 0.7$  ;  $\phi_{12} = 0$  ;  $\phi_{21} = 0$  ;  $\phi_{22} = -0.4$ ;  $\rho = 0.3$ ;
- DGP 2 :  $\phi_{11} = 0.7$  ;  $\phi_{12} = 0.2$  ;  $\phi_{21} = 0.32$  ;  $\phi_{22} = 0.3$ ;  $\rho = 0.3$ ;
- DGP 3 :  $\phi_{11} = 0.1$  ;  $\phi_{12} = 0.8$  ;  $\phi_{21} = 0.8$  ;  $\phi_{22} = 0.1$ ;  $\rho = 0.3$ ;

All the processes are invertible. The DGP 1 represents the theoretical case when the disaggregate outperforms the aggregate process. On the other hand, the DGP 3 is the case when the aggregate process performs better than the disaggregate process. DGP 2 is the object of interest of this paper, since it satisfies condition (26). In fact, in this case, we have shown that the aggregate and the disaggregate processes have exactly the same one-step ahead forecasting performance in terms of MSE. For all the DGPs, the number of observations used to estimate the model in-sample is  $T = 30, 50, 100, 200, 500$ . Five observations are kept for out-of-sample evaluation.

[TABLE A1 ABOUT HERE]

Table A1 reports the Monte Carlo results using the three different DGPs: each cell contains the ratio of the aggregate MSE relative to the disaggregate one; values greater than one, for instance, indicate that the aggregate MSE is larger than the disaggregate one. With DGP 1, results are clearly in favor of the disaggregate process for one-step ahead forecasts. This is true using both the Akaike and Schwartz information criteria. In general, it can be seen that the forecasts of the two competitive predictors get closer when T increases, and for five-steps ahead.

Focusing on DGP 3, on the other hand, we face the opposite situation in which the aggregate predictor outperforms the disaggregate: this is particularly evident for one-step ahead forecasts and for any T, whereas for five-steps ahead, the differences vanish. We stress that no specific process outperforms its competitors for five-steps ahead forecasts.

Looking at DGP 2, for which the condition of equality of predictors (26) holds, the disaggregate tends to perform slightly better than the aggregate in very small samples, when  $T = 30$ , but not as much as observed for DGP 1. On the other hand, when  $T \geq 50$ , the aggregate and the disaggregate have the same forecasting performance. In general, when the number of observations increases,

the MSEs of the aggregate and disaggregate predictors become almost identical. This is true for one-step ahead and for five-steps ahead forecasts. In summary, the differences between MSEs are very small, especially in large samples, where estimation uncertainty is reduced.

Overall, from this Monte Carlo experiment, we can conclude that the simulation results confirm our theoretical findings and shed further light on the  $\rho$  parameter's influence on the accuracy of the competing predictors. In particular, Table DGP 1 and DGP 3 represent two opposite frameworks in which one forecasting method clearly outperforms its competitor. Moreover, Table A1 shows that the condition of equal forecasting performance in terms of MSE in (26) is validated by simulations.

## 5 An empirical application: forecasting M1 in Italy in the pre-EMU period

In this section, we present an empirical application that involves the problem of forecasting the Italian monetary aggregate M1 on the basis of annual time series ranging from 1948 to 1998, prior to the creation of the European Economic and Monetary Union (EMU). This is of course an issue of major interest for researchers and practitioners.

According to the definition given by the Eurosystem, M1 is a narrow monetary aggregate that comprises two disaggregate components: overnight deposits ( $OV_t$ ) and currency in circulation ( $CC_t$ ) issued by the monetary financial institutions (MFIs) sector and by entities belonging to the central government. The same definition is used in this empirical application, that is, M1 is defined as

$$M1_t = OV_t + CC_t, \quad t = 1948, \dots, 1998. \quad (29)$$

The series in (29) was reconstructed by the Bank of Italy following the Eurosystem's definition (with minor differences concerning mainly the perimeter of the variables and the money-holding sector). Notice that the data for M1, overnight deposits and currency in circulation are not perfectly coincident with the series calculated since the first years of the sixties up to 1974, of which there is a trace in some historical publications (Bank of Italy's Annual Report, Economic Bulletin). Up to 1975, indeed, the M1 aggregate included a part of Treasury bonds (BOTs) and deposits that credit institutions had to hold to meet the obligatory reserve requirements. Since 1975, with the cash payment of the obligatory reserve by credit institutions, the definition of M1 used in this empirical application coincides with the definition employed in other studies and publications.

[FIGURE A3 ABOUT HERE]

The three series M1, overnight deposits and currency in circulation for the pre-EMU period are shown in levels in Figure A3. Data are expressed in billions of lire.

[TABLE A2 ABOUT HERE]

Table A2 presents some descriptive statistics about overnight deposits and currency in circulation (expressed in first differences of the logged series). As is evident, both series have non-zero means and are positively skewed. Moreover, overnight deposits have a standard deviation which is roughly two times the standard deviation of currency in circulation.

As an illustrative example, we fit the bivariate VMA(1) in (11) to (demeaned) overnight deposits and currency in circulation:

$$\mathbf{x}_t = \begin{bmatrix} \Delta \log(OV_t) \\ \Delta \log(CC_t) \end{bmatrix} = \begin{bmatrix} 1 + \phi_{11}L & \phi_{12}L \\ \phi_{21}L & 1 + \phi_{22}L \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}. \quad (30)$$

To fit the data to the model in (30), we generalize the DGP in (11) by assuming that  $\boldsymbol{\varepsilon}_t$  is a vector white noise innovation sequence with  $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ ,  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}$  and  $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \mathbf{0}$  for  $s \neq t$ . That is, innovations are heteroskedastic and cross-correlated. This is the most general framework we can set up. With these assumptions, the two time series can be adequately represented by the bivariate VMA(1) in (30). The resulting model fits the data reasonably well.<sup>12</sup>

Estimation is conducted by maximum-likelihood. Four different samples are employed, recursively: 1948-1995, 1948-1996, 1948-1997 and 1948-1998. The years 1996-1998 are used as forecasting period and for forecasting evaluation. We get four sets of estimated parameters for each sample:  $\hat{\phi}_{11}$ ,  $\hat{\phi}_{12}$ ,  $\hat{\phi}_{21}$ ,  $\hat{\phi}_{22}$ , together with the variance of the residuals  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$ , and the residuals covariance,  $\hat{\rho}$ . Estimation results are available in Table A3, for each sample used.<sup>13</sup>

[TABLE A3 ABOUT HERE]

Suppose we are interested in predicting the percentage changes of the aggregate variable M1 from period to period. Expressed differently, the aim is to forecast a linear transformation of the process in (30), i.e.

$$y_t = \mathbf{F} \mathbf{x}_t.$$

To do this, as already detailed, forecasts can be obtained using an aggregate or a disaggregate predictor, built by forecasting overnight deposits and currency in circulation separately and aggregating the forecasts ex-post.

Note, however, that the model in (30) is estimated on log transformed variables. This is rather usual in applied econometrics and it is mainly done to stabilize the variance of the time series before modeling. Yet, all the results so far discussed concerning the ranking of different predictors for aggregated variables refer to linear contemporaneous aggregation. Whether or not these results apply to nonlinear aggregation schemes is not easy to say at this stage. This is a relevant caveat

<sup>12</sup>Additional results are available from the authors upon request.

<sup>13</sup>Note that, for each sample used, the estimated parameters give invertible bivariate MA(1) models.

to be kept in mind.<sup>14</sup> To our knowledge, the topic of forecasting nonlinearly contemporaneously aggregated variables is still unexplored and, as recognized by Lütkepohl and Xu (2009), is a very promising avenue for future research.

As argued by Wesche (1997), Fagan and Henry (1998), Dedola, Gaiotti and Silipo (2001) and Sbrana (2008a), the following approximate relation holds between the logarithm of the aggregate series ( $M1$ ) and the sum of the logarithms of its individual components, properly weighted,

$$\Delta \log(M1_t) \approx \eta_1 \Delta \log(OV_t) + \eta_2 \Delta \log(CC_t), \quad (31)$$

where  $\eta_{1t} = \frac{OV_t}{M1_t}$ ,  $\eta_{2t} = \frac{CC_t}{M1_t}$ ,  $\forall t$ . In addition,  $\eta_1 = \frac{\sum_{t=1948}^{1998} \eta_{1t}}{50}$  and  $\eta_2 = \frac{\sum_{t=1948}^{1998} \eta_{2t}}{50}$  are the constant average shares of overnight deposits and currency in circulation on  $M1$ . Figure A4 displays the RHS and the LHS of (31). As it can be clearly seen, the approximation in (31) is not so rough, since the two lines almost overlap across the sample, with two exceptions in 1948 and 1951 (where the difference is  $10^{-2}$ ).<sup>15</sup>

[FIGURE A4 ABOUT HERE]

Note that, by construction,  $\eta_1 + \eta_2 = 1$ . Therefore, the aggregation vector that operates on the RHS of (31) is  $\mathbf{F} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$ .

First, we focus on the parameters of the disaggregate predictor. Consider the process in (30), that can be re-parameterized as in (12). Due to the covariance structure of the innovations in (30), that allows for heteroskedasticity, we get these expressions for the moving average parameters  $\theta_1$  and  $\theta_2$ , i.e. a generalized version of (14):

$$\begin{aligned} \theta_1 &= \frac{(\sigma_1^2 + \phi_{11}^2 \sigma_1^2 + \phi_{12}^2 \sigma_2^2 + 2\rho\phi_{11}\phi_{12})}{2(\phi_{11}\sigma_1^2 + \rho\phi_{12})} \\ &\quad - \frac{\sqrt{(\sigma_1^4 + \phi_{11}^4 \sigma_1^4 + \phi_{12}^4 \sigma_2^4 + 4\rho^2 \phi_{11}^2 \phi_{12}^2 - 2\phi_{11}^2 \sigma_1^4 + 2\phi_{12}^2 \sigma_1^2 \sigma_2^2 + 2\phi_{11}^2 \phi_{12}^2 \sigma_1^2 \sigma_2^2 - 4\rho\phi_{11}\phi_{12}\sigma_1^2 + 4\rho\phi_{11}^3 \phi_{12}\sigma_1^2 + 4\rho\phi_{11}\phi_{12}^3 \sigma_2^2 - 4\rho^2 \phi_{12}^2)}{2(\phi_{11}\sigma_1^2 + \rho\phi_{12})} \\ \theta_2 &= \frac{(\sigma_2^2 + \phi_{22}^2 \sigma_2^2 + \phi_{21}^2 \sigma_1^2 + 2\rho\phi_{22}\phi_{21})}{2(\phi_{22}\sigma_2^2 + \rho\phi_{21})} \\ &\quad - \frac{\sqrt{(\sigma_2^4 + \phi_{22}^4 \sigma_2^4 + \phi_{21}^4 \sigma_1^4 + 4\rho^2 \phi_{22}^2 \phi_{21}^2 - 2\phi_{22}^2 \sigma_2^4 + 2\phi_{21}^2 \sigma_1^2 \sigma_2^2 + 2\phi_{22}^2 \phi_{21}^2 \sigma_1^2 \sigma_2^2 - 4\rho\phi_{22}\phi_{21}\sigma_2^2 + 4\rho\phi_{22}^3 \phi_{21}\sigma_2^2 + 4\rho\phi_{22}\phi_{21}^3 \sigma_1^2 - 4\rho^2 \phi_{21}^2)}}{2(\phi_{22}\sigma_2^2 + \rho\phi_{21})}. \end{aligned} \quad (32)$$

The covariance between  $w_{1t}$  and  $w_{2t}$  is equal to

$$\text{cov}(w_{1t}, w_{2t}) = \frac{(\phi_{11} - \theta_1)\phi_{21}\sigma_1^2 + (\phi_{22} - \theta_2)\phi_{12}\sigma_2^2 + (1 - \theta_2\phi_{11} - \theta_1\phi_{22} + \phi_{12}\phi_{21} + \phi_{11}\phi_{22})\rho}{1 - \theta_1\theta_2}.$$

<sup>14</sup>Contributions focusing on nonlinearities issues for aggregated variables are not numerous. Among them, we refer to: Granger and Lee (1999), for an analysis of the effects of three aggregation schemes (contemporaneous and temporal aggregation, systematic sampling) on nonlinearity; Proietti (2006), for a derivation of the estimator of the fixed and random effects of a nonlinearly temporally aggregated mixed model.

<sup>15</sup>In all the other cases, the difference is around  $10^{-3}$ .

As a result, since  $E(w_{1t}^2) = \frac{\phi_{11}\sigma_1^2 + \rho\phi_{12}}{\theta_1}$  and  $E(w_{2t}^2) = \frac{\phi_{22}\sigma_2^2 + \rho\phi_{21}}{\theta_2}$ , the MSE of the disaggregate predictor of  $y_t$ , that is the predictor based on univariate components of  $\mathbf{x}_t$ , is

$$\begin{aligned} \Sigma_y^u(1) &= \eta_1^2 \left( \frac{\phi_{11}\sigma_1^2 + \rho\phi_{12}}{\theta_1} \right) + \eta_2^2 \left( \frac{\phi_{22}\sigma_2^2 + \rho\phi_{21}}{\theta_2} \right) \\ &+ 2\eta_1\eta_2 \left( \frac{(\phi_{11} - \theta_1)\phi_{21}\sigma_1^2 + (\phi_{22} - \theta_2)\phi_{12}\sigma_2^2 + (1 - \theta_2\phi_{11} - \theta_1\phi_{22} + \phi_{12}\phi_{21} + \phi_{11}\phi_{22})\rho}{1 - \theta_1\theta_2} \right), \end{aligned} \quad (33)$$

where  $\theta_1$  and  $\theta_2$  are given in (32). Notice that (33) is a more complicated expression than (16), since it takes into account the heteroskedasticity of the innovations and some generic weights,  $\eta_1$  and  $\eta_2$ .

Second, we focus on the parameters of the aggregate predictor: this latter corresponds to the contemporaneously aggregated model in (18), with weights equal to  $\eta_1$  and  $\eta_2$ . The moving average parameter  $\psi$  in (18) can be recovered as in Section 3.2, with

$$\delta := \frac{E(y_t^2)}{2E(y_t y_{t-1})} = \frac{\eta_1^2(1 + \alpha_1^2)\sigma_1^2 + \eta_2^2(1 + \alpha_2^2)\sigma_2^2 + 2\eta_1\eta_2(1 + \alpha_1\alpha_2)\rho}{2(\eta_1^2\alpha_1\sigma_1^2 + \eta_2^2\alpha_2\sigma_2^2 + \eta_1\eta_2(\alpha_1 + \alpha_2)\rho)}.$$

After some algebra, we get the following expression for  $\psi$ :

$$\begin{aligned} \psi &= \frac{\eta_1^2(1 + \alpha_1^2)\sigma_1^2 + \eta_2^2(1 + \alpha_2^2)\sigma_2^2 + 2\eta_1\eta_2(1 + \alpha_1\alpha_2)\rho}{2(\eta_1^2\alpha_1\sigma_1^2 + \eta_2^2\alpha_2\sigma_2^2 + \eta_1\eta_2(\alpha_1 + \alpha_2)\rho)} \\ &- \frac{\sqrt{(\eta_1^2(\alpha_1 - 1)^2\sigma_1^2 + \eta_2^2(\alpha_2 - 1)^2\sigma_2^2 + 2\eta_1\eta_2(\alpha_1 - 1)(\alpha_2 - 1)\rho)(\eta_1^2(\alpha_1 + 1)^2\sigma_1^2 + \eta_2^2(\alpha_2 + 1)^2\sigma_2^2 + 2\eta_1\eta_2(\alpha_1 + 1)(\alpha_2 + 1)\rho)}}{2(\eta_1^2\alpha_1\sigma_1^2 + \eta_2^2\alpha_2\sigma_2^2 + \eta_1\eta_2(\alpha_1 + \alpha_2)\rho)}. \end{aligned} \quad (34)$$

Furthermore, the variance of the innovations of the aggregate predictor is

$$\sigma_v^2 = \frac{\eta_1^2\alpha_1\sigma_1^2 + \eta_2^2\alpha_2\sigma_2^2 + \eta_1\eta_2(\alpha_1 + \alpha_2)\rho}{\psi}, \quad (35)$$

where  $\psi$  is given in (34). Notice that, with respect to (21) and (20), (35) and (34) contain the extra parameters  $\sigma_1^2$ ,  $\sigma_2^2$ ,  $\eta_1$  and  $\eta_2$ .

We now turn the attention to the parameters of the disaggregate predictor. Plugging the estimated VMA's parameters  $\hat{\phi}_{11}$ ,  $\hat{\phi}_{12}$ ,  $\hat{\phi}_{21}$ ,  $\hat{\phi}_{22}$ ,  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  and  $\hat{\rho}$  in (32) and (33), we get the “implied” parameters of the disaggregate predictor, i.e.  $\theta_1$  and  $\theta_2$ , together with the “implied” root MSEs. Results for  $\theta_1$  and  $\theta_2$  are provided in the first two columns of Table A4, where the “estimated” counterparts based on the 1948-1995, 1948-1996, 1948-1997, 1948-1998 samples are also reported. These “estimated” parameters are obtained by fitting univariate MA(1) models separately to overnight deposits and currency in circulation.<sup>16</sup>

[TABLE A4 ABOUT HERE]

<sup>16</sup>Estimation results are available from the authors upon request.



The implied and estimated values are extremely close, particularly for  $\theta_1$ , and for each estimation sample used. These results are striking, in our view, since the estimated models that are used to calculate the “implied” and “estimated” parameters are based on less than 50 observations of annual data.

The first column of Table A5 presents the “implied” and “estimated” root MSEs of the disaggregate predictor. The “implied” and “estimated” root MSEs are in general close to each other, particularly working with the 1948-1996 and 1948-1997 samples.

[TABLE A5 ABOUT HERE]

Back to the parameters of the aggregate predictor. The “implied” parameters are calculated by plugging the estimated  $\hat{\phi}_{11}$ ,  $\hat{\phi}_{12}$ ,  $\hat{\phi}_{21}$ ,  $\hat{\phi}_{22}$ ,  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  and  $\hat{\rho}$  in (34) and (35), for each estimation sample. The third column of Table A4 displays the “implied” and “estimated”  $\psi$  parameters of the aggregate variable, i.e. the percentage changes of M1. For  $\psi$ , in every estimation sample, the estimated values are roughly the values implied by (34). The second column of Table A5 shows the “implied” and “estimated” root MSEs of the aggregate predictor (for each estimation sample), which are also very close.

The third column of Table A5 ranks the two competing predictors in terms of root MSE. The aggregate predictor ranks first for every estimation sample and on the basis of the “implied” and “estimated” root MSEs: therefore, our results are fully coherent and demonstrate that, on the basis of the available sample, forecasting M1 directly is more efficient (in mean squared error sense) than aggregating ex-post disaggregate univariate forecasts for overnight deposits and currency in circulation.

Of course, we stress that our results hold true using the years 1996-1998 as forecasting period and working with a data set relatively small in size. Moreover, a central assumption is that the bivariate VMA(1) in (30) is an adequate approximation of the “true” data generating vector process, which is unknown. To check the robustness of our empirical results, we could present more formal test of forecast comparison and use alternative metrics for determining which forecasting method is expected to be most accurate (rather than the forecast MSE). This is out of the scope of the paper and is left for future research.

## 6 Conclusions and outlook

In this study, we address the issue of forecasting a contemporaneously aggregated vector process. To this aim, we focus on predictors based on the aggregate process directly and on the individual components of the disaggregate vector process. If the DGP is known and no parameters uncertainty is faced, these two predictors are sub-optimal (in mean squared error sense) with respect to the optimal predictor that can be built by aggregating forecasts of the original data generating vector process.

In a bivariate framework of a vector MA(1), which is the benchmark used by the whole of the aggregation literature, we provide the necessary and sufficient condition for the equality of MSEs associated with the two competing predictive procedures. We explain how the MSEs can be expressed as a function of the parameters of the original DGP. In this sense, the structure of the disaggregate process determines the relative forecasting accuracy of the two predictors, as recognized but not fully investigated by the early contributors to the aggregation debate.

Furthermore, we show that the condition given by Lütkepohl (1984b, 1987, 2004), although necessary and sufficient for the equality of the predictors, is sufficient but not necessary for the equality of the corresponding forecast MSEs. Finally, we provide sufficient conditions for the equality of MSEs. With these conditions, we give evidence that specific assumptions on the parameters of the VMA(1) guarantee equal forecasting accuracy for the prediction approaches under scrutiny. Monte Carlo simulations seem to confirm our findings. An empirical application to M1 forecasting illustrates the main issues and how the techniques can be applied in practice.

## References

- [1] Baffigi A., Golinelli R. and Parigi G. (2004). Bridge Models to Forecast the Euro area GDP. *International Journal of Forecasting* 20, 447-460.
- [2] Benalal N., Diaz del Hoyo J. L., Landau B., Roma M. and Skudelny F. (2004). To Aggregate or not to Aggregate? Euro Area Inflation Forecasting. European Central Bank Working Paper Seris No. 374.
- [3] Bodo G., Golinelli R. and Parigi G. (2000). Forecasting Industrial Production in the Euro Area. *Empirical Economics* 25, 541-561.
- [4] Dedola L., Gaiotti E. and Silipo L. (2001). Money Demand in the Euro Area: Do National Differences Matter? Banca d'Italia, Temi di Discussione n. 405.
- [5] Espasa A., Senra E. and Albacete R. (2002). Forecasting Inflation in the European Monetary Union: A Disaggregated Approach by Countries and by Sectors. *The European Journal of Finance* 8, 402-421.
- [6] Fagan G. and Henry J. (1998). Long Run Money Demand in the EU: Evidence for Area-Wide Aggregates. *Empirical Economics* 23, 483-506.
- [7] Giacomini R. and Granger C. W. J. (2004). Aggregation of Space-Time Processes. *Journal of Econometrics* 118, 7-26.
- [8] Granger C. W. J. (1990). Aggregation of Time Series Variables: A Survey. In *Disaggregation in Econometric Modelling*, edited by T. Barker and H. Pesaran, London: Routledge.
- [9] Granger C.W.J. and Lee T.H. (1999). The Effect of Aggregation on Nonlinearity. *Econometric Reviews* 18, 259-269.
- [10] Granger C. W. J. and Newbold P. (1986). *Forecasting Economic Time Series*. Second Edition. New York: Academic Press.
- [11] Grunfeld Y. and Griliches Z. (1960). Is Aggregation Necessarily Bad? *The Review of Economics and Statistics* 42, 1-13.
- [12] Hamilton J. D. (1994). *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.
- [13] Hendry D. and Hubrich K. (2007). Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate. Mimeo.
- [14] Hubrich K. (2005). Forecasting Euro Area Inflation: Does Aggregating Forecasts by HICP Component Improve Forecast Accuracy? *International Journal of Forecasting* 21, 119-136.

- [15] Kohn R. (1982). When is an Aggregate of a Time Series Efficiently Forecast by its Past? *Journal of Econometrics* 18, 337-349.
- [16] Lütkepohl H. (1984a). Linear Aggregation of Vector Autoregressive Moving Average Processes. *Economics Letters* 14, 345-350.
- [17] Lütkepohl H. (1984b). Linear Transformations of Vector ARMA Processes. *Journal of Econometrics* 26, 283-293.
- [18] Lütkepohl H. (1984c). Forecasting Contemporaneously Aggregated Vector ARMA Processes. *Journal of Business & Economic Statistics* 2, 201-214.
- [19] Lütkepohl H. (1987). *Forecasting Aggregated Vector ARMA Processes*. Berlin: Springer-Verlag.
- [20] Lütkepohl H. (2004). Forecasting with VARMA Models. EUI Working Paper ECO No. 2004/25.
- [21] Lütkepohl H. (2007). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.
- [22] Lütkepohl H. and Xu F. (2009). The Role of the Log Transformation in Forecasting Economic Variables. Mimeo.
- [23] Man K. S. (2004). Linear Prediction of Temporal Aggregates Under Model Misspecification. *International Journal of Forecasting* 20, 659-670.
- [24] Marcellino M., Stock J. H. and Watson M. W. (2003). Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-Wide Information. *European Economic Review* 47, 1-18.
- [25] Pino F. A., Moretton P. A. and Mentz R. P. (1987). Modelling and Forecasting Linear Combinations of Time Series. *International Statistical Review* 55, 295-313.
- [26] Proietti T. (2006). On the Estimation of Nonlinearly Aggregated Mixed Models. *Journal of Computational and Graphical Statistics* 15, 18-38.
- [27] Rose D. E. (1977). Forecasting Aggregates of Independent ARIMA Processes. *Journal of Econometrics* 5, 323-345.
- [28] Sbrana G. (2008a). On the Use of Area-Wide Models in the Euro-zone: The Money Demand Case. *Statistical Methods and Applications* 17, 499-518.
- [29] Sbrana G. (2008b). Aggregation of Vector ARMA Processes: Some Further Results. Mimeo.
- [30] Theil H. (1954). *Linear Aggregation of Economic Relations*. Amsterdam: North Holland.
- [31] Tiao G. C. and Guttman I. (1980). Forecasting Contemporaneous Aggregates of Multiple Time Series. *Journal of Econometrics* 12, 219-230.
- [32] Van Garderen K. J., Lee K. and Pesaran M. K. (2000). Cross-sectional aggregation of non-linear models. *Journal of Econometrics* 95, 285-331.
- [33] Wei W. W. S. and Abraham B. (1981). Forecasting Contemporaneous Time Series Aggregates. *Communications in Statistics - Theory and Methods* A10, 1335-1344.
- [34] Wesche K. (1997). The Stability of European Money Demand: An Investigation of M3H. *Open Economies Review* 8, 371-391.
- [35] Zellner A. and Tobias J. (2000). A Note on Aggregation, Disaggregation and Forecasting Performance. *Journal of Forecasting* 19, 457-469.
- [36] Zizza R. (2002). Forecasting the Industrial Production Index for the Euro Area Through Forecasts for the Main Countries. Banca d'Italia, Temi di Discussione n. 441.

# Appendix

## Proof of Proposition 3

We assume  $\phi_{21} = (\phi_{11} - \phi_{22})(\frac{1}{2} + \rho)$ ,  $\phi_{12} = \frac{\phi_{11} - \phi_{22}}{2}$  and  $\rho \neq 0$ . As stated in Proposition 3, in what follows we are going to show that this linear combination of the DGP's parameters guarantees the equality of forecasting performance of the competitive processes in (6) and (9), no matter the values of  $\phi_{11}$ ,  $\phi_{22}$  and  $\rho$ .

Bearing in mind equations (14), it is easy to see that  $\theta_1$  and  $\theta_2$  simplify to

$$\theta_1 = \frac{4 + (5 + 4\rho)\phi_{11}^2 + \phi_{22}^2 - 2\phi_{11}\phi_{22}(1 + 2\rho)}{4((2 + \rho)\phi_{11} - \rho\phi_{22})} - \frac{\sqrt{(4 + 8\phi_{11} + 4\rho\phi_{11} + 5\phi_{11}^2 + 4\rho\phi_{11}^2 - 4\rho\phi_{22} - 2\phi_{11}\phi_{22} - 4\rho\phi_{11}\phi_{22} + \phi_{22}^2)(4 - 8\phi_{11} - 4\rho\phi_{11} + 5\phi_{11}^2 + 4\rho\phi_{11}^2 + 4\rho\phi_{22} - 2\phi_{11}\phi_{22} - 4\rho\phi_{11}\phi_{22} + \phi_{22}^2)}}{4((2 + \rho)\phi_{11} - \rho\phi_{22})}, \quad (\text{A1})$$

and

$$\theta_2 = \frac{4 + (\phi_{11} + 2\rho\phi_{11})^2 + (5 - 4\rho^2)\phi_{22}^2 - 2\phi_{11}\phi_{22}(1 + 2\rho)}{(4\rho(1 + 2\rho)(\phi_{11} - \phi_{22}) + 8\phi_{22})} - \frac{\sqrt{(4 + 4\rho\phi_{11} + 8\rho^2\phi_{11} + \phi_{11}^2 + 4\rho\phi_{11}^2 + 4\rho^2\phi_{11}^2 + 8\phi_{22} - 4\rho\phi_{22} - 8\rho^2\phi_{22} - 2\phi_{11}\phi_{22} - 4\rho\phi_{11}\phi_{22} + 5\phi_{22}^2 - 4\rho^2\phi_{22}^2)}}{\sqrt{(4\rho(1 + 2\rho)(\phi_{11} - \phi_{22}) + 8\phi_{22})}} \times \frac{\sqrt{(4 - 4\rho\phi_{11} - 8\rho^2\phi_{11} + \phi_{11}^2 + 4\rho\phi_{11}^2 + 4\rho^2\phi_{11}^2 - 8\phi_{22} + 4\rho\phi_{22} + 8\rho^2\phi_{22} - 2\phi_{11}\phi_{22} - 4\rho\phi_{11}\phi_{22} + 5\phi_{22}^2 - 4\rho^2\phi_{22}^2)}}{\sqrt{(4\rho(1 + 2\rho)(\phi_{11} - \phi_{22}) + 8\phi_{22})}}. \quad (\text{A2})$$

As a consequence  $\theta_1 \neq \theta_2$ . Some straightforward calculations show that the  $\psi$  parameter in (20) is equal to  $\theta_1$ .

Furthermore, the variance of the aggregate process is provided in (21), which for  $\phi_{12} = \frac{\phi_{11} - \phi_{22}}{2}$  and  $\phi_{21} = (\phi_{11} - \phi_{22})(1/2 + \rho)$  is

$$\Sigma_y(1) = \sigma_v^2 = \frac{(1 + \rho)((2 + \rho)\phi_{11} - \rho\phi_{22})}{\theta_1}. \quad (\text{A3})$$

The MSE of the optimal one-step ahead predictor of  $y_t$  based on the univariate components of  $\mathbf{x}_t$  is given in (16), which for  $\phi_{12} = \frac{\phi_{11} - \phi_{22}}{2}$  and  $\phi_{21} = (\phi_{11} - \phi_{22})(1/2 + \rho)$  becomes

$$\Sigma_y^u(1) = \frac{\phi_{11} + \frac{1}{2}\rho(\phi_{11} - \phi_{22})}{\theta_1} + \frac{\rho(\frac{1}{2} + \rho)(\phi_{11} - \phi_{22}) + \phi_{22}}{\theta_2} + \frac{\frac{1}{2}(2\rho^2(\phi_{11} - \phi_{22})^2 + 2(\phi_{11} - \phi_{22})(\phi_{11} + \phi_{22} - \theta_1 - \theta_2) + \rho(4 + 5\phi_{11}^2 + \phi_{22}^2 - 2\phi_{11}(\phi_{22} + 2(\theta_1 + \theta_2))))}{1 - \theta_1\theta_2}} \quad (\text{A4})$$

To have equal forecasting performance, it has to be  $\Sigma_y(1) = \sigma_v^2 = \Sigma_y^u(1)$ . For this condition to be verified, on the basis of (A3) and (A4), it must hold

$$\frac{\frac{1}{2}(1 + 2\rho)((2 + \rho)\phi_{11} - \rho\phi_{22})}{\theta_1} - \frac{\rho(\frac{1}{2} + \rho)(\phi_{11} - \phi_{22}) + \phi_{22}}{\theta_2} = \frac{\frac{1}{2}(2\rho^2(\phi_{11} - \phi_{22})^2 + \rho(4 + 5\phi_{11}^2 + \phi_{22}^2 - 2\phi_{11}(\phi_{22} + 2(\theta_1 + \theta_2))) + 2(\phi_{11} - \phi_{22})(\phi_{11} + \phi_{22} - \theta_1 - \theta_2))}{1 - \theta_1\theta_2}},$$

which yields

$$\frac{\theta_1\theta_2}{1 - \theta_1\theta_2} = \frac{-2\rho^2(\phi_{11} - \phi_{22})(\theta_1 - \theta_2) - \rho(\phi_{11}\theta_1 - \phi_{22}\theta_1 - 5\phi_{11}\theta_2 + \phi_{22}\theta_2) + 2\phi_{11}\theta_2 - 2\phi_{22}\theta_1}{2\rho^2(\phi_{11} - \phi_{22})^2 + \rho(4 + 5\phi_{11}^2 + \phi_{22}^2 - 2\phi_{11}(\phi_{22} + 2(\theta_1 + \theta_2))) + 2(\phi_{11} - \phi_{22})(\phi_{11} + \phi_{22} - \theta_1 - \theta_2)}.$$

Let us focus on the first ratio  $\frac{\theta_1\theta_2}{1 - \theta_1\theta_2}$ . Substituting for (A1) and (A2), we get

$$\begin{aligned}
& \frac{\theta_1 \theta_2}{1 - \theta_1 \theta_2} \\
& = \\
& -\frac{1}{2} + \frac{(\rho(1+2\rho)(\phi_{11} - \phi_{22}) + 2\phi_{22})\sqrt{(4+8\phi_{11}+4\rho\phi_{11}+5\phi_{11}^2+4\rho\phi_{11}^2-4\rho\phi_{22}-2\phi_{11}\phi_{22}-4\rho\phi_{11}\phi_{22}+\phi_{22}^2)}}{\sqrt{(4(\rho-1)(\rho+1)(\phi_{11}-\phi_{22})(4+(1+2\rho)\phi_{11}^2-2(3+2\rho)\phi_{11}\phi_{22}+(1+2\rho)\phi_{22}^2))}} \\
& \quad \times \frac{\sqrt{(4-8\phi_{11}-4\rho\phi_{11}+5\phi_{11}^2+4\rho\phi_{11}^2+4\rho\phi_{22}-2\phi_{11}\phi_{22}-4\rho\phi_{11}\phi_{22}+\phi_{22}^2)}}{\sqrt{(4(\rho-1)(\rho+1)(\phi_{11}-\phi_{22})(4+(1+2\rho)\phi_{11}^2-2(3+2\rho)\phi_{11}\phi_{22}+(1+2\rho)\phi_{22}^2))}} \\
& + \frac{(-(2+\rho)\phi_{11} + \rho\phi_{22})\sqrt{(4+4\rho\phi_{11}+8\rho^2\phi_{11}+\phi_{11}^2+4\rho\phi_{11}^2+4\rho^2\phi_{11}^2+8\phi_{22}-4\rho\phi_{22}-8\rho^2\phi_{22}-2\phi_{11}\phi_{22}-4\rho\phi_{11}\phi_{22}+5\phi_{22}^2-4\rho^2\phi_{22}^2)}}{\sqrt{(4(\rho-1)(\rho+1)(\phi_{11}-\phi_{22})(4+(1+2\rho)\phi_{11}^2-2(3+2\rho)\phi_{11}\phi_{22}+(1+2\rho)\phi_{22}^2))}} \\
& \quad \times \frac{\sqrt{(4-4\rho\phi_{11}-8\rho^2\phi_{11}+\phi_{11}^2+4\rho\phi_{11}^2+4\rho^2\phi_{11}^2-8\phi_{22}+4\rho\phi_{22}+8\rho^2\phi_{22}-2\phi_{11}\phi_{22}-4\rho\phi_{11}\phi_{22}+5\phi_{22}^2-4\rho^2\phi_{22}^2)}}{\sqrt{(4(\rho-1)(\rho+1)(\phi_{11}-\phi_{22})(4+(1+2\rho)\phi_{11}^2-2(3+2\rho)\phi_{11}\phi_{22}+(1+2\rho)\phi_{22}^2))}}.
\end{aligned}$$

After tedious calculations it is possible to see that we get exactly the same expression as above after plugging (A1) and (A2) in the second ratio

$$\frac{-2\rho^2(\phi_{11}-\phi_{22})(\theta_1-\theta_2)-\rho(\phi_{11}\theta_1-\phi_{22}\theta_1-5\phi_{11}\theta_2+\phi_{22}\theta_2)+2\phi_{11}\theta_2-2\phi_{22}\theta_1}{2\rho^2(\phi_{11}-\phi_{22})^2+\rho(4+5\phi_{11}^2+\phi_{22}^2-2\phi_{11}(\phi_{22}+2(\theta_1+\theta_2)))+2(\phi_{11}-\phi_{22})(\phi_{11}+\phi_{22}-\theta_1-\theta_2)}.$$

This completes the proof. □

## Proof of Proposition 4

Let  $\phi_{12} = \phi_{21} = \frac{\phi_{22}-\phi_{11}}{2}$  and  $\rho = 0$ . Similar results, *mutatis mutandis*, hold for  $\phi_{12} = \phi_{21} = \frac{\phi_{11}-\phi_{22}}{2}$  and  $\rho = 0$ .

Bearing in mind equations (14), it is easy to see that  $\theta_1$  and  $\theta_2$  simplify to

$$\theta_1 = \frac{4 + 5\phi_{11}^2 - 2\phi_{11}\phi_{22} + \phi_{22}^2}{8\phi_{11}} - \sqrt{\frac{(4 + 5\phi_{11}^2 - 2\phi_{11}\phi_{22} + \phi_{22}^2)^2}{64\phi_{11}^2} - 1}$$

and

$$\theta_2 = \frac{4 + \phi_{11}^2 - 2\phi_{11}\phi_{22} + 5\phi_{22}^2}{8\phi_{22}} - \sqrt{\frac{(4 + \phi_{11}^2 - 2\phi_{11}\phi_{22} + 5\phi_{22}^2)^2}{64\phi_{22}^2} - 1},$$

hence,  $\theta_1 \neq \theta_2$ . The  $\psi$  parameter in (20) is equal to

$$\psi = \frac{1}{8\phi_{22}} \left( (4 + \phi_{11}^2 - 2\phi_{11}\phi_{22} + 5\phi_{22}^2) - \sqrt{(4 + \phi_{11}^2 - 2\phi_{11}\phi_{22} + 5\phi_{22}^2 - 8\phi_{22})(4 + \phi_{11}^2 - 2\phi_{11}\phi_{22} + 5\phi_{22}^2 + 8\phi_{22})} \right).$$

It differs from  $\theta_1$ . Moreover  $\psi = \theta_2$ .

Furthermore, when  $\rho = 0$ , the variance of the aggregate process is  $\sigma_v^2 = \frac{2\phi_{22}}{\psi}$ , that is,

$$\Sigma_y(1) = \sigma_v^2 = \frac{2\phi_{22}}{\theta_2}.$$

The MSE of the optimal one-step ahead predictor of  $y_t$  based on the univariate components of  $\mathbf{x}_t$  is given in (16), which for  $\phi_{12} = \phi_{21} = \frac{\phi_{22}-\phi_{11}}{2}$  and  $\rho = 0$  becomes

$$\Sigma_y^u(1) = \frac{\phi_{11}}{\theta_1} + \frac{\phi_{22}}{\theta_2} + \frac{\phi_{22} - \phi_{11}}{1 - \theta_1\theta_2} (\phi_{11} + \phi_{22} - \theta_1 - \theta_2).$$

To have equal forecasting performance, it has to be  $\Sigma_y^u(1) = \Sigma_y(1) = \sigma_v^2$ . For this condition to be verified, it must hold

$$\frac{\phi_{22}}{\theta_2} - \frac{\phi_{11}}{\theta_1} = \frac{\phi_{22} - \phi_{11}}{1 - \theta_1\theta_2} (\phi_{11} + \phi_{22} - \theta_1 - \theta_2).$$

Hence, to be  $\Sigma_y^u(1) = \Sigma_y(1) = \sigma_v^2$ , we need to show that

$$\frac{\phi_{22}\theta_1}{\phi_{11}\theta_2} = \frac{\theta_1^2 - \phi_{11}\theta_1 + 1}{\theta_2^2 - \phi_{22}\theta_2 + 1}. \quad (\text{A5})$$

Let us focus on the right-hand side (RHS) of (A5). After some tedious calculations, we notice that the numerator  $\theta_1^2 - \phi_{11}\theta_1 + 1$  can be factorized as

$$\begin{aligned} & \frac{1}{32\phi_{11}^2} (-2\phi_{22}\phi_{11} + \phi_{22}^2 + 4 + \phi_{11}^2) \\ & \times \left( 4 + 5\phi_{11}^2 + \phi_{22}^2 - 2\phi_{22}\phi_{11} - \sqrt{\frac{(\phi_{22}^2 + 4 - 8\phi_{11} - 2\phi_{22}\phi_{11} + 5\phi_{11}^2)(\phi_{22}^2 + 4 + 8\phi_{11} - 2\phi_{22}\phi_{11} + 5\phi_{11}^2)}{\phi_{11}^2}} \phi_{11} \right). \end{aligned}$$

Similarly, the denominator  $\theta_2^2 - \phi_{22}\theta_2 + 1$  may be factorized as

$$\begin{aligned} & \frac{1}{32\phi_{22}^2} (-2\phi_{22}\phi_{11} + \phi_{22}^2 + 4 + \phi_{11}^2) \\ & \times \left( 4 + 5\phi_{22}^2 + \phi_{11}^2 - 2\phi_{22}\phi_{11} - \sqrt{\frac{(\phi_{11}^2 + 4 - 8\phi_{22} - 2\phi_{22}\phi_{11} + 5\phi_{22}^2)(\phi_{11}^2 + 4 + 8\phi_{22} - 2\phi_{22}\phi_{11} + 5\phi_{22}^2)}{\phi_{22}^2}} \phi_{22} \right). \end{aligned}$$

Consequently we can express the ratio  $\frac{\theta_1^2 - \phi_{11}\theta_1 + 1}{\theta_2^2 - \phi_{22}\theta_2 + 1}$  as

$$\frac{\phi_{22}^2 \left( 4 + 5\phi_{11}^2 + \phi_{22}^2 - 2\phi_{22}\phi_{11} - \sqrt{\frac{(\phi_{22}^2 + 4 - 8\phi_{11} - 2\phi_{22}\phi_{11} + 5\phi_{11}^2)(\phi_{22}^2 + 4 + 8\phi_{11} - 2\phi_{22}\phi_{11} + 5\phi_{11}^2)}{\phi_{11}^2}} \phi_{11} \right)}{\phi_{11}^2 \left( 4 + 5\phi_{22}^2 + \phi_{11}^2 - 2\phi_{22}\phi_{11} - \sqrt{\frac{(\phi_{11}^2 + 4 - 8\phi_{22} - 2\phi_{22}\phi_{11} + 5\phi_{22}^2)(\phi_{11}^2 + 4 + 8\phi_{22} - 2\phi_{22}\phi_{11} + 5\phi_{22}^2)}{\phi_{22}^2}} \phi_{22} \right)}.$$

In addition, focus on the left-hand side (LHS) of (A5). We can express  $\theta_2$  as

$$\theta_2 = \frac{1}{8\phi_{22}} \left( (4 + \phi_{11}^2 - 2\phi_{11}\phi_{22} + 5\phi_{22}^2) - \sqrt{(4 + \phi_{11}^2 - 2\phi_{11}\phi_{22} + 5\phi_{22}^2 - 8\phi_{22})(4 + \phi_{11}^2 - 2\phi_{11}\phi_{22} + 5\phi_{22}^2 + 8\phi_{22})} \right). \quad (\text{A6})$$

Similarly to  $\theta_2$ , we can express  $\theta_1$  as

$$\theta_1 = \frac{1}{8\phi_{11}} \left( (4 + 5\phi_{11}^2 - 2\phi_{11}\phi_{22} + \phi_{22}^2) - \sqrt{(4 + 5\phi_{11}^2 - 2\phi_{11}\phi_{22} + \phi_{22}^2 - 8\phi_{11})(4 + 5\phi_{11}^2 - 2\phi_{11}\phi_{22} + \phi_{22}^2 + 8\phi_{11})} \right). \quad (\text{A7})$$

From (A7) and (A6), it is immediately evident that the LHS and the RHS of (A5) are equal, and  $\Sigma_y^u(1) = \Sigma_y(1)$ . Therefore, the result follows.  $\square$

Table A1: MSEs of competitive models using DGPs 1, 2, 3

MSE of competitive models using DGPs 1, 2, 3							
Model Sample	Selection Steps	DGP 1		DGP 2		DGP 3	
		AIC	BIC	AIC	BIC	AIC	BIC
T=30	h=1	1.17	1.13	1.05	1.03	0.87	0.90
	h=5	1.00	1.00	1.04	1.03	0.99	0.98
T=50	h=1	1.13	1.12	1.01	1.00	0.88	0.90
	h=5	1.00	1.00	1.01	1.01	1.00	1.00
T=100	h=1	1.09	1.08	1.00	0.99	0.90	0.92
	h=5	1.02	1.00	1.01	1.01	1.00	0.99
T=200	h=1	1.08	1.08	1.00	0.99	0.89	0.91
	h=5	1.00	1.00	1.01	1.01	1.00	1.00
T=500	h=1	1.10	1.10	1.00	1.00	0.88	0.89
	h=5	1.01	1.00	1.01	1.00	1.01	1.00

Table A2: Descriptive statistics

Variables	Mean	St. dev.	Max	Min	Skewness	Kurtosis
$\Delta \log(OV_t)$	0.1276	0.0639	0.2743	-0.0032	0.1696	2.6243
$\Delta \log(CC_t)$	0.0957	0.0389	0.1835	0.0196	0.2584	2.5626

The top part of the table is for overnight deposits (delta logs), the bottom part for currency in circulation (delta logs). For both series the sample goes from 1948 until 1998.

Table A3: Estimated VMA(1) models

Parameters	Estimates	Standard error
Sample: 1948-1995		
$\hat{\phi}_{11}$	0.4788	0.1435
$\hat{\phi}_{12}$	0.1073	0.1075
$\hat{\phi}_{21}$	0.0311	0.2500
$\hat{\phi}_{22}$	0.1271	0.1459
$\hat{\sigma}_1^2$	0.0030	0.0006
$\hat{\rho}$	0.0008	0.0003
$\hat{\sigma}_2^2$	0.0013	0.0003
Sample: 1948-1996		
$\hat{\phi}_{11}$	0.4629	0.1377
$\hat{\phi}_{12}$	0.1080	0.1096
$\hat{\phi}_{21}$	0.0767	0.2432
$\hat{\phi}_{22}$	0.1668	0.1480
$\hat{\sigma}_1^2$	0.0030	0.0006
$\hat{\rho}$	0.0008	0.0003
$\hat{\sigma}_2^2$	0.0014	0.0003
Sample: 1948-1997		
$\hat{\phi}_{11}$	0.4683	0.1401
$\hat{\phi}_{12}$	0.1085	0.1079
$\hat{\phi}_{21}$	0.0882	0.2433
$\hat{\phi}_{22}$	0.1670	0.1437
$\hat{\sigma}_1^2$	0.0030	0.0006
$\hat{\rho}$	0.0008	0.0003
$\hat{\sigma}_2^2$	0.0013	0.0003
Sample: 1948-1998		
$\hat{\phi}_{11}$	0.4723	0.1367
$\hat{\phi}_{12}$	0.1131	0.1075
$\hat{\phi}_{21}$	0.0798	0.2417
$\hat{\phi}_{22}$	0.1633	0.1433
$\hat{\sigma}_1^2$	0.0029	0.0006
$\hat{\rho}$	0.0008	0.0003
$\hat{\sigma}_2^2$	0.0013	0.0002

The estimated model is the VMA(1) in (30). Estimation is conducted by maximum-likelihood. The covariance of the parameters is computed by the following method: inverse of computed Hessian. The GAUSS program `vecarma.e`, written by Ron Schoenberg, is used for estimation. Four different samples are employed: 1948-1995, 1948-1996, 1948-1997 and 1948-1998.



Table A4: Estimated and implied parameters

<b>Sample</b>	$\theta_1$	$\theta_2$	$\psi$
	<b>estimated</b>	<b>estimated</b>	<b>estimated</b>
1948-1995	0.4918	0.2187	0.4889
1948-1996	0.4915	0.2689	0.4914
1948-1997	0.5014	0.2729	0.4996
1948-1998	0.5025	0.2725	0.5004
	<b>implied</b>	<b>implied</b>	<b>implied</b>
1948-1995	0.5046	0.1460	0.4771
1948-1996	0.4887	0.2103	0.5072
1948-1997	0.4944	0.2178	0.5228
1948-1998	0.5003	0.2098	0.5182
<b>Model</b>	<b>ARMA(0,1)</b>	<b>ARMA(0,1)</b>	<b>ARMA(0,1)</b>
<b>Constant</b>	<b>no</b>	<b>no</b>	<b>no</b>

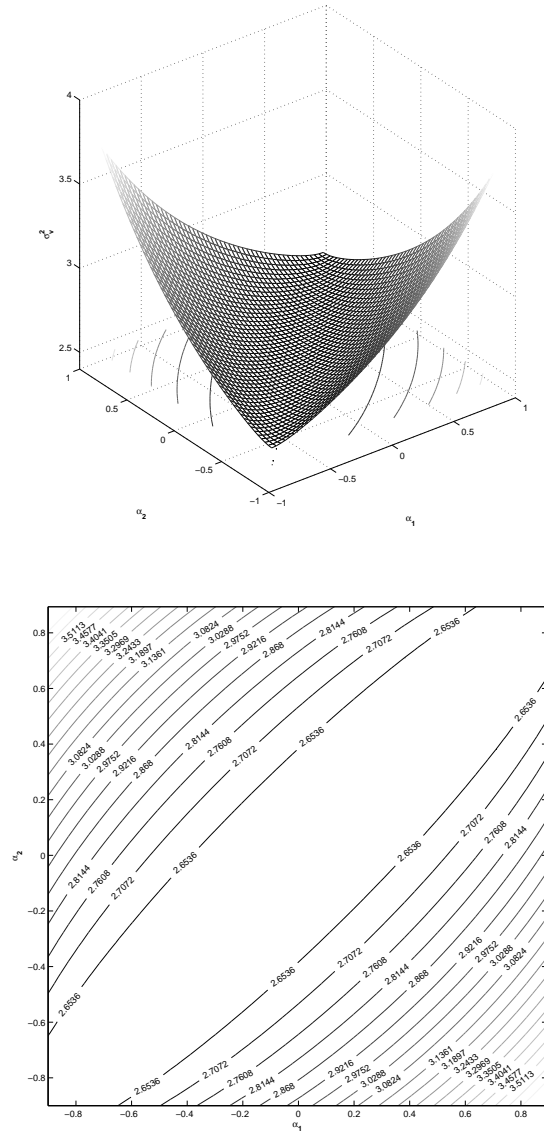
Estimated and implied parameters for the univariate disaggregate models in (12), where  $x_{1t} = \Delta \log(OV_t)$  and  $x_{2t} = \Delta \log(CC_t)$ , and for the contemporaneously aggregated model in (18), where  $y_t = \Delta \log(M1_t)$ . All the variables are demeaned.

Table A5: Estimated and implied forecast root MSEs

<b>Estimation sample</b>	<b>Disaggregate predictor</b>	<b>Aggregate predictor</b>	<b>Best predictor</b>
	<b>estimated</b>	<b>estimated</b>	<b>estimated</b>
1948-1995	0.09184	0.05262	Aggregate
1948-1996	0.04264	0.03106	Aggregate
1948-1997	0.04479	0.02830	Aggregate
	<b>implied</b>	<b>implied</b>	<b>implied</b>
1948-1995	0.04545	0.04542	Aggregate
1948-1996	0.04553	0.04549	Aggregate
1948-1997	0.04545	0.04542	Aggregate

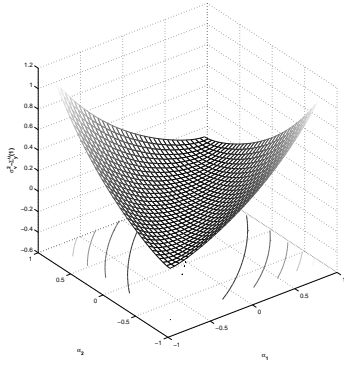
Forecast root MSEs (estimated and implied) for the aggregate and disaggregate predictors. For each estimation sample, the “best” predictor is the one with the lowest root MSE.

Figure A1: Aggregate predictor: contour and three-dimensional plots of  $\sigma_v^2$

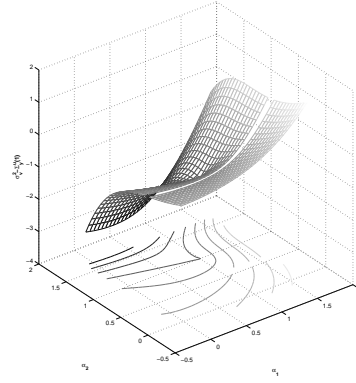


From top to down: three-dimensional and contour plots of  $\sigma_v^2$ , i.e., the variance of the aggregate predictor, as a function of the parameters of the bivariate VMA(1) in (11):  $\alpha_1 = (\phi_{11} + \phi_{21})$  and  $\alpha_2 = (\phi_{12} + \phi_{22})$ . In both figures, the extra-diagonal element of the covariance matrix of the innovations  $\rho$  is set equal to 0.3.

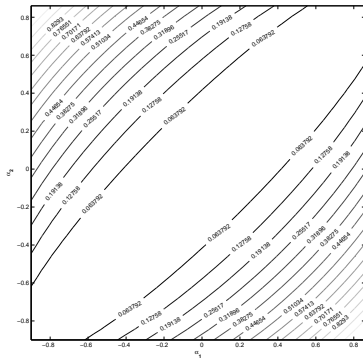
Figure A2: Proposition 1. Three-dimensional and contour plots of  $\sigma_v^2 - \Sigma_y^u(1)$



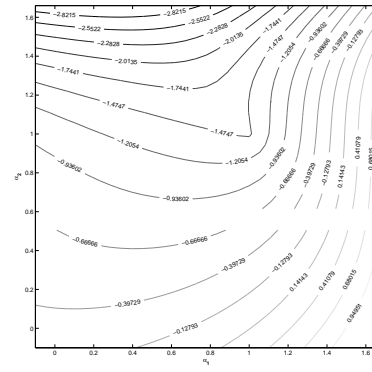
(a)  $\rho = 0.3, \phi_{12} = 0, \phi_{21} = 0;$



(b)  $\rho = 0.3, \phi_{12} = 0.8, \phi_{21} = 0.8;$



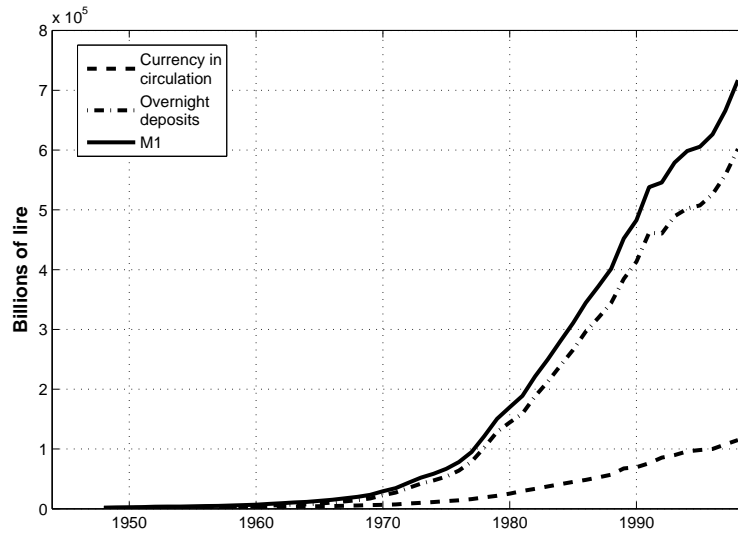
(c)  $\rho = 0.3, \phi_{12} = 0, \phi_{21} = 0;$



(d)  $\rho = 0.3, \phi_{12} = 0.8, \phi_{21} = 0.8$

Three-dimensional and contour plots of  $\sigma_v^2 - \Sigma_y^u(1)$ , i.e., the variance of the aggregate predictor minus the variance of the disaggregate predictor, as a function of the parameters of the bivariate VMA(1) in (11):  $\alpha_1 = (\phi_{11} + \phi_{21})$  and  $\alpha_2 = (\phi_{12} + \phi_{22})$ . In the figures, two of the parameters (i.e.,  $\phi_{11}, \phi_{22}$ ) vary while the other three parameters (i.e.,  $\rho, \phi_{12}, \phi_{21}$ ) are set equal to the values below each panel.

Figure A3: Main components of M1



Outstanding amounts of M1, currency in circulation and overnight deposits. End-of-year data ranging from 1948 until 1998. Source: Bank of Italy.

Figure A4: Relation (31)

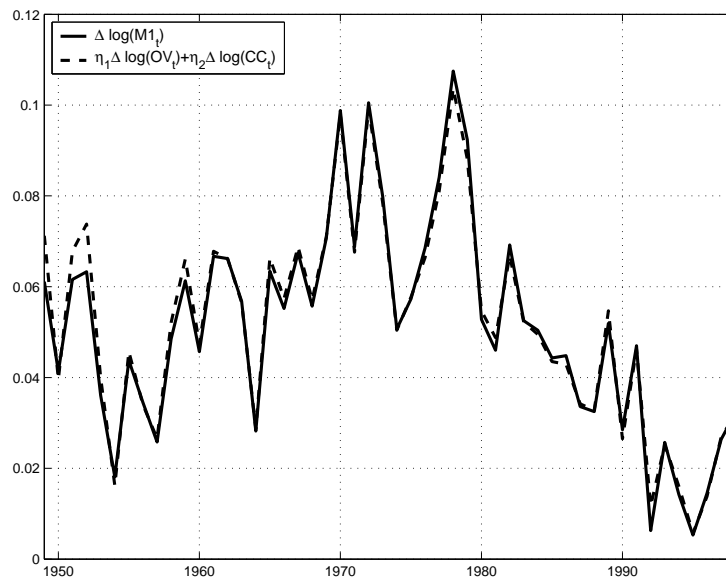


Illustration of the approximate relation in (31), where logarithms of M1, overnight deposits and currency in circulation are involved. End-of-year data ranging from 1948 until 1998. Source: Bank of Italy.

## Recent titles

### CORE Discussion Papers

- 2008/66. Jean-François CAULIER, Ana MAULEON and Vincent VANNETELBOSCH. Contractually stable networks.
- 2008/67. Jean J. GABSZEWICZ, Filomena GARCIA, Joana PAIS and Joana RESENDE. On Gale and Shapley '*College admissions and stability of marriage*'.
- 2008/68. Axel GAUTIER and Anne YVRANDE-BILLON. Contract renewal as an incentive device. An application to the French urban public transport sector.
- 2008/69. Yuri YATSENKO and Natali HRITONENKO. Discrete-continuous analysis of optimal equipment replacement.
- 2008/70. Michel JOURNÉE, Yurii NESTEROV, Peter RICHTÁRIK and Rodolphe SEPULCHRE. Generalized power method for sparse principal component analysis.
- 2008/71. Toshihiro OKUBO and Pierre M. PICARD. Firms' location under taste and demand heterogeneity.
- 2008/72. Iwan MEIER and Jeroen V.K. ROMBOOTS. Style rotation and performance persistence of mutual funds.
- 2008/73. Shin-Huei WANG and Christian M. HAFNER. Estimating autocorrelations in the presence of deterministic trends.
- 2008/74. Yuri YATSENKO and Natali HRITONENKO. Technological breakthroughs and asset replacement.
- 2008/75. Julio DÁVILA. The taxation of capital returns in overlapping generations economies without financial assets.
- 2008/76. Giorgia OGGIONI and Yves SMEERS. Equilibrium models for the carbon leakage problem.
- 2008/77. Jean-François MERTENS and Anna RUBINCHIK. Intergenerational equity and the discount rate for cost-benefit analysis.
- 2008/78. Claire DUJARDIN and Florence GOFFETTE-NAGOT. Does public housing occupancy increase unemployment?
- 2008/79. Sandra PONCET, Walter STEINGRESS and Hylke VANDENBUSSCHE. Financial constraints in China: firm-level evidence.
- 2008/80. Jean GABSZEWICZ, Salome GVETADZE, Didier LAUSSEL and Patrice PIERETTI. Public goods' attractiveness and migrations.
- 2008/81. Karen CRABBE and Hylke VANDENBUSSCHE. Are your firm's taxes set in Warsaw? Spatial tax competition in Europe.
- 2008/82. Jean-Sébastien TANCREZ, Benoît ROLAND, Jean-Philippe CORDIER and Fouad RIANE. How stochasticity and emergencies disrupt the surgical schedule.
- 2008/83. Peter RICHTÁRIK. Approximate level method.
- 2008/84. Çağatay KAYI and Eve RAMAEKERS. Characterizations of Pareto-efficient, fair, and strategy-proof allocation rules in queueing problems.
- 2009/1. Carlo ROSA. Forecasting the direction of policy rate changes: The importance of ECB words.
- 2009/2. Sébastien LAURENT, Jeroen V.K. ROMBOOTS and Francesco VIOLANTE. Consistent ranking of multivariate volatility models.
- 2009/3. Dunia LÓPEZ-PINTADO and Juan D. MORENO-TERNERO. The principal's dilemma.
- 2009/4. Jacek B. KRAWCZYK and Oana-Silvia SEREA. A viability theory approach to a two-stage optimal control problem of technology adoption.
- 2009/5. Jean-François MERTENS and Anna RUBINCHIK. Regularity and stability of equilibria in an overlapping generations model with exogenous growth.
- 2009/6. Nicolas GILLIS and François GLINEUR. Using underapproximations for sparse nonnegative matrix factorization.
- 2009/7. Michel M. DENUIT, Louis EECKHOUDT and Mario MENEGATTI. Correlated risks, bivariate utility and optimal choices.
- 2009/8. Michel M. DENUIT, Louis EECKHOUDT and Mario MENEGATTI. Adding independent risks in an insurance portfolio: which shape for the insurers' preferences?

## Recent titles

### CORE Discussion Papers - continued

- 2009/9. Antoine BOMMIER and Stéphane ZUBER. The Pareto principle of optimal inequality.
- 2009/10. Raouf BOUCEKKINE, Jacek B. KRAWCZYK and Thomas VALLEE. Environmental negotiations as dynamic games: Why so selfish?
- 2009/11. Théophile T. AZOMAHOU, Raouf BOUCEKKINE and Phu NGUYEN-VAN. Promoting clean technologies under imperfect competition.
- 2009/12. Patrice PIERETTI and Skerdilajda ZANAJ. On tax competition, public goods provision and jurisdictions' size.
- 2009/13. Jeroen V.K. ROMBOUTS and Lars STENTOFT. Bayesian option pricing using mixed normal heteroskedasticity models.
- 2009/14. Gauthier de MAERE d'AERTRYCKE and Yves SMEERS. The valuation of power futures based on optimal dispatch.
- 2009/15. Thierry BRECHET, Tsvetomir TSACHEV and Vladimir M. VELIOV. Prices versus quantities in a vintage capital model.
- 2009/16. François VANDERBECK and Laurence A. WOLSEY. Reformulation and decomposition of integer programs.
- 2009/17. Marc FLEURBAEY, Erik SCHOKKAERT and Koen DECANCO. What good is happiness?
- 2009/18. David DE LA CROIX and Michel LUBRANO. The tradeoff between growth and redistribution: ELIE in an overlapping generations model.
- 2009/19. Thierry BRECHET and Fabien PRIEUR. Can education be good for both growth and the environment?
- 2009/20. Giacomo SBRANA and Andrea SILVESTRINI. What do we know about comparing aggregate and disaggregate forecasts?

### Books

- H. TULKENS (ed.) (2006), *Public goods, environmental externalities and fiscal competition*. New York, Springer-Verlag.
- V. GINSBURGH and D. THROSBY (eds.) (2006), *Handbook of the economics of art and culture*. Amsterdam, Elsevier.
- J. GABSZEWICZ (ed.) (2006), *La différenciation des produits*. Paris, La découverte.
- L. BAUWENS, W. POHLMEIER and D. VEREDAS (eds.) (2008), *High frequency financial econometrics: recent developments*. Heidelberg, Physica-Verlag.
- P. VAN HENTENRYCKE and L. WOLSEY (eds.) (2007), *Integration of AI and OR techniques in constraint programming for combinatorial optimization problems*. Berlin, Springer.
- P-P. COMBES, Th. MAYER and J-F. THISSE (eds.) (2008), *Economic geography: the integration of regions and nations*. Princeton, Princeton University Press.
- J. HINDRIKS (ed.) (2008), *Au-delà de Copernic: de la confusion au consensus ?* Brussels, Academic and Scientific Publishers.

### CORE Lecture Series

- C. GOURIÉROUX and A. MONFORT (1995), *Simulation Based Econometric Methods*.
- A. RUBINSTEIN (1996), *Lectures on Modeling Bounded Rationality*.
- J. RENEGAR (1999), *A Mathematical View of Interior-Point Methods in Convex Optimization*.
- B.D. BERNHEIM and M.D. WHINSTON (1999), *Anticompetitive Exclusion and Foreclosure Through Vertical Agreements*.
- D. BIENSTOCK (2001), *Potential function methods for approximately solving linear programming problems: theory and practice*.
- R. AMIR (2002), *Supermodularity and complementarity in economics*.
- R. WEISMANTEL (2006), *Lectures on mixed nonlinear programming*.