University of Kent

School of Economics Discussion Papers

# EuroMInd-C: a Disaggregate Monthly Indicator of Economic Activity for the Euro Area and member countries

Stefano Grassi, Tommaso Proietti, Cecilia Frale,
Massimiliano Marcellino and Gianluigi Mazzi

May 2014

KDPE 1406

# EuroMInd-C: a Disaggregate Monthly Indicator of Economic Activity for the Euro Area and member countries[1]


**Stefano Grassi**[2]
School of Economics, University of Kent and CREATES

**Tommaso Proietti**[3]
Università di Roma "Tor Vergata" and CREATES

**Cecilia Frale**[4]
Italian Ministry of the Economy and Finance

**Massimiliano Marcellino**[5]
European University Institute

**Gianluigi Mazzi**[6]
Eurostat



[1]We would like to thank the Editor, an Associate Editor and two anonymous Referees for helpful comments on a previous draft.



[2]School of Economics, University of Kent and CREATES. E-mail: S.Grassi@kent.ac.uk

[3]Corresponding author. Dipartimento di Economia e Finanza, Via Columbia 2, 00133 Roma, Italy. E-mail: tommaso.proietti@uniroma2.it.

[4]Ministry of the Economy and Finance, via XX Settembre 97, 00185 Roma, Italy. E-mail: cecilia.frale@tesoro.it.

[5]Bocconi University, IGIER and CEPR. E-mail: massimiliano.marcellino@unibocconi.it

[6]Eurostat, Luxembourg, E-mail: Gianluigi.Mazzi@ec.europa.eu

**Abstract**

The paper deals with the estimation of monthly indicators of economic activity for the Euro area and its largest member countries that possess the following attributes: relevance, representativeness and timeliness. Relevance is obtained by referring our monthly indicators to gross domestic product at chained volumes, the most important measure of the level of economic activity. Representativeness is achieved by entertaining a very large number of (timely) time series on monthly indicators relating to the level of economic activity, providing a more or less complete coverage. The indicators are modelled with a large scale parametric factor model. We discuss its specification and provide details on the statistical treatment. Computational efficiency is crucial for estimating a large scale parametric factor model of the dimension considered in our application (considering about 170 series). To achieve it we apply state of the art state space methods that can handle temporal aggregation, and any pattern of missing values.

*Keywords:* Index of coincident indicators. Temporal Disaggregation. Multivariate State Space Models. Dynamic factor Models. Quarterly National accounts.

*J.E.L. Classification:* E32, E37, C53

## 1 Introduction

Large scale factor models aim at extracting the main economic signals from a very large number of time series. The underlying idea is that the comovements among economic time series can be ascribed to a small number of common factors. The two most prominent areas of applications deal with the construction of synthetic indicators, such as coincident indicators of real economic activity (**?**, **?**, **?**) and core inflation (**?**), and forecasting macroeconomic variables (**?**, **?**). See **?** for a survey.

Most of the factor based studies use non-parametric techniques, which make estimation feasible even with a large dataset. Parametric large scale factor models have been recently entertained by **?**, **?**, and **?**. Most of these studies rely on the Kalman filter and smoothing algorithms to be able to handle mixed data frequencies and generic patterns of missing observations.

In this paper we present a generalization of EuroMInd, see **?**, called EuroMInd-C, which allows for the simultaneous calculation of monthly indicators of the economic activity for the Euro Area and its largest member states. EuroMInd-C is based on a parametric large scale factor model handling a very large set of time series, with mixed frequency, and subject to missing values, due eminently to ragged-edged data structure, which in turn depends on the publication and dissemination schedule of the data producers.

The construction of EuroMInd-C is based on more than 100 monthly time series and 55 quarterly national accounts series. The latter concern the decomposition of gross domestic product according to the output and expenditure approaches, for the Euro area as a whole and for the four largest countries (Germany, France, Italy and Spain). Indeed, a distinctive trait of our approach is the consideration of the quarterly national accounts estimates.[1] As a matter of fact, our dataset features a total of 11 GDP components, seven of which are drawn from the following decomposition of GDP from the output side:

| *Label* | *Value added of branch* | |
|---|---|---|
| A–B | Agriculture, hunting, forestry and fishing | + |
| C–D–E | Industry, incl. Energy | + |
| F | Construction | + |
| G–H–I | Trade, transport and communication services | + |
| J–K | Financial services and business activities | + |
| L–P | Other services | = |
| | Total Gross Value Added | + |
| TlS | Taxes less subsidies on products | = |
| | *GDP at market prices* | |

The breakdown of total GDP from the expenditure side is the following:

[1]Typically, the information set used for the estimation of factor models is strongly unbalanced towards the series collected from the supply side of the economy, that is establishments surveys (e.g., industrial production and turnover, retail sales, financial statistics), and from administrative records (e.g. building permits and car registration). On the contrary, important information from other institutional units and economic agents, namely households, is missed, just because the underlying measurement process is more complex, with the consequence that the information becomes available with longer delays. Notable examples are the Labor Force Survey and the Consumer Expenditure Survey, which are carried out by the Euro area member states and represent essential sources for the labor market and consumption. On the other hand, this information is incorporated in the national accounts estimates of gross domestic product (GDP) and its components.

| *Label* | *Component* | |
|---|---|---|
| FCE | Final consumption expenditure | + |
| GCF | Gross capital formation | + |
| EXP | Exports of goods and services | - |
| IMP | Imports of goods and services | = |
| | *GDP at market prices* | |

From the methodological point of view, we propose a large scale factor model such that the comovements among the series are synthesized by a set of sixteen common factors, representing the Euro area common trend, 4 country specific factors and 11 components specific factors. The statistical treatment is based on likelihood inferences for a suitable state space model that is able to accommodate temporal aggregation and any pattern of missing data that arise, also taking into consideration the ragged-edge structure of the dataset and the release schedule of the economic indicators. Computationally efficient algorithms have been implemented to avoid the curse of dimensionality and to address specific issues related to the available data. For instance, for the extraction of the common factors we adopt a reduction technique that decreases the dimension of the model to the number of common factors.

Our model is an extension of the single index model originally proposed by **?**, and generalized by **?**, jointly modelling quarterly GDP and monthly indicators, see also **?** and **?**. Our model also extends **?**, who link together in a state space framework a set of parametric mixed frequency factor models, one for each GDP demand and supply component of Euro area GDP, providing an indicator called EuroMInd. Unlike the original EuroMInd, here the estimation of the monthly indicators of GDP and its components is not carried out componentwise, but simultaneously. The model in fact includes all the components of the breakdown of GDP, rather than performing separate estimations for each different component (sector or expenditure type). Moreover, not only Euro area, but also country information is used, and country indicators of economic conditions are also produced. Compared to **?**,where a large scale factor model for the Euro area is estimated with an ex post identification of the factors, in this model factors are identified a priori and ascribed to a particular effect (country specific factors, factors pertaining to a specific GDP component).

The model is specified at the monthly frequency and for the logarithms of the original series. The presence of national accounts aggregates, observed only quarterly, imposes constraints related to temporal aggregation. Furthermore, due to the logarithmic specification, a nonlinear smoothing algorithm has to be applied for the disaggregation of the quarterly national accounts aggregates. The corresponding state space representation is modified according to the observational constraints to handle the temporal aggregation and the ragged-edged data structure and missing values at the beginning of the sample period.

Maximum likelihood estimation is carried out using the EM algorithm. As the number of time series that are handled simultaneously is very large (around 170), we implement efficient methods of inference, focusing on the treatment of missing values and maximum likelihood estimation, as documented in the next sections. Furthermore, the reduction technique proposed by **?** and **?** is applied to the temporal disaggregation case. This further extends **?** and **?**.

The plan of the paper is the following: section 2 illustrates the specification of the dynamic factor model and in particular the definition and the identification of the common factors; section 3 describes the state space representation of the model; the statistical treatment of temporal aggregation is dealt with in section 4, while section 5 considers the missing values problem; section 6 discusses estimation of common factors and section 7 of the other model parameters. The empirical application starts in section 8, which provides a description of the dataset; section 9 presents the estimation results and the resulting EuroMInd-C estimates. Section 10 provides a comparative assessment of EuroMInd-C in terms of its predictive accuracy for quarterly GDP and its components, by means of a recursive forecasting experiment. Finally, section 11 summarizes our main findings and concludes.

## 2 Model Specification

This section outlines the basic structure of the dynamic factor model used to estimate EuroMInd-C. The specification can be considered as an extension of the single index model proposed by **?**, introducing multiple indices relating to the countries and the sectors.

For the purposes of illustrating the model we assume that a complete set of monthly time series is available. Let $c = 1, \ldots, C$, index a country, with $c = 1$ indexing the Euro Area, and let $s = 1, \ldots, S$ denote a particular component (e.g. the construction sector, or an expenditure component such as exports); in our particular application, $C = 5$ and $S = 11$. Denoting by $y_{cs,t}$ the monthly series (usually transformed into logarithms), $t = 0, \ldots, n$, for an indicator referring to country $c$ and component $s$, we write:

$$y_{cs,t} = \theta_{cs,0}\mu_t + \theta_{cs,1}\mu_{ct} + \theta_{cs,2}\mu_{st} + \mu^*_{cs,t}, \tag{1}$$

where $\mu_t$ is the common Euro area factor, specified as an ARIMA(1,1,0) process:

$$\Delta\mu_t = \phi\Delta\mu_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, 1).$$

The factor $\mu_{ct}$ is specific to country $c$, whereas $\mu_{st}$ is a factor specific to sector $s$. Both are specified as ARIMA(1,1,0) processes:

$$\begin{array}{rcll} \Delta\mu_{ct} & = & \phi_c\Delta\mu_{c,t-1} + \eta_{ct}, & \eta_{ct} \sim \text{NID}(0, 1), \\ \Delta\mu_{st} & = & \phi_s\Delta\mu_{s,t-1} + \eta_{st}, & \eta_{st} \sim \text{NID}(0, 1). \end{array} \tag{2}$$

Finally, the idiosyncratic component is formulated as follows:

$$\Delta\mu^*_{cs,t} = (1 - \psi_{cs})\delta_{cs} + \psi_{cs}\Delta\mu^*_{cs,t-1} + \epsilon_{cs,t}, \quad \epsilon_{cs,t} \sim \text{NID}(0, \sigma^2_{\epsilon,cs}). \tag{3}$$

The model assumes that each individual time series is made stationary by the transformation $\Delta y_{cs,t} = y_{cs,t} - y_{cs,t-1}$ for $t = 1, \ldots, n$. If $y_{cs,t}$ is a survey series, then the specification refers to the cumulated values of the series (which are by construction integrated of order 1). Notice that the drift $\delta_{cs}$ is assigned to the idiosyncratic component and thus it is a series specific feature.

For the Euro area as a whole ($c = 1$), the $j$-th series results from the aggregation of the series for all the member countries. Thus, it will load on the global factor, the sector specific factor, as well as the country specific factors of the individual member states under evaluation (the four largest economies). The specific factors of the countries not included in the model will contribute to the idiosyncratic component of the Euro area. However, the latter can be correlated with the idiosyncratic components of the countries included in the model. The equation for the Euro area series is thus specified as follows:

$$y_{1s,t} = \theta_{1s,0}\mu_t + \sum_{c=2}^{C}\theta_{1c,1}\mu_{ct} + \theta_{1s,2}\mu_{st} + \mu^*_{1s,t}. \tag{4}$$

In general, to show the identification restrictions imposed on the loadings matrix, we report an example for 5 countries (EA, G, S, F and I, respectively the Euro area, Germany, Spain, France and Italy) and 2 sectors ($a$ and $b$), (in our study we have a total of 5 countries and 11 sectors). The value added series for the two sectors are listed first, and three monthly indicators are available: one refers to both sectors ($Ind1_{a,b}$), while the other two refer respectively to the first ($Ind2_a$) and second sector ($Ind3_b$).

$$\left(\begin{array}{c} EA_a \\ G_a \\ S_a \\ F_a \\ I_a \\ EA_b \\ G_b \\ S_b \\ F_b \\ I_b \\ \hline EA-Ind1_{a,b} \\ G-Ind1_{a,b} \\ S-Ind1_{a,b} \\ F-Ind1_{a,b} \\ I-Ind1_{a,b} \\ \hline EA-Ind2_a \\ G-Ind2_a \\ S-Ind2_a \\ F-Ind2_a \\ I-Ind2_a \\ \hline EA-Ind3_b \\ G-Ind3_b \\ S-Ind3_b \\ F-Ind3_b \\ I-Ind3_b \end{array}\right) = \left(\begin{array}{ccccccc} x & x & x & x & x & x & 0 \\ x & x & 0 & 0 & 0 & x & 0 \\ x & 0 & x & 0 & 0 & x & 0 \\ x & 0 & 0 & x & 0 & x & 0 \\ x & 0 & 0 & 0 & x & x & 0 \\ x & x & x & x & x & 0 & x \\ x & x & 0 & 0 & 0 & 0 & x \\ x & 0 & x & 0 & 0 & 0 & x \\ x & 0 & 0 & x & 0 & 0 & x \\ x & 0 & 0 & 0 & x & 0 & x \\ \hline x & x & x & x & x & x & x \\ x & x & 0 & 0 & 0 & x & x \\ x & 0 & x & 0 & 0 & x & x \\ x & 0 & 0 & x & 0 & x & x \\ x & 0 & 0 & 0 & x & x & x \\ \hline x & x & x & x & x & x & 0 \\ x & x & 0 & 0 & 0 & x & 0 \\ x & 0 & x & 0 & 0 & x & 0 \\ x & 0 & 0 & x & 0 & x & 0 \\ x & 0 & 0 & 0 & x & x & 0 \\ \hline x & x & x & x & x & 0 & x \\ x & x & 0 & 0 & 0 & 0 & x \\ x & 0 & x & 0 & 0 & 0 & x \\ x & 0 & 0 & x & 0 & 0 & x \\ x & 0 & 0 & 0 & x & 0 & x \end{array}\right) \begin{pmatrix} \mu_t \\ \mu_{G,t} \\ \mu_{S,t} \\ \mu_{F,t} \\ \mu_{I,t} \\ \mu_{a,t} \\ \mu_{b,t} \end{pmatrix} + \boldsymbol{\mu}_t^* \qquad (5)$$

Here, $\boldsymbol{\mu}_t^*$ is a vector of idiosyncratic components. The first factor, $\mu_t$, is interpreted as the Euro area common factor. The factors $\mu_{G,t}, \mu_{S,t}, \mu_{F,t}, \mu_{I,t}$ are country specific, and finally the factors $\mu_{a,t}$ and $\mu_{b,t}$ are sector specific. All the time series load on the Euro area factor and on the factor of the component to which they refer. Hence, for instance, the German value added for sector $b$, $G_b$, as well as $G - Ind3_b$, load on the common Euro area factor, the German factor and the industry $b$ factor. The German and French sector $a$ indicators are related only via the sector specific factor $\mu_{a,t}$ and the common Euro area factor $\mu_t$. All the series are related via the latter.

The Euro area is treated as a separate country - this treatment is similar to that assigned to the U.S. in **?** and **?**. Another possibility is modelling the rest of the Euro area as an additional entity and obtaining the Euro area aggregates by summing up the five series pertaining to the 4 countries and the residual unit. While this is technically feasible, the main issue is finding a set of representative indicators for the Euro area remaining countries.

Finally, the model is specified at the monthly frequency. However, the national accounts series, and, more generally, the quarterly series are not observed at that frequency. Sections 3–7 deal with more technical aspects of model specification and estimation and can be skipped by the reader uninterested in those details.

## 3 The Dynamic Factor Model

First, we present the specification of the model as a dynamic factor model in the first log-differences of the series assuming that a complete data set of monthly observations is available at times $t = 1, \dots, n$.

We start by denoting $\mathbf{Y}_t, t = 0, \dots, n$, the stack of the $N$ individual monthly time series $Y_{cs,t}, c = 1, \dots, C, s = 1, \dots, S$, in their original scale of measurement. Further, let us denote by $\mathbf{y}_t$ a transformation of $\mathbf{Y}_t$. We assume throughout that $y_{cs,t} = \ln Y_{cs,t}$ (log transformation) for all the series measured on a ratio scale, such as the components of GDP, the index of industrial production, etc. For the survey variables (which can take on negative values) we always set $y_{cs,t} = Y_{cs,t}$ (no transformation).

For the sake of notation, henceforth we shall adopt a single index $i$ to refer to the $i$-th element of the $N \times 1$ vector $\mathbf{Y}_t$. Stacking the individual equations for $\Delta y_{it}$, the model is formulated as follows:

$$\begin{array}{lcl} \Delta \mathbf{y}_t & = & \boldsymbol{\delta} + \boldsymbol{\Theta} \mathbf{f}_t + \mathbf{u}_t, \quad t = 1, \dots, n, \\ \mathbf{f}_t & = & \boldsymbol{\Phi} \mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \\ \mathbf{u}_t & = & \boldsymbol{\Psi} \mathbf{u}_{t-1} + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \text{NID}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon), \end{array} \tag{6}$$

where $\boldsymbol{\delta} = \{\delta_{cs}\}$, contains the drifts in (3), $\mathbf{f}_t$ is a $K \times 1$ vector of stationary common factors, with zero mean, stacking the first differences $[\Delta\mu_t, \Delta\mu_{ct}, c = 2, \dots, C, \Delta\mu_{st}, s = 1, \dots, S]$. The total number of common factors is denoted by $K = C + S$, $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_N]'$ is the $N \times K$ matrix of factor loadings,

$$\boldsymbol{\Phi} = \text{diag}\{\phi_1, \dots, \phi_k, \dots, \phi_K\},$$

$\boldsymbol{\Sigma}_\eta = \mathbf{I}_K$, where $\mathbf{I}_K$ is the identity matrix of order $K$. Hence, $f_{kt} = \phi_k f_{k,t-1} + \eta_{kt}, \quad \eta_{kt} \sim \text{NID}(0, 1)$. Finally, $\mathbf{u}_t$ is the $N \times 1$ vector of stationary idiosyncratic components, which arises from the decomposition

$$\Delta \boldsymbol{\mu}_t^* = \boldsymbol{\delta} + \mathbf{u}_t.$$

Hence, the components of the vector $\mathbf{u}_t$ follow scalar independent AR(1) processes, $u_{it} = \psi_i u_{i,t-1} + \epsilon_{it}, \quad \epsilon_{it} \sim \text{NID}(0, \sigma_i^2)$, according to (3), so that

$$\boldsymbol{\Psi} = \text{diag}\{\psi_1, \dots, \psi_i, \dots, \psi_N\}, \quad \boldsymbol{\Sigma}_\epsilon = \text{diag}\{\sigma_1^2, \dots, \sigma_i^2, \dots, \sigma_N^2\}.$$

For later use we also define the covariance matrix of the factors, $\boldsymbol{\Sigma}_f$, which is given by $\boldsymbol{\Sigma}_f = \text{diag}\{(1 - \phi_k^2)^{-1}\}, \quad k = 1, \dots, K$.

Model (6) is identified by imposing exclusion restrictions on the matrix $\boldsymbol{\Theta}$, that are actually overidentifying the model. The number of common factors is fixed to $K = 16$, since we postulate the presence of a common global factor, and country and sector specific factors. The assumption that the common factors and the idiosyncratic components are uncorrelated, $\text{E}(\boldsymbol{\epsilon}_t \boldsymbol{\eta}_t') = \mathbf{0}$, completes the specification.

Let us define the parameter vector $\boldsymbol{\Xi} = [\boldsymbol{\vartheta}', \boldsymbol{\delta}', \boldsymbol{\phi}', \boldsymbol{\psi}', \boldsymbol{\varsigma}_\epsilon']'$, where $\boldsymbol{\vartheta}$ stacks the nonzero elements of the loadings matrix, $\boldsymbol{\phi} = [\phi_1, \dots, \phi_k]'$, $\boldsymbol{\Phi} = \text{diag}(\boldsymbol{\phi})$, $\boldsymbol{\psi} = [\psi_1, \dots, \psi_N]'$, $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\psi})$, and $\boldsymbol{\varsigma}_\epsilon = [\sigma_1^2, \ldots, \sigma_N^2]'$, $\boldsymbol{\Sigma}_\epsilon = \text{diag}(\boldsymbol{\varsigma}_\epsilon)$. Using the above restrictions and $\sigma_i^2 > 0, \forall i$, the vector $\boldsymbol{\Xi}$ is globally identified.

If $\Delta \mathbf{y}_t$ were fully observed, inference concerning model (6) would be straightforward, although computationally demanding. For small $N$ (less than 10, say) the parameters can be estimated by maximum likelihood, where the likelihood is evaluated by the Kalman filter (KF) using a numerical quasi-Newton method. With large $N$, the high dimensionality of $\boldsymbol{\Xi}$ prevents maximising the likelihood via gradient based methods. A computationally viable alternative is to use the Expectation- Maximization (EM) algorithm of **?**. See **?** and **?**.

Another issue is that, according to (6), $\mathbf{u}_t$ has a first order Markovian representation. If the model is written in state space form, the state vector should feature $N$ elements to account for the idiosyncratic factors. Filtering and smoothing then becomes infeasible if $N$ has the dimension considered in this paper.

A transformation of the model that will be useful in the sequel is the quasi-difference form:

$$\Delta \mathbf{y}_t = \boldsymbol{\Psi} \Delta \mathbf{y}_{t-1} + \boldsymbol{\delta} - \boldsymbol{\Psi}\boldsymbol{\delta} + \boldsymbol{\Theta}\mathbf{f}_t - \boldsymbol{\Psi}\boldsymbol{\Theta}\mathbf{f}_{t-1} + \boldsymbol{\epsilon}_t. \tag{7}$$

The main virtue of this representation is that the idiosyncratic VAR(1) process $\mathbf{u}_t$ has been removed, and only the idiosyncratic disturbances $\boldsymbol{\epsilon}_t$ are present, which are however serially uncorrelated. The dimensionality of the unobserved components with a dynamic structure depends solely on the number of common factors and thus it is much smaller than $N$. The inclusion of $\mathbf{u}_t$ in the state vector will be necessary only for setting up the initial conditions. The reduction is achieved at the cost of conditioning upon $\Delta \mathbf{y}_{t-1}$; this will have certain consequences on the way we handle missing values, as we will see later.

New computationally efficient methods of inference need to be applied to estimate the factors and the parameters. First and foremost the specification of the model needs to take into consideration temporal aggregation and the presence of missing values and ragged edge structure of the data. The final state space representation will now be obtained in two stages:

- The original state space form suitable for (7) is modified to take into consideration temporal aggregation: the observational constraints are sequentially enforced by solving a nonlinear smoothing problem using a linearised Gaussian model, in section 4.

- Missing values are entertained by a time varying state space form, in section 5.

# 4 Temporal Aggregation

Our data set contains $N_1 = 55$ quarterly national accounts (NA) aggregates that are subject to temporal aggregation: the 11 GDP components for the 5 countries. These series represent a set of constraints for the monthly estimates. The remaining $N_2 = N - N_1$ time series are a set of monthly coincident indicators referring to the Euro area or to the individual countries, (e.g. index of industrial production, turnover and so forth). They may be subject to isolated missing values and to ragged edge structure, depending on their timeliness and publishing schedule. Accordingly, we partition the vector $\mathbf{Y}_t$ in two blocks, $\mathbf{Y}_t = [\mathbf{Y}_{1t}', \mathbf{Y}_{2t}']'$. Correspondingly, also $\mathbf{y}_t$ and $\Delta \mathbf{y}_t$ will be partitioned in two blocks, the first referring to the quarterly national accounts series and the second to the monthly coincident indicators.

The monthly series $\mathbf{Y}_{1t}$, and thus $\mathbf{y}_{it} = \ln \mathbf{Y}_{1t}, t = 0, 1, \ldots, n$, and $\Delta \mathbf{y}_{it}$, for $i = 1, \ldots, N_1$ are not observed for the first block of $N_1$ time series belonging to the national accounts. Only

*Table 1: Nonlinear temporal aggregation*

| $t$ | Unobserved $Y_{it}$ | Logarithms | Cumulator | Availability of $Y^c_{it}$ |
|---|---|---|---|---|
| 0 | $Y_{i0}$ | $y_{i0}$ | $Y^c_{i0} = Y_{i0} = \exp(y_{i0})$ | missing |
| 1 | $Y_{i1}$ | $y_{i1}$ | $Y^c_{i1} = Y_{i0} + Y_{i1} = \exp(y_{i0}) + \exp(y_{i1})$ | missing |
| 2 | $Y_{i2}$ | $y_{i2}$ | $Y^c_{i2} = Y_{i0} + Y_{i1} + Y_{i2} = \exp(y_{i0}) + \exp(y_{i1}) + \exp(y_{i2})$ | observed |
| 3 | $Y_{i3}$ | $y_{i3}$ | $Y^c_{i3} = Y_{i3} = \exp(y_{i3})$ | missing |
| 4 | $Y_{i4}$ | $y_{i4}$ | $Y^c_{i4} = Y_{i3} + Y_{i4} = \exp(y_{i3}) + \exp(y_{i4})$ | missing |
| 5 | $Y_{i5}$ | $y_{i5}$ | $Y^c_{i5} = Y_{i3} + Y_{i4} + Y_{i5} = \exp(y_{i3}) + \exp(y_{i4}) + \exp(y_{i5})$ | observed |
| 6 | $Y_{i6}$ | $y_{i6}$ | $Y^c_{i6} = Y_{i6} = \exp(y_{i6})$ | missing |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

the quarterly totals are observed, instead, which are supposed to constraint the monthly estimates. This observational constraint can be incorporated in the specification of the model and later we will derive the relevant state space representation of the model. The latter has two fundamental uses: for a given parameter configuration it provides the optimal (posterior mode) estimates of the disaggregate monthly time series, the factors and the idiosyncratic components. It is also instrumental in performing the expectation step of the EM algorithm for the maximum likelihood estimation of the parameters. These evaluations are done via a nonlinear smoothing algorithm that is discussed in section 4.1. For $Y_{it}, i = 1, \dots, N_1$, subject to temporal aggregation, we observe the quarterly totals:

$$Y_{i\tau} = \sum_{j=1}^{3} Y_{i,3\tau-j}, \quad \tau = 1, 2, \dots, [(n+1)/3], \tag{8}$$

where $[\cdot]$ is the integer part of the argument.

For the statistical treatment it is useful to convert temporal aggregation into a systematic sampling problem; this can be done by constructing a cumulator variable, generated by the recursive formula (see **?**, and **?**):

$$\begin{aligned} Y^c_{it} &= \rho_t Y^c_{i,t-1} + Y_{it}, \quad t = 0, \dots, n \\ &= \rho_t Y^c_{i,t-1} + \exp(y_{it}), \end{aligned} \tag{9}$$

where $\rho_t$ is the cumulator coefficient, equal to zero for $t$ corresponding to the first month in the quarter and 1 otherwise, i.e.,

$$\rho_t = \begin{cases} 0 & t = 3(\tau - 1), \quad \tau = 1, \dots, [(n+1)/3] \\ 1 & \text{otherwise.} \end{cases}$$

Only a systematic sample of the cumulator variable $Y^c_{it}$ is available; in particular, if the sample period starts with the first month of the quarter at $t = 0$, the observed end of quarter values occur at times $t = 3\tau - 1, \tau = 1, 2, \dots, [(n+1)/3]$. In our particular case, $Y^c_{i0} = \exp(y_{i0})$, $Y^c_{i1} = \exp(y_{i0}) + \exp(y_{i1})$, $Y^c_{i2} = \exp(y_{i0}) + \exp(y_{i1}) + \exp(y_{i2})$, $Y^c_{i3} = \exp(y_{i3})$, $Y^c_{i4} = \exp(y_{i3}) + \exp(y_{i4})$, $Y^c_{i5} = \exp(y_{i3}) + \exp(y_{i4}) + \exp(y_{i5})$, $\dots$. As illustrated in table 1, only the values $Y^c_{i2}, Y^c_{i5}, \dots$ are observed, while the intermediate ones will be missing. Temporal aggregation yields a nonlinear observational constraint since the quarterly totals are a nonlinear function of the underlying monthly values $y_{it}$.

## 4.1 An Approximating Linear Gaussian Model

The estimation of the common factors $\mathbf{f}_t$, and the missing values $\mathbf{y}_{it}, i = 1, \dots, N_1$, conditional on $\boldsymbol{\Xi}$ and the available information, which consists of $Y_{it}^c, i = 1, \dots, N_1, t = 3\tau - 1, \tau = 1, 2, \dots, [(n+1)/3]$, for the quarterly time series and $y_{it}$ for $i = N_1 + 1, \dots, N$, is a nonlinear smoothing problem that can be solved by iterating the Kalman filter and smoother (KFS) adapted to a sequentially linearized state space model, see **?**.

Let us partition the vectors $\mathbf{Y}_t = [\mathbf{Y}_{1t}', \mathbf{Y}_{2t}']'$, $\mathbf{y}_t = [\mathbf{y}_{1t}', \mathbf{y}_{2t}']'$, such that $Y_{it} = \exp(y_{it})$ for all the series measured on a ratio scale (all the series except the business survey variables), $\Delta\mathbf{y}_t = [\Delta\mathbf{y}_{1t}', \Delta\mathbf{y}_{2t}']'$, $\boldsymbol{\delta} = [\boldsymbol{\delta}_1, \boldsymbol{\delta}_2]'$, $\mathbf{u}_t = [\mathbf{u}_{1t}, \mathbf{u}_{2t}]'$ and the matrices $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1', \boldsymbol{\Theta}_2']'$, $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2)$, $\boldsymbol{\Sigma}_\epsilon = \text{diag}(\boldsymbol{\Sigma}_{\epsilon 1}, \boldsymbol{\Sigma}_{\epsilon 2})$, where the subscript 1 indexes the national accounts series, and the dimension of the blocks are respectively $N_1$ and $N_2$. Also, define $\mathbf{u} = [\mathbf{u}_1', \dots, \mathbf{u}_n']'$, and $\mathbf{f} = [\mathbf{f}_1', \dots, \mathbf{f}_n']'$. Although $\Delta\mathbf{y}_{2t}$ is available, $\Delta y_{1t}, t = 1, \dots, n$, is not, but we observe a systematic sample of

$$\begin{aligned} \mathbf{Y}_{1t}^c &= \rho_t \mathbf{Y}_{1,t-1}^c + \mathbf{Y}_{1t}, \\ &= \rho_t \mathbf{Y}_{1,t-1}^c + \exp(\mathbf{y}_{1t}). \end{aligned}$$

To obtain the values $\mathbf{f}$ and $\mathbf{u}$ that maximize the posterior density $g(\mathbf{f}, \mathbf{u}|\mathbf{x}; \boldsymbol{\Xi})$, given the available data,we linearize $\mathbf{Y}_{1t}^c = \rho_t \mathbf{Y}_{1,t-1}^c + \exp(\mathbf{y}_{1t})$ around a trial value $\mathbf{y}_{1t}^*$, by a first order Taylor series expansion. This yields a linear state space model and the corresponding KFS provides a new trial value for the disaggregate series. This sequence of linearisations is iterated until convergence and the end result is a set of disaggregate monthly estimates $\mathbf{Y}_1$ and factor scores $\mathbf{f}$ which incorporate the temporal aggregation constraints. As a by-product, disaggregate (monthly) estimates of the missing values $\mathbf{y}_{1t}$ and thus $\mathbf{Y}_{it} = \exp(\mathbf{y}_{1t})$ will be made available. Using the trial value $\mathbf{y}_{1t}^* = \log(\mathbf{Y}_{1t}^*)$, the linearisation operates as follows:

$$\mathbf{Y}_{1t}^c = \rho_t \mathbf{Y}_{1,t-1}^c + \exp(\mathbf{y}_{1t}^*) + \mathbf{U}_{1t}^*(\mathbf{y}_{1t} - \mathbf{y}_{1t}^*),$$

where the $N_1 \times N_1$ matrix $\mathbf{U}_{1t}^*$ is a diagonal matrix with the derivatives of the exponential transformation on the main diagonal, $\mathbf{U}_{1t}^* = \text{diag}\left(\exp(\mathbf{y}_{1t}^*)\right)$. Writing

$$\begin{aligned} \mathbf{y}_{1t} &= \mathbf{y}_{1,t-1} + \Delta\mathbf{y}_{1t} \\ &= \mathbf{y}_{1,t-2} + \Delta\mathbf{y}_{1,t-1} + \Delta\mathbf{y}_{1t}, \end{aligned}$$

and replacing $\Delta\mathbf{y}_{1t} = \boldsymbol{\Psi}_1 \Delta\mathbf{y}_{1t-1} + (\mathbf{I} - \boldsymbol{\Psi}_1)\boldsymbol{\delta}_1 + \boldsymbol{\Theta}_1 \mathbf{f}_t - \boldsymbol{\Psi}_1 \boldsymbol{\Theta}_1 \mathbf{f}_{t-1} + \boldsymbol{\epsilon}_{1t}$, yields

$$\mathbf{Y}_{1t}^c = \rho_t \mathbf{Y}_{1t-1}^c + \tilde{\mathbf{d}}_t^* + \mathbf{U}_{1t}^*[\mathbf{y}_{1t-2} + (\mathbf{I} + \boldsymbol{\Psi}_1)\Delta\mathbf{y}_{1t-1} + (\boldsymbol{\Theta}_1 \boldsymbol{\Phi} - \boldsymbol{\Psi}_1 \boldsymbol{\Theta}_1)\mathbf{f}_{t-1} + \boldsymbol{\epsilon}_{1t} + \boldsymbol{\Theta}_1 \boldsymbol{\eta}_t] \tag{10}$$

where $\tilde{\mathbf{d}}_t^* = \exp(\mathbf{y}_{1t}^*) - \mathbf{U}_{1t}^* \mathbf{y}_{1t} + \mathbf{U}_{1t}^*(\mathbf{I} - \boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1$. Equation (10) expresses $\mathbf{Y}_{1t}^c$ as a time-varying linear combination of $\mathbf{Y}_{1,t-1}^c, \mathbf{y}_{1,t-2}, \Delta\mathbf{y}_{1,t-1}, \mathbf{f}_{t-1}$, which will constitute the elements of the state vector at time $t-1$, denoted $\boldsymbol{\alpha}_{t-1}$, and of the disturbances of the factor model.

## 4.2 A convenient state space formulation under temporal aggregation constraints

The state space representation is conveniently formulated for the vector $\mathbf{y}_t^\dagger$, where, for $t \geq 1$, and if no element of $\mathbf{y}_{2t}$ is missing,

$$\mathbf{y}_t^\dagger = \begin{bmatrix} \mathbf{Y}_{1t}^c \\ \Delta\mathbf{y}_{2t} \end{bmatrix}, \quad t = 1, 2, \dots, n,$$

whereas for $t = 0$, $\mathbf{y}_0^\dagger = \mathbf{Y}_{10}^c$. The length of the observation vector varies with time and will be denoted by $N_t$. When one or more elements of $\mathbf{y}_{2t}$ are missing, the measurement equation will be modified suitably, as it will be discussed later.

The state vector for $t \geq 1$, when $\Delta\mathbf{y}_{2t}$ is fully observed, is defined as follows:

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \mathbf{Y}_{1,t}^c \\ \mathbf{y}_{1,t-1} \\ \Delta\mathbf{y}_{1t} \\ \Delta\mathbf{y}_{1,t-1} \\ \mathbf{f}_t \\ \mathbf{f}_{t-1} \end{bmatrix}.$$

The length of the state vector will also vary according to $t$ and to the presence of missing values.

The measurement equation is formulated as follows:

$$\begin{bmatrix} \mathbf{Y}_{1t}^c \\ \Delta\mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_2 & -\boldsymbol{\Psi}_2\boldsymbol{\Theta}_2 \end{bmatrix} \boldsymbol{\alpha}_t + \begin{bmatrix} \mathbf{0} \\ (\mathbf{I} - \boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 + \boldsymbol{\Psi}_2\Delta\mathbf{y}_{2t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}. \tag{11}$$

The transition equation is:

$$\begin{bmatrix} \mathbf{Y}_{1t}^c \\ \mathbf{y}_{1t-1} \\ \Delta\mathbf{y}_{1t} \\ \Delta\mathbf{y}_{1t-1} \\ \mathbf{f}_t \\ \mathbf{f}_{t-1} \end{bmatrix} = \begin{bmatrix} \rho_t\mathbf{I}_{N1} & \mathbf{U}_{1t}^* & \mathbf{U}_{1t}^*(\mathbf{I} + \boldsymbol{\Psi}_1) & \mathbf{0} & \mathbf{U}_{1t}^*[\boldsymbol{\Theta}_1\boldsymbol{\Phi} - \boldsymbol{\Psi}_1\boldsymbol{\Theta}_1] & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N1} & \mathbf{I}_{N1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{N_1} & \mathbf{0} & [\boldsymbol{\Theta}_1\boldsymbol{\Phi} - \boldsymbol{\Psi}_1\boldsymbol{\Theta}_1] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{1t-1}^c \\ \mathbf{y}_{1t-2} \\ \Delta\mathbf{y}_{1t-1} \\ \Delta\mathbf{y}_{1t-2} \\ \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \end{bmatrix}$$
$$+ \begin{bmatrix} \mathbf{Y}_{1t}^* - \mathbf{U}_{1t}^*\mathbf{y}_{1t}^* + \mathbf{U}_{1t}^*(\mathbf{I} - \boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1 \\ \mathbf{0} \\ (\mathbf{I} - \boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{U}_{1t}\boldsymbol{\Theta}_1 & \mathbf{U}_{1t}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_1 & \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}. \tag{12}$$

The treatment of initial conditions is rather technical and it is thus relegated to the Appendix C.

# 5 The treatment of missing values

Expressing the model in quasi differences as in (7) offers the advantage of dropping the $N$ dimensional vector $\mathbf{u}_t$ from the unobserved states. However, it introduces a new difficulty when the monthly indicator series $\mathbf{y}_{2t}$, or any element thereof, are missing at time $t$. As a matter of fact, a missing value at time $t$ implies not only that the pair $(\Delta\mathbf{y}_t, \Delta\mathbf{y}_{t+1})$ are subject to missing values, but also that the conditioning elements on the right hand side of the equation (7) are missing.

To overcome this difficulty the state space formulation needs to be adapted by augmenting the state vector by the minimal set of components needed to preserve the dynamic structure of the system and its Markovian structure. Our treatment is based on an adaptation of the approach recently proposed by **?**. The main advantages of this formulation is the treatment of arbitrary patterns of missing values.

We need to discriminate between the following situations: in the case when both $\mathbf{y}_t$ and $\mathbf{y}_{t-1}$ are observed, the state space formulation considered in the previous section applies. The state space form has to be modified in case of missing values. The main three cases of interest are:

1. Case 1: $\mathbf{y}_t$ (or any component series thereof) is missing at time $t$, but is observed at time $t-1$.

2. Case 2: $\mathbf{y}_t$ (or any component series thereof) is missing at time $t$, and it is also missing at time $t-1$.

3. Case 3: $\mathbf{y}_t$ (or any component series thereof) is observed at time $t$, but is missing at time $t-1$.

The detailed treatment of the three cases is provided in Appendix C.

# 6 A Reduction Technique for the Estimation of the Common Factors

The smoothed estimates of the factors conditional on the parameters vector $\boldsymbol{\Xi}$ can be efficiently obtained by an adaptation of a reduction technique entailing a linear transformation of the model that operates a projection into the space of the factors. This will be useful for the E-step of the EM algorithm.

The computational effort for the KFS depends on the dimensions of both the state vector and the observation vectors. In most practical applications of the dynamic factor model, the dimension of $\mathbf{y}_t$ is significantly larger than the dimension of $\boldsymbol{\alpha}_t$. **?** demonstrated that in such circumstances the computational effort of KFS can significantly be improved by a transformation of the model. Recently, **?** used this computational device for the estimation of the common factors for large scale models.

We illustrate the technique with respect to a generic state space form represented in the following way:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{c}_t + \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{G}_t\boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\omega), \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t\boldsymbol{\alpha}_t + \mathbf{d}_t + \mathbf{H}_t\boldsymbol{\omega}_t, \end{aligned} \tag{13}$$

where for a $p$ dimensional state vector, $\mathbf{c}_t, \mathbf{d}_t$ are suitable vectors of size respectively $N \times 1$ and $p \times 1$, and $\mathbf{Z}_t, \mathbf{G}_t, \boldsymbol{\Sigma}_\omega$, $\mathbf{T}_t$ and $\mathbf{H}_t$ are suitable system matrices that are possibly time-varying.

The reduction technique is based on a linear transformation of the original model which projects the observations on the space spanned by the states $\boldsymbol{\alpha}_t$ and its orthogonal complement. The information that is relevant for extracting the factors lies in the former, which is typically low-dimensional.

Define the transformation matrix

$$\mathbf{A}_t = \begin{pmatrix} \mathbf{A}_t^L \\ \mathbf{A}_t^H \end{pmatrix} \qquad \mathbf{A}_t^L = \boldsymbol{C}_t^{-1}\mathbf{Z}_t'\widetilde{\boldsymbol{\Sigma}}_t^{-1}, \tag{14}$$

where $\widetilde{\boldsymbol{\Sigma}}_t = \mathbf{G}_t \boldsymbol{\Sigma}_\omega \mathbf{G}_t'$, $\boldsymbol{C}_t$ is an invertible matrix, $\mathbf{A}_t^H$ is chosen such that matrix $\mathbf{A}_t$ is full rank, and $\mathbf{A}_t^L \widetilde{\boldsymbol{\Sigma}}_t \mathbf{A}_t^{H'} = \mathbf{0}$, so that $\mathbf{A}_t^H \mathbf{Z}_t = \mathbf{0}$, see **?** and **?**. Moreover by choosing $\boldsymbol{C}_t$ such that $\boldsymbol{C}_t \boldsymbol{C}_t' = \mathbf{Z}_t' \boldsymbol{\Sigma}_\omega^{-1} \mathbf{Z}_t$ and, premultiplying the measurement equation by $\mathbf{A}_t$, we get:

$$\begin{pmatrix} \mathbf{A}_t^L \mathbf{y}_t \\ \mathbf{A}_t^H \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} \mathbf{A}_t^L \mathbf{c}_t \\ \mathbf{A}_t^H \mathbf{c}_t \end{pmatrix} + \begin{pmatrix} \boldsymbol{C}_t' \mathbf{Z}_t \\ \mathbf{0} \end{pmatrix} \boldsymbol{\alpha}_t + \begin{pmatrix} \mathbf{A}_t^L \mathbf{G}_t \boldsymbol{\omega}_t \\ \mathbf{A}_t^H \mathbf{G}_t \boldsymbol{\omega}_t \end{pmatrix}, \quad \begin{pmatrix} \mathbf{A}_t^L \mathbf{G}_t \boldsymbol{\omega}_t \\ \mathbf{A}_t^H \mathbf{G}_t \boldsymbol{\omega}_t \end{pmatrix} \sim \mathrm{N}\left( \mathbf{0} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_t^H \widetilde{\boldsymbol{\Sigma}}_t \mathbf{A}_t^{H'} \end{pmatrix} \right). \tag{15}$$

The unobserved factors are then calculated by applying the KFS to the first block of equations that depends on $\boldsymbol{\alpha}_t$. It is clear that this reduction technique speeds up signal extraction quite dramatically.

This computational device can be adapted to our problem. The reduction is carried out just for the monthly series. This leads, in case of no missing values, to the following formulation:

$$\begin{pmatrix} \mathbf{Y}_{1t}^c \\ A^L \Delta \mathbf{y}_t \\ A^H \Delta \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ A_t^L \mathbf{c}_t \\ A_t^H \mathbf{c}_t \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_1 \\ A_t^L \mathbf{Z}_2 \\ \mathbf{0} \end{pmatrix} \boldsymbol{\alpha}_t + \begin{pmatrix} \mathbf{0} \\ A_t^L \boldsymbol{\epsilon}_{2t} \\ A_t^H \boldsymbol{\epsilon}_{2t} \end{pmatrix}. \tag{16}$$

We cannot apply the reduction to observations missing at time $t$ and time $t-1$; moreover, we do not apply the reduction to case 3 of the previous section, as it does not lead to significant computational gains.

## 7 Estimation of the hyperparameters by the EM algorithm

Maximum likelihood estimation of the model is carried out by the EM algorithm (**?**, **?**). Let $\mathbf{y}_j = [\mathbf{y}_{j0}', \dots, \mathbf{y}_{jn}']', j = 1, 2$, in accordance with the partitioning made in section 4, and $\Delta \mathbf{y}_j = [\Delta \mathbf{y}_{j1}', \dots, \Delta \mathbf{y}_{jn}']', j = 1, 2$; only $\mathbf{y}_2$ is observed at the monthly frequency. The complete data likelihood is, in our application, the joint density of the observed $\Delta \mathbf{y}_2$ and the missing data $\Delta \mathbf{y}_1$ and $\mathbf{f} = [\mathbf{f}_0', \dots, \mathbf{f}_n']'$

$$g(\Delta \mathbf{y}_1, \Delta \mathbf{y}_2, \mathbf{f}; \boldsymbol{\Xi}) = g(\Delta \mathbf{y}_1 | \mathbf{f}; \boldsymbol{\Xi}) g(\Delta \mathbf{y}_2 | \mathbf{f}; \boldsymbol{\Xi}) g(\mathbf{f}; \boldsymbol{\Xi}).$$

Defining the quasi-differences

$$w_{it} = \begin{cases} \sqrt{1 - \psi_i^2} \Delta y_{it}, & t = 1, \\ \Delta y_{it} - \psi_i \Delta y_{i,t-1}, & t = 2, \dots, n, \end{cases}$$

and

$$\mathbf{x}_{it} = \begin{cases} \sqrt{1 - \psi_i^2} [1, \mathbf{f}_t']', & t = 1, \\ \left[ (1 - \psi_i), (\mathbf{f}_t - \psi_i \mathbf{f}_{t-1})' \right]', & t = 2, \dots, n, \end{cases}$$

we can write

$$\ln g(\Delta \mathbf{y}_j | \mathbf{f}; \boldsymbol{\Xi}) = -\frac{1}{2} \sum_{i=1}^{N_j} \left\{ n \ln \sigma_i^2 + \frac{1}{\sigma_i^2} \sum_{t=1}^{n} (w_{it} - \mathbf{x}_{it}' \mathbf{m}_i)^2 \right\}, \tag{17}$$

where

$$\mathbf{m}_i = \left[ \begin{array}{c} \delta_i \\ \boldsymbol{\theta}_i \end{array} \right].$$

This result arises as for the $i$-th series,

$$\Delta y_{it} = \mu_i + \mathbf{f}_t'\boldsymbol{\theta}_i + u_{it}, \qquad u_{it} = \psi_i u_{i,t-1} + \epsilon_{it},$$

with $\epsilon_{it} \sim \text{NID}(0, \sigma_i^2)$. Hence, $\Delta y_{it}, i = 1, \ldots, N$, are conditionally independent given $\mathbf{f}_t$, and are characterised by errors that follow a first order AR model. As a result, if a Choleski orthogonalisation is performed, we get (17). Moreover,

$$\begin{array}{rcl} \ln g(\mathbf{f}; \boldsymbol{\Xi}) & = & -\frac{1}{2}\left[n \ln |\boldsymbol{\Sigma}_\eta| + \text{tr}\left\{\boldsymbol{\Sigma}_\eta^{-1} \sum_{t=1}^n (\mathbf{f}_t - \boldsymbol{\Phi}\mathbf{f}_{t-1})(\mathbf{f}_t - \boldsymbol{\Phi}\mathbf{f}_{t-1})'\right\}\right] \\ & & -\frac{1}{2}\left[\ln |\mathbf{P}_0| + \text{tr}\left\{\mathbf{P}_0^{-1}\mathbf{f}_0\mathbf{f}_0'\right\}\right], \end{array}$$

where $\mathbf{P}_0$ satisfies the matrix equation $\mathbf{P}_0 = \boldsymbol{\Phi}\mathbf{P}_0\boldsymbol{\Phi}' + \boldsymbol{\Sigma}_\eta$. The restrictions imposed on the factor structure imply:

$$\ln g(\mathbf{f}; \boldsymbol{\Xi}) = -\frac{1}{2}\sum_{k=1}^{K}\left\{n + \sum_{t=1}^{n}(f_{kt} - \phi_k f_{k,t-1})^2 - \ln(1 - \phi_k^2) + \frac{f_{k0}^2}{1 - \phi_k^2}\right\}. \tag{18}$$

Given an initial parameter value, $\boldsymbol{\Xi}^*$, the EM algorithm iteratively maximizes, with respect to $\boldsymbol{\Xi}$, the intermediate quantity (**?**):

$$\begin{array}{rcl} Q(\boldsymbol{\Xi}; \boldsymbol{\Xi}^*) & = & \text{E}_{\boldsymbol{\Xi}^*}\left[\ln g(\Delta\mathbf{y}_1, \Delta\mathbf{y}_2, \mathbf{f}; \boldsymbol{\Xi})\right] \\ & = & \int \ln g(\Delta\mathbf{y}_1, \Delta\mathbf{y}_2, \mathbf{f}; \boldsymbol{\Xi}) g(\Delta\mathbf{y}_1, \mathbf{f}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \boldsymbol{\Xi}^*) d(\Delta\mathbf{y}_1, \mathbf{f}), \end{array} \tag{19}$$

which is interpreted as the expectation of the complete data log-likelihood with respect to $g(\Delta\mathbf{y}_1, \mathbf{f}| \Delta\mathbf{y}_2, \mathbf{Y}_1^c; \boldsymbol{\Xi}^*)$, which is the conditional probability density function of the unobservable states, given the observations, evaluated at $\boldsymbol{\Xi}^*$.

**?** show that the parameter estimates maximising the log-likelihood $\log L(\boldsymbol{\Xi})$, can be obtained by a sequence of iterations, each consisting of an expectation step (E-step) and a maximization step (M-step), that aim at locating a stationary point of $Q(\boldsymbol{\Xi}; \boldsymbol{\Xi}^*)$.

## 7.1 The E step

At iteration $m$, given the estimate $\boldsymbol{\Xi}^{(m)}$, the E-step deals with the evaluation of $Q(\boldsymbol{\Xi}; \boldsymbol{\Xi}^{(m)})$; this is carried out with the support of the KFS applied to the state space representation provided in section (4.2) with hyperparameters $\boldsymbol{\Xi}^{(m)}$.

As it is evident from (17)-(19) the E-step computes

$$\text{E}_{\boldsymbol{\Xi}^{(m)}}\left(\sum_{t=1}^{n}\mathbf{x}_{it}\mathbf{x}_{it}'\right), \quad \text{E}_{\boldsymbol{\Xi}^{(m)}}\left(\sum_{t=1}^{n}\mathbf{x}_{it}w_{it}\right), \quad \text{E}_{\boldsymbol{\Xi}^{(m)}}\left(\sum_{t=1}^{n}w_{it}^2\right),$$

$$\text{E}_{\boldsymbol{\Xi}^{(m)}}\left(\sum_{t=2}^{n}f_{k,t-1}^2\right), \quad \text{E}_{\boldsymbol{\Xi}^{(m)}}\left(\sum_{t=1}^{n}f_{kt}^2\right), \quad \text{E}_{\boldsymbol{\Xi}^{(m)}}\left(\sum_{t=2}^{n}f_{kt}f_{k,t-1}\right),$$

where the expectation is taken with respect to the density $g(\boldsymbol{\Delta}\mathbf{y}_1, \mathbf{f}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \boldsymbol{\Xi}^{(m)})$. As a by product,

$$\mathrm{E}_{\Xi^{(m)}}\left(\sum_{t=2}^{n} u_{i,t-1}^2\right), \quad \mathrm{E}_{\Xi^{(m)}}\left(\sum_{t=1}^{n} u_{it}^2\right), \quad \mathrm{E}_{\Xi^{(m)}}\left(\sum_{t=2}^{n} u_{it}u_{i,t-1}\right),$$

are made available. It must be remarked that for $i = N_1 + 1, \ldots, N$, $w_{it}$ is observed and thus $\mathrm{E}_{\Xi^{(m)}}\left(\sum_t w_{it}^2\right) = \sum_t w_{it}^2$,$\mathrm{E}_{\Xi^{(m)}}\left(\sum_t \mathbf{x}_{it}w_{it}\right) = \sum_t \mathrm{E}_{\Xi^{(m)}} w_{it}\left(\mathbf{x}_{it}\right)$.

In order to be able to evaluate the covariance between the current and past states, e.g $\mathrm{E}_{\Xi^{(*)}}(\sum_{t=2}^{n} f_{kt}f_{k,t-1})$, the KFS recursion needs to be augmented as in **?** and **?**.

In sum, the E step consists of evaluating the conditional expectation of cross-products of random quantities. For instance,

$$\mathrm{E}_{\Xi^{(m)}}\left(\sum_{t=1}^{n} \mathbf{x}_{it}\mathbf{x}_{it}'\right) = \sum_{t=1}^{n}\left[\tilde{\mathbf{x}}_{it|n}\tilde{\mathbf{x}}_{it|n}' + \mathbf{V}_{x|n}\right],$$

where $\tilde{\mathbf{x}}_{it|n} = \mathrm{E}(\mathbf{x}_{it}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$ and $\mathbf{V}_{x|n} = \mathrm{Var}(\mathbf{x}_{it}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$.

## 7.2 The M-step

The M-step amounts to choosing a new value $\boldsymbol{\Xi}^{(m+1)}$, so as to maximize with respect to $\boldsymbol{\Xi}$ the criterion $Q(\boldsymbol{\Xi}; \boldsymbol{\Xi}^{(m)})$, i.e., $Q(\boldsymbol{\Xi}^{(m+1)}; \boldsymbol{\Xi}^{(m)}) \geq Q(\boldsymbol{\Xi}^{(m)}; \boldsymbol{\Xi}^{(m)})$. The maximization is in closed form, if, as it is customary, we ignore the term $\ln(1-\phi_k^2) + \frac{f_{k0}^2}{1-\phi_k^2}$ in (18) and drop the component $(w_{i1} - \mathbf{x}_{i1}'\boldsymbol{\delta}_i)^2$ for $t = 1$ in (17).

As a result of this simplification, the new estimates of the means and the loadings of the $i$-th series, $i = 1, \ldots, N_1$, are

$$\begin{bmatrix} \hat{\delta}_i^{(m+1)} \\ \hat{\boldsymbol{\theta}}_i^{(m+1)} \end{bmatrix} = \left[\sum_{t=1}^{n}\left(\tilde{\mathbf{x}}_{it|n}\tilde{\mathbf{x}}_{it|n}' + \mathbf{V}_{x|n}\right)\right]^{-1}\sum_{t=1}^{n}\left(\tilde{\mathbf{x}}_{it|n}\tilde{w}_{it|n} + \mathbf{V}_{xw|n}\right),$$

where $\tilde{w}_{it|n} = \mathrm{E}(w_{it}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$ and $\mathbf{V}_{xw|n} = \mathrm{Cov}(\mathbf{x}_{it}, w_{it}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$.

For the series belonging to second group, $i = N_1 + 1, \ldots, N$,

$$\begin{bmatrix} \hat{\delta}_i^{(m+1)} \\ \hat{\boldsymbol{\theta}}_i^{(m+1)} \end{bmatrix} = \left[\sum_{t=1}^{n}\left(\tilde{\mathbf{x}}_{it|n}\tilde{\mathbf{x}}_{it|n}' + \mathbf{V}_{x|n}\right)\right]^{-1}(\mathbf{I} - \boldsymbol{\Delta}_t)\sum_{t=1}^{n}\tilde{\mathbf{x}}_{it|n}w_{it} + \boldsymbol{\Delta}_t\sum_{t=1}^{n}\tilde{\mathbf{x}}_{it|n}\mathbf{V}_{xm|n},$$

where $\mathbf{I}$ is a unitary matrix and $\boldsymbol{\Delta}_t$ is a matrix with element $\Delta_{i,j} = 1$ when an observation is missing. Finally $\mathbf{V}_{xm|n}$ is the smooth covariance between missing values that has to be calculated for the maximization step.

Since $w_{it}$ is observed

$$\hat{\sigma}_i^{(m+1)} = \frac{1}{n}\sum_{t=1}^{n}\mathrm{E}_{\Xi^{(m)}}\left[(w_{it} - \mathbf{x}_{it}'\boldsymbol{\delta}_i^{(m+1)})^2\right],$$

where, if $i \leq N_1$,

$$\mathrm{E}_{\Xi^{(m)}}\left[(w_{it} - \mathbf{x}_{it}'\boldsymbol{\delta}_i^{(m)})^2\right] = (\tilde{w}_{it|n} - \tilde{\mathbf{x}}_{it|n}'\boldsymbol{\delta}_i^{(m)})^2 + V_{e|n}$$

and $V_{e|n} = \text{Var}(w_{it} - \mathbf{x}'_{it}\boldsymbol{\delta}_i^{(m)}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$; else, $w_{it}$ is available, and thus,

$$\text{E}_{\Xi^{(m)}}\left[(w_{it} - \mathbf{x}'_{it}\boldsymbol{\delta}_i^{(m)})^2\right] = (w_{it} - \tilde{\mathbf{x}}'_{it|n}\boldsymbol{\delta}_i^{(m)})^2 + V_{e|n}.$$

The estimates of the factor autoregressive coefficients are

$$\hat{\phi}_k^{(m+1)} = \frac{\sum_{t=1}^n \left(\tilde{f}_{kt|n}\tilde{f}_{k,t-1|n} + V_{ff_1|n}\right)}{\sum_{t=1}^n \left(\tilde{f}_{k,t-1|n}^2 + V_{f_1f_1|n}\right)},$$

where $\tilde{f}_{kt|n} = \text{E}(f_{kt}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$, $V_{ff_1|n} = \text{Cov}(f_{kt}, f_{k,t-1}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$, and $V_{f_1f_1|n} = \text{Var}(f_{k,t-1}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$.

As for the parameters of the idiosyncratic components,

$$\hat{\psi}_i^{(m+1)} = \frac{\sum_{t=2}^n \left(\tilde{u}_{it|n}\tilde{u}_{i,t-1|n} + V_{uu_1|n}\right)}{\sum_{t=2}^n \left(\tilde{u}_{i,t-1|n}^2 + V_{u_1u_1|n}\right)},$$

where $\tilde{u}_{it|n} = \text{E}(u_{it}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$, $V_{uu_1|n} = \text{Cov}(u_{it}, u_{i,t-1}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$, and $V_{u_1u_1|n} = \text{Var}(u_{i,t-1}|\Delta\mathbf{y}_2, \mathbf{Y}_1^c; \Xi^{(m)})$.

When $\Delta\mathbf{y}_{2t}$ is subject to missing values the algorithm has to be modified for taking into account the conditional covariances between the missing values and the latent states.

## 8 Description of the dataset

We now apply the methodology described in the previous sections to the construction of Euromind-C, an indicator of economic conditions in the Euro area and its largest member states. The set of time series making up the system has dimension $N = 171$. The first 55 are the national accounts quarterly series concerning the value added chained volumes with reference year 2000 for all the sectors and the main expenditure components of the GDP, referred to the Euro area and the four selected countries (see paragraph 1). The remaining time series are monthly indicators.

We have carried out a systematic review of data availability by geographical entity which has led to the recognition of important information gaps due to a variety of reasons: the production of the new orders indicator in industry has been discontinued; the collection of the new car registration has been discontinued by Eurostat while the series continue to be made available by other sources; some labour market indicators in short term business statistics, are not anymore regularly produced. Nevertheless, since our methods allow for arbitrary patterns of missing values, we decided to retain the series in the database, even if they are out of date, to support the historical disaggregation of value added.

Table 4 provides a list of the time series that have been used in the estimation of EuroMInd-C organized by GDP component. The series are organized according to the reference sector or GDP component:

- Agriculture: the monthly indicators concern the production of milk, bovine meat production in tons, pigs meat production in tons.
- Industry: the monthly indicators concern the index of industrial production, the index of turnover and new orders; we also include the industrial confidence indicator compiled by the European Commission.

- Construction: The monthly indicators concern the volume index of production for construction, employment and hours worked (Germany) and building permits.
- Trade, transport and communication: the monthly indicators concern the total turnover of the retail sector, the index of deflated turnover, employment (Germany) and car registration, along with the survey based retail confidence indicator.
- Financial services and business: the monthly indicators concern the index of industrial production as a proxy measure of the state of the economy, the stock market indices and capitalization (when available), a set of 5 monetary aggregates for the Euro area (M1-M3, loans, etc.) and financial ones (deposits of residents held at monetary financial institutions, and credit to total residents granted by monetary financial institutions).
- Other services: the monthly indicators concern the index of industrial production, the index of retail turnover and the monthly unemployment rate as proxy measures of the state of the economy.
- Taxes less subsidies on products: the monthly indicators concern the index of industrial production and the index of retail turnover as proxy measures of the state of the economy.
- Final Consumption Expenditures: the monthly indicators concern the index of deflated turnover of the retail sector, car registration and two survey indicators, the index of consumer confidence and the index of major purchases at present, compiled by the European Commission.
- Gross Capital Formation: the monthly indicators concern the index of industrial production for capital goods and the index of economic sentiment compiled by the European Commission.
- Exports of goods and services: the monthly indicators concern the index of industrial production and the volume index of exports.
- Imports of goods and services: the monthly indicators concern the index of industrial production and the volume index of imports.

For a complete list of the series in the dataset and their reference period see Appendix A.

## 9 Estimation results

The dynamic factor model is estimated for the sample period January 1995 - September 2012. The computational complexity of the model is quite substantial, with a convergence of the EM algorithm after 250 iterations. Conditional on the maximum likelihood estimates of the hyperparameters, we can perform signal extraction according to the nonlinear smoothing algorithm outlined in the previous sections. This yields the monthly estimates of the components of GDP (disaggregated also by country) by economic activity and expenditure components, from which the estimates of total GDP at market prices are compiled, according to the contemporaneous aggregation procedure outlined in section Appendix C.

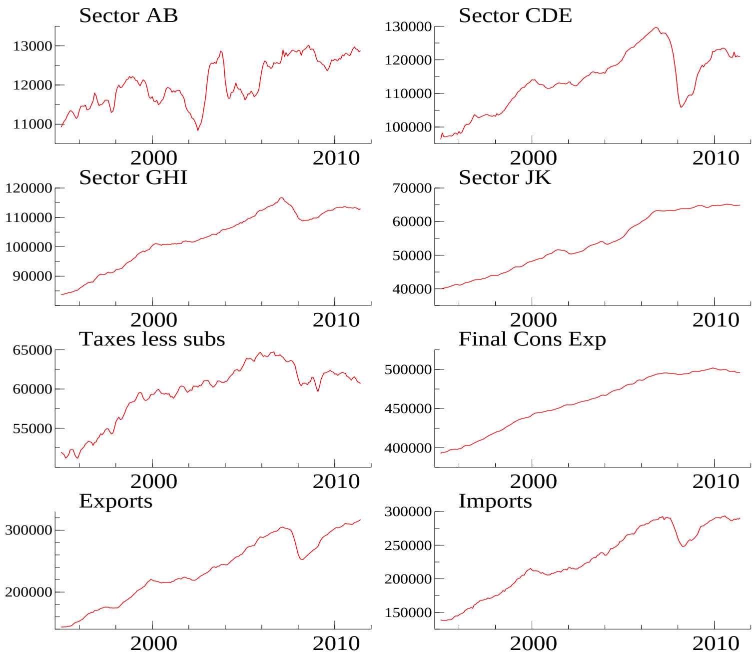


***Figure 1:*** *Monthly indicators of the 11 components of Euro area GDP, levels (chained 2000 volumes).*

EuroMInd-C makes available the monthly estimates of the 11 GDP components for the Euro area and the 4 largest economies; these are relevant for the assessment of the state of the economy from the point of view of sectorial value added and by expenditure component.

Figure 1 displays the levels in millions of chained euros, whereas figure 2 presents the growth rates for the Euro area as a whole. It is noticeable that the real effects of the global financial crisis hit the manufacturing sector quite dramatically, although the recovery was sustained. The effects on the fall in the output of the construction sector have been prolonged and, until the end of 2012,there was no exit from the recession for this sector.

## 9.1 The Reconciliation of the Estimates of GDP from the Output and the Expenditure approach

As stated in section 4, 11 quarterly GDP components make up the dataset for each of the 4 countries and the Euro area. Our model produces the disaggregate estimates of the monthly components conditional on the information available, the estimates can be combined to produce the final estimate of monthly GDP. Two issues are posed by the contemporaneous aggregation of the output and expenditure components of GDP. First and foremost the quarterly national accounts series are subject to the following accounting deterministic constraints,

$$\begin{aligned} \text{GDP}_O \text{ at market prices} &= \textstyle\sum \text{Value added of the 6 branches (A-B, C-D-E, F, G-H-I, J-K, L-P)} \\ &\quad + \text{Taxes less subsidies} \\ \text{GDP}_E \text{ at market prices} &= \text{FCE+GCF+EXP-IMP} \end{aligned}$$

only when the aggregates are expressed at current prices and at the average prices of the previous year. Secondly, the values of GDP at market prices from the output and expenditure approaches

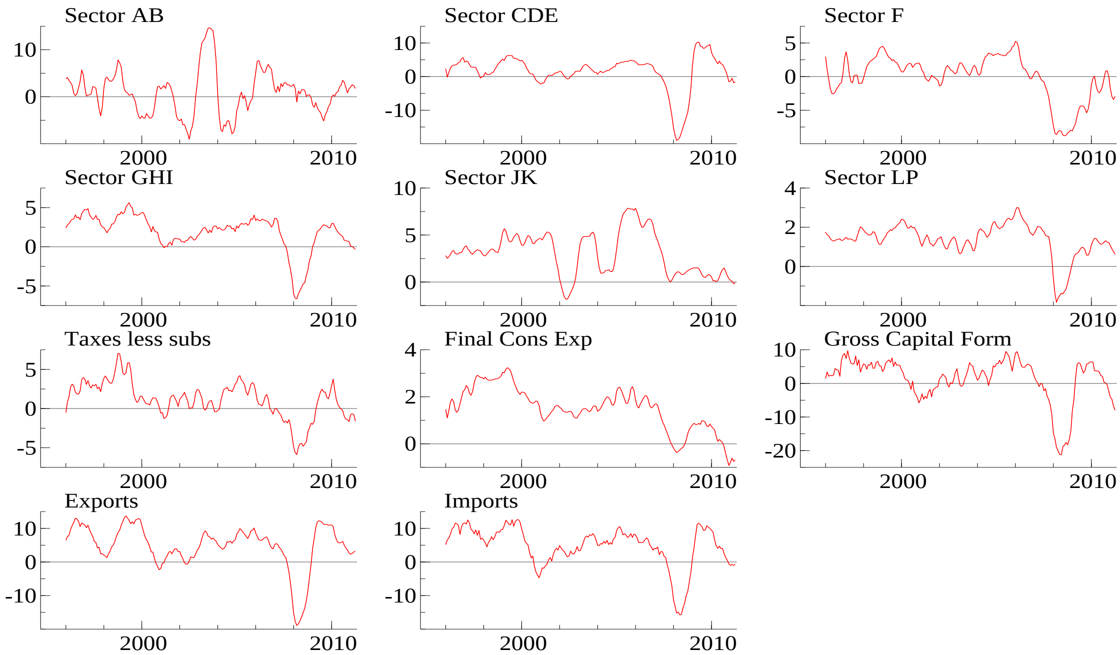


***Figure 2:*** *Monthly indicators of the 11 components of Euro area GDP, yearly growth rates (chained 2000 volumes).*

not necessarily coincide and have to be reconciled.

The first issue is a consequence of the current method of production of chained linked national accounts estimates. Chaining, which is a multiplicative operation, prevents the additivity of the national accounts constraints, and a nonzero discrepancy arises. GDP and its main components are expressed in chain-linked volumes (millions of euros), with reference year 2000, which implies that the constraints hold exactly only for the four quarters of the year 2001.

Interestingly, due to the application of the *annual overlap technique* for the compilation of volume data of the quarterly national accounts, constraints are not entirely lost, but they continue to hold after a transformation of the data, to which we refer as "dechaining", which aims at expressing the chained values at the prices of the previous year. According to the annual overlap technique, (see **?**, Chapter IX), the estimates for each quarter are compiled using the weighted annual average prices of the previous year; this produces quarterly volume estimates that sum up exactly to the corresponding annual aggregate. As a result, the disaggregated (monthly and quarterly) volume measures expressed at the prices of the previous year preserve both the temporal and cross-sectional additivity.

The cross-sectional constraints can be enforced by a multistep procedure that de-chains the estimated monthly values, expressing them at the average prices of the previous year, and projects the estimates on the subspace of the constraints. The procedure is described in details in **?** and is reproduced in Appendix D.

The monthly GDP estimates for the Euro area and the four largest economies arising from the output approach are displayed in figure 3, along with their approximate 95% confidence limits.

The confidence intervals take into account what is often referred to as filtering uncertainty, i.e. they reflect the estimation error variance of the unobserved monthly time series, conditional on the

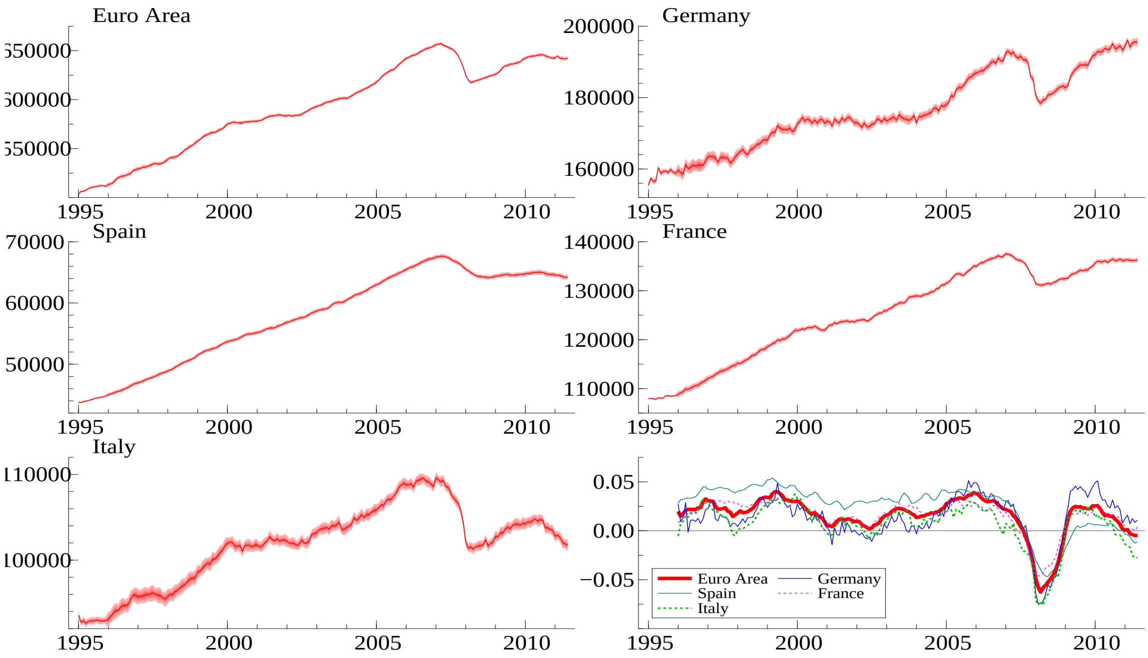


***Figure 3:*** *EuroMInd-C estimates: Monthly indicator of GDP at market prices (chained 2000 volumes). Estimates from the output approach.*

full available observed sample and the maximum likelihood estimates of the parameters. Hence, parameter uncertainty is not reflected in the confidence regions.

It should be noticed that the mean square estimation error of the unobserved monthly series is higher for Germany and Italy; this is so since the two countries have a higher share of value added of the manufacturing sector (branch C-D-E), which is more volatile than the other sectors (excluding agriculture). The bottom right panel displays the underlying yearly growth rates.

The estimates from the expenditure approach are obtained from the sum of consumption, investment and net exports (FCE+GCF+EXP-IMP). They are displayed in figure 4, along with their 95% confidence interval and the underlying yearly growth rates (bottom right panel).

The estimates arising from the output and the expenditure approaches are not coincident, as it is well known, but they differ only slightly and can be reconciled by a suitable combination, which is carried out as follows. Denoting the GDP estimates obtained by the output approach by $Y_t^o$, and those from the expenditure side by $Y_t^e$, by $S_t^{2o}$ and $S_t^{2e}$ the estimation error variance of the output and expenditure estimates, we use a set of weights proportional to the relative precision of the output estimates, that is

$$w_{ot} = \frac{1/S_t^{2o}}{1/S_t^{2o} + 1/S_t^{2e}}.$$

The combined estimates are then balanced according to the procedure outlined in Appendix C. The combined estimates of the yearly and monthly growth rates are displayed in figures 5. Obviously, it turns out that the combined estimate is more precise than $Y_t^o$ and $Y_t^e$. The last panel illustrates quite effectively the average and differential pace at which the recovery took place after the global financial crises and the depth of the recessionary movement.

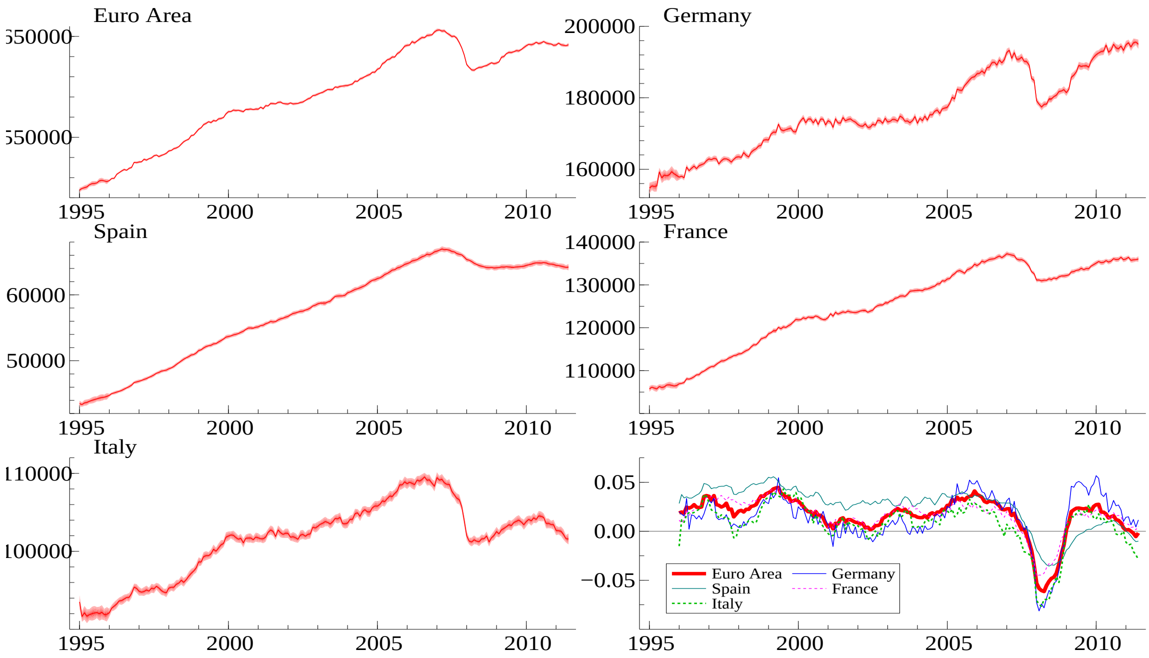


***Figure 4:*** *EuroMInd-C estimates: Monthly indicator of GDP at market prices (chained 2000 volumes). Estimates from the expenditure approach.*

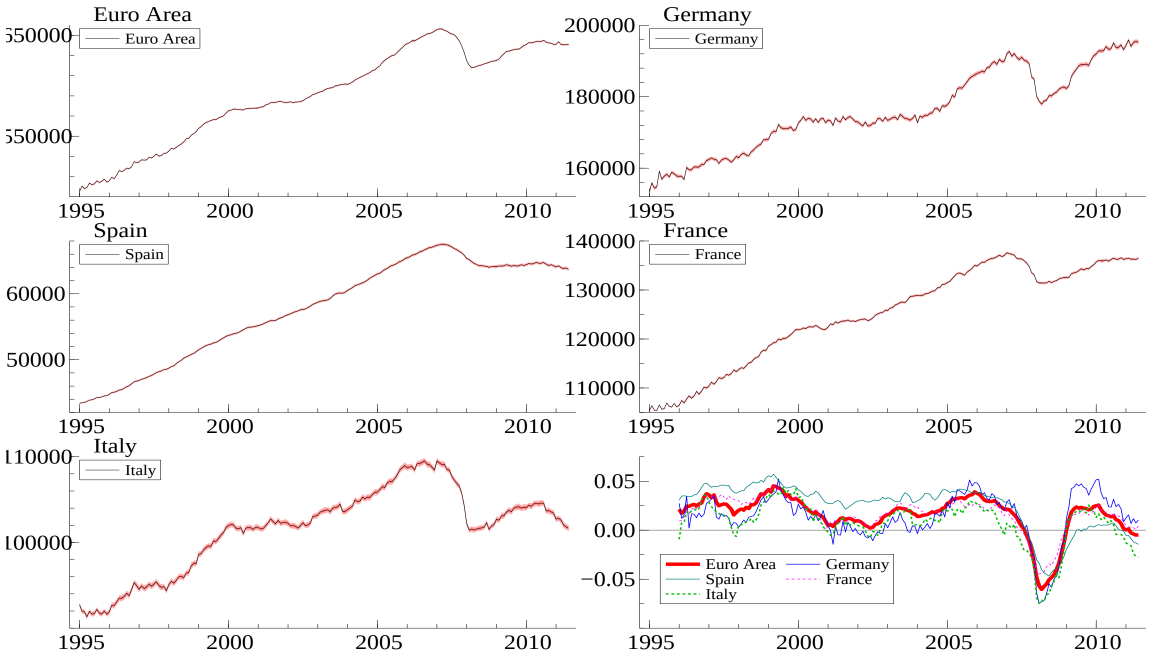


***Figure 5:*** *EuroMInd-C: GDP at market prices (chained 2000 volumes). Combined estimates.*

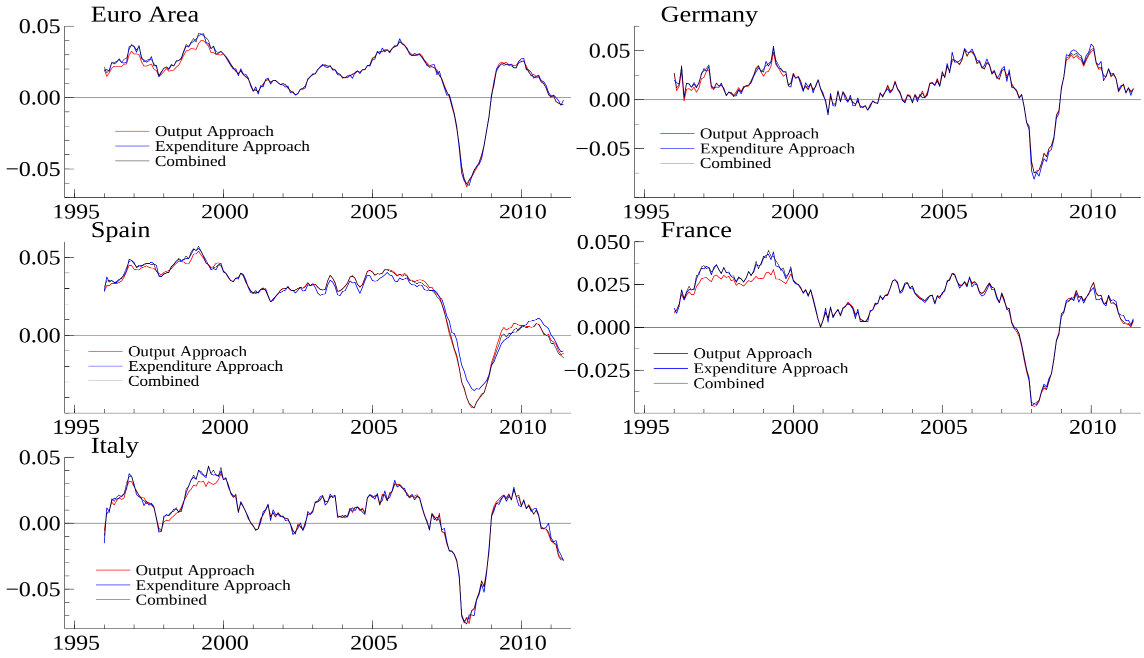


***Figure 6:*** *EuroMInd-C estimates: Comparison of the estimates of the yearly growth rates arising from the expenditure and the output approach.*

Figure 6 provides a comparison of the estimates of the yearly growth rates arising from the expenditure and the output approach, and superimposes also the final Euromind-C estimates of the growth rates. The most sizable differences concern the growth rates for Spain. They relate to the statistical discrepancy between the quarterly estimates of GDP at market prices obtained from the output and expenditure approach. As such, it is a characteristic of the raw data, rather than of our methodology. In general, the estimates arising from the expenditure approach are more volatile.

## 9.2 Variance decompositions

The monthly indicator of GDP draws information from the latent factors. Each variable (monthly indicators and national accounts series) contributes to the definition of the factors, according to the model specification. *Coeteris paribus*, it is desirable that the individual variables load significantly on the common factors and that the latter contribute significantly to the national accounts series. For country $c = 2, \ldots, C$, the relative contribution of the factors in explaining the variability of the series can be obtained from the following variance decomposition:

$$\text{Var}(\Delta y_{cs,t}) = \theta^2_{cs,0}\text{Var}(\Delta\mu_t) + \theta^2_{cs,1}\text{Var}(\Delta\mu_{ct}) + \theta^2_{cs,2}\text{Var}(\Delta\mu_{st}) + \text{Var}(\Delta\mu^*_{cs,t}).$$

For the Euro area,

$$\text{Var}(\Delta y_{1s,t}) = \theta^2_{1s,0}\text{Var}(\Delta\mu_t) + \sum_{c=2}^{C}\theta^2_{1c,1}\text{Var}(\Delta\mu_{ct}) + \theta^2_{1s,2}\text{Var}(\Delta\mu_{st}) + \text{Var}(\Delta\mu^*_{1s,t}).$$

The four addends on the right, expressed as a percentage of the total variance on the left hand side, are reported in table 2, which displays, along with the median value (middle columns), also the 33th (1/3) and 66 (2/3) percentiles of the sampling distribution of the statistic. The table is divided into separate panels, referring to the different series, the first 11 referring to the national accounts series. The following panels refer to the different groups of monthly indicators. In the interpretation of these numbers it has to be considered that the larger the variance share that is absorbed by the common and the sector specific factors, the greater the contribution of an indicator to EuroMind-C.

An interesting fact is that the value added of the sectors C-D-E (manufacturing), G-H-I (Trade, transport and communication services), J-K (Financial services and business activities) display the highest degree of commonality, whereas Construction and Other Services are predominantly idiosyncratic. As far as the expenditure components are concerned, comovements play a larger role for GCF, whereas the other components (Consumption, Exports and Imports) are predominantly idiosyncratic. This is perhaps the reason why the EuroMInd-C estimates of GDP according to the expenditure approach are less reliable with respect to those arising from the output approach.

Another conclusion we may draw from table 2 is that the Spanish economy behaves more idiosyncratically. Among the monthly indicators, the industrial production series (sts_impr_msa) contribute most sizably to the common and the component specific factors. Similar considerations hold for industrial turnover and deflated turnover of the retail sector (sts_trtu_defltv).

The least significant contributions to the common factors are associated with the indicators for the construction sector (sts_colb and sts_copr), the financial indicators (mfma_m) and the survey indicators for the expenditure components (bssi_GCF, bssi_FCE) and the sector G-H-I. A possible use of this information is the deletion of series that do not contribute significantly to the common factors.

## 10 A Pseudo Real-Time Comparative Assessment of Euromind-C

We now provide a comparative assessment of EuroMInd-C by performing a recursive forecasting experiment focusing on the ability to predict GDP and its components for the Euro area referring to quarters $\tau = 2005.q1, 2005.q2, \dots, 2011.q4$, using the information set available at the end of the month preceding the end of the quarter (February 2005 for 2005.q1, May 2005 for 2005.q2, etc.). The prediction of the 11 components of GDP arises from the aggregation of the nowcasts for the estimation month and the previous month (e.g. February 2005 and January 2005 when predicting 2005.q1) and the one-step-ahead forecast for the last month of the quarter (March 2005). The model is re-estimated every time using the most recent data.

The same experiment is carried out for EuroMInd, which uses the componentwise estimation methodology described in Frale et al. (2010). This yields $M = 28$ forecasts $\hat{Y}_{i\tau}$, where $i$ indexes a particular GDP component and $\tau = 2005.q1, \dots 2011.q4$, for each method.

We also consider as a reference the naïve (”no change” or random walk) predictor $\hat{Y}_{i\tau} = Y_{i,\tau-1}$, which will be referred to henceforth as RW. We expect the EuroMInd and EuroMInd-C predictions to be more informative as they exploit the monthly information made available for the first two months of the quarter.

For the three predictors we compare the forecasts with the observed value, $Y_{i\tau}$ and summarise the distribution of the prediction errors $Y_{i\tau} - \hat{Y}_{i\tau}, \tau = 2005.q1, \dots 2011.q4$, by means of the mean

*Table 2: Variance Decomposition for the National Account series. The table reports the median value of the percentage share of variance explained by each factor, along with the 33th and 66th percentiles of the sampling distribution of the statistic.*

| | **Common European Factor** | | | **Country specific Factors** | | | **Component specific Factors** | | | **Idiosyncratic** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1/3 | Med | 2/3 | 1/3 | Med | 2/3 | 1/3 | Med | 2/3 | 1/3 | Med | 2/3 |
| **Sector A-B** | | | | | | | | | | | | |
| EA | 26.69 | 47.66 | 67.31 | 2.36 | 4.22 | 5.97 | 2.05 | 3.66 | 5.17 | 24.89 | 44.44 | 62.77 |
| Germany | 11.68 | 20.86 | 29.46 | 0.49 | 0.89 | 1.25 | 2.16 | 3.85 | 5.45 | 41.66 | 74.38 | 105.06 |
| France | 1.16 | 2.07 | 2.92 | 0.37 | 0.67 | 0.94 | 6.56 | 11.71 | 16.54 | 47.90 | 85.53 | 120.82 |
| Spain | 21.22 | 37.89 | 53.52 | 1.32 | 2.36 | 3.33 | 5.24 | 9.36 | 13.22 | 28.21 | 50.37 | 71.15 |
| Italy | 27.14 | 48.47 | 68.46 | 0.20 | 0.35 | 0.50 | 0.11 | 0.20 | 0.29 | 28.54 | 50.96 | 71.97 |
| **Sector C-D-E** | | | | | | | | | | | | |
| EA | 32.27 | 57.62 | 81.38 | 8.51 | 15.20 | 21.47 | 0.99 | 1.78 | 2.51 | 14.21 | 25.38 | 35.86 |
| Germany | 29.76 | 53.14 | 75.06 | 14.29 | 25.52 | 36.04 | 0.59 | 1.05 | 1.49 | 11.35 | 20.27 | 28.63 |
| France | 31.22 | 55.75 | 78.74 | 0.56 | 1.01 | 1.42 | 0.36 | 0.65 | 0.92 | 23.85 | 42.58 | 60.15 |
| Spain | 31.02 | 55.38 | 78.22 | 5.33 | 9.53 | 13.46 | 3.74 | 6.69 | 9.45 | 15.90 | 28.39 | 40.10 |
| Italy | 17.90 | 31.96 | 45.14 | 9.27 | 16.56 | 23.39 | 8.13 | 14.51 | 20.5 | 20.69 | 36.95 | 52.19 |
| **Sector F** | | | | | | | | | | | | |
| EA | 5.33 | 9.53 | 13.46 | 4.29 | 7.66 | 10.82 | 10.68 | 19.07 | 26.94 | 35.69 | 63.73 | 90.01 |
| Germany | 0.2 | 0.36 | 0.52 | 3.45 | 6.16 | 8.70 | 11.36 | 20.29 | 28.67 | 40.98 | 73.16 | 103.34 |
| France | 8.56 | 15.30 | 21.61 | 7.97 | 14.23 | 20.11 | 0.28 | 0.50 | 0.71 | 39.18 | 69.95 | 98.8 |
| Spain | 0.49 | 0.89 | 1.25 | 14.01 | 25.02 | 35.34 | 1.72 | 3.08 | 4.35 | 39.76 | 71.00 | 100.28 |
| Italy | 4.15 | 7.42 | 10.49 | 6.84 | 12.22 | 17.27 | 2.50 | 4.47 | 6.31 | 42.49 | 75.87 | 107.16 |
| **Sector G-H-I** | | | | | | | | | | | | |
| EA | 31.17 | 55.66 | 78.62 | 6.17 | 11.02 | 15.56 | 0.86 | 1.55 | 2.19 | 17.78 | 31.75 | 44.85 |
| Germany | 25.88 | 46.21 | 65.27 | 5.12 | 9.15 | 12.93 | 1.50 | 2.68 | 3.79 | 23.49 | 41.94 | 59.24 |
| France | 22.35 | 39.9 | 56.36 | 1.23 | 2.20 | 3.11 | 0.91 | 1.64 | 2.31 | 31.50 | 56.24 | 79.44 |
| Spain | 29.14 | 52.04 | 73.50 | 8.68 | 15.51 | 21.91 | 0.27 | 0.48 | 0.68 | 17.90 | 31.96 | 45.14 |
| Italy | 26.64 | 47.56 | 67.18 | 4.59 | 8.19 | 11.57 | 0.51 | 0.92 | 1.30 | 24.25 | 43.31 | 61.17 |
| **Sector J-K** | | | | | | | | | | | | |
| EA | 38.90 | 69.46 | 98.11 | 4.32 | 7.71 | 10.90 | 0.01 | 0.03 | 0.05 | 12.76 | 22.78 | 32.18 |
| Germany | 38.54 | 68.81 | 97.20 | 0.54 | 0.98 | 1.38 | 0.37 | 0.66 | 0.93 | 16.54 | 29.53 | 41.71 |
| France | 12.90 | 23.04 | 32.54 | 6.38 | 11.40 | 16.10 | 1.64 | 2.92 | 4.13 | 35.07 | 62.62 | 88.46 |
| Spain | 26.00 | 46.43 | 65.58 | 11.32 | 20.21 | 28.55 | 0.12 | 0.22 | 0.31 | 18.55 | 33.12 | 46.79 |
| Italy | 30.71 | 54.84 | 77.46 | 3.76 | 6.71 | 9.48 | 0.13 | 0.23 | 0.33 | 21.39 | 38.20 | 53.95 |
| **Sector L-P** | | | | | | | | | | | | |
| EA | 4.39 | 7.84 | 11.08 | 9.39 | 16.78 | 23.70 | 0.11 | 0.19 | 0.28 | 42.10 | 75.17 | 106.17 |
| Germany | 7.91 | 14.13 | 19.95 | 4.53 | 8.09 | 11.43 | 0.04 | 0.07 | 0.10 | 43.51 | 77.69 | 109.74 |
| France | 6.47 | 11.56 | 16.33 | 0.18 | 0.33 | 0.46 | 1.30 | 2.33 | 3.29 | 48.03 | 85.76 | 121.14 |
| Spain | 29.81 | 53.23 | 75.19 | 2.61 | 4.66 | 6.58 | 0.20 | 0.37 | 0.52 | 23.37 | 41.72 | 58.93 |
| Italy | 11.72 | 20.94 | 29.57 | 0.38 | 0.69 | 0.97 | 4.62 | 8.24 | 11.65 | 39.27 | 70.11 | 99.04 |
| **Sector TlS** | | | | | | | | | | | | |
| EA | 18.08 | 32.28 | 45.59 | 4.04 | 7.22 | 10.20 | 15.68 | 28.00 | 39.56 | 18.19 | 32.48 | 45.87 |
| Germany | 3.07 | 5.49 | 7.75 | 3.87 | 6.91 | 9.76 | 18.80 | 33.57 | 47.42 | 30.25 | 54.01 | 76.29 |
| France | 6.43 | 11.49 | 16.23 | 7.11 | 12.70 | 17.94 | 8.18 | 14.61 | 20.63 | 34.27 | 61.19 | 86.43 |
| Spain | 18.73 | 33.45 | 47.25 | 9.99 | 17.83 | 25.19 | 2.94 | 5.25 | 7.42 | 24.33 | 43.45 | 61.37 |
| Italy | 21.00 | 37.50 | 52.96 | 2.88 | 5.14 | 7.27 | 11.68 | 20.86 | 29.46 | 20.43 | 36.49 | 51.54 |
| **Sector FCE** | | | | | | | | | | | | |
| EA | 3.53 | 6.30 | 8.90 | 9.18 | 16.4 | 23.16 | 0.41 | 0.73 | 1.04 | 42.87 | 76.55 | 108.12 |
| Germany | 0.04 | 0.08 | 0.11 | 0.00 | 0.00 | 0.00 | 7.16 | 12.79 | 18.07 | 48.79 | 87.11 | 123.05 |
| France | 2.53 | 4.53 | 6.40 | 7.71 | 13.77 | 19.45 | 0.10 | 0.18 | 0.25 | 45.65 | 81.51 | 115.13 |
| Spain | 17.20 | 30.71 | 43.38 | 0.39 | 0.70 | 0.99 | 0.14 | 0.25 | 0.36 | 38.26 | 68.32 | 96.50 |
| Italy | 14.96 | 26.71 | 37.72 | 3.86 | 6.89 | 9.73 | 0.00 | 0.00 | 0.00 | 37.18 | 66.39 | 93.77 |
| **Sector GCF** | | | | | | | | | | | | |
| EA | 31.85 | 56.86 | 80.32 | 5.88 | 10.49 | 14.82 | 0.50 | 0.90 | 1.28 | 17.76 | 31.72 | 44.80 |
| Germany | 9.58 | 17.10 | 24.16 | 0.01 | 0.03 | 0.04 | 1.59 | 2.85 | 4.02 | 44.80 | 80.00 | 113.00 |
| France | 8.55 | 15.27 | 21.56 | 2.85 | 5.10 | 7.20 | 0.08 | 0.15 | 0.21 | 44.51 | 79.47 | 112.25 |
| Spain | 29.38 | 52.45 | 74.09 | 9.64 | 17.22 | 24.32 | 0.52 | 0.93 | 1.31 | 16.45 | 29.38 | 41.50 |
| Italy | 35.43 | 63.26 | 89.35 | 3.17 | 5.66 | 7.99 | 0.44 | 0.79 | 1.12 | 16.95 | 30.27 | 42.76 |
| **Sector EXP** | | | | | | | | | | | | |
| EA | 0.88 | 1.57 | 2.22 | 2.90 | 5.18 | 7.31 | 14.31 | 25.56 | 36.10 | 37.90 | 67.68 | 95.59 |
| Germany | 0.29 | 0.52 | 0.74 | 0.12 | 0.22 | 0.31 | 14.35 | 25.63 | 36.20 | 41.23 | 73.61 | 103.97 |
| France | 4.37 | 7.81 | 11.04 | 0.33 | 0.59 | 0.83 | 3.89 | 6.95 | 9.82 | 47.40 | 84.63 | 119.54 |
| Spain | 0.00 | 0.00 | 0.00 | 1.61 | 2.88 | 4.07 | 2.00 | 3.58 | 5.06 | 52.38 | 93.52 | 132.10 |
| Italy | 2.75 | 4.92 | 6.95 | 1.55 | 2.77 | 3.92 | 16.13 | 28.80 | 40.68 | 35.56 | 63.50 | 89.69 |
| **Sector IMP** | | | | | | | | | | | | |
| EA | 0.49 | 0.89 | 1.25 | 12.44 | 22.22 | 31.39 | 6.98 | 12.47 | 17.62 | 36.07 | 64.40 | 90.97 |
| Germany | 3.22 | 5.76 | 8.14 | 13.71 | 24.48 | 34.58 | 3.67 | 6.55 | 9.25 | 35.39 | 63.19 | 89.26 |
| France | 19.31 | 34.47 | 48.70 | 2.48 | 4.43 | 6.25 | 3.07 | 5.48 | 7.74 | 31.14 | 55.60 | 78.54 |
| Spain | 25.01 | 44.66 | 63.09 | 8.09 | 14.46 | 20.42 | 0.13 | 0.23 | 0.33 | 22.75 | 40.63 | 57.39 |
| Italy | 4.99 | 8.91 | 12.59 | 3.18 | 5.69 | 8.03 | 0.32 | 0.57 | 0.81 | 47.50 | 84.82 | 119.80 |

***Table 2: Continued**: Variance Decomposition for the each monthly indicator, see table 4 and 5 for the name and definition. The table reports the median value of the percentage share of variance explained by each factor, along with the 33th and 66th percentiles of the sampling distribution of the statistic.*

| | **Common European Factor** | | | **Country specific Factors** | | | **Component specific Factors** | | | **Idiosyncratic** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1/3 | Med | 2/3 | 1/3 | Med | 2/3 | 1/3 | Med | 2/3 | 1/3 | Med | 2/3 |
| **Series: sts_inpr_msa** | | | | | | | | | | | | |
| EA | 17.81 | 31.81 | 44.93 | 12.64 | 22.57 | 31.88 | 25.52 | 45.57 | 64.36 | 0.02 | 0.04 | 0.06 |
| Germany | 16.07 | 28.69 | 40.53 | 17.48 | 31.21 | 44.09 | 22.39 | 39.98 | 56.47 | 0.05 | 0.10 | 0.14 |
| Spain | 0.00 | 0.00 | 0.00 | 5.91 | 10.56 | 14.92 | 2.36 | 4.22 | 5.96 | 47.72 | 85.20 | 120.34 |
| France | 3.79 | 6.76 | 9.56 | 13.52 | 24.14 | 34.10 | 16.22 | 28.96 | 40.91 | 22.46 | 40.11 | 56.66 |
| Italy | 2.44 | 4.35 | 6.15 | 6.47 | 11.56 | 16.33 | 25.76 | 45.99 | 64.96 | 21.32 | 38.08 | 53.78 |
| **Series: sts_trtu_defltv** | | | | | | | | | | | | |
| EA | 8.19 | 14.62 | 20.66 | 0.28 | 0.51 | 0.72 | 47.42 | 84.68 | 119.61 | 0.09 | 0.17 | 0.25 |
| Germany | 5.76 | 10.29 | 14.54 | 0.42 | 0.75 | 1.06 | 27.37 | 48.87 | 69.03 | 22.44 | 40.07 | 56.60 |
| Spain | 0.00 | 0.00 | 0.00 | 1.23 | 2.20 | 3.10 | 9.02 | 16.12 | 22.76 | 45.74 | 81.67 | 115.35 |
| France | 2.85 | 5.10 | 7.20 | 0.87 | 1.55 | 2.19 | 19.75 | 35.27 | 49.81 | 32.52 | 58.07 | 82.02 |
| Italy | 3.55 | 6.34 | 8.96 | 0.01 | 0.02 | 0.03 | 16.87 | 30.13 | 42.56 | 35.56 | 63.49 | 89.68 |
| **Series: sts_intv_m** | | | | | | | | | | | | |
| EA | 22.31 | 39.83 | 56.26 | 25.74 | 45.96 | 64.92 | 3.58 | 6.39 | 9.03 | 4.37 | 7.80 | 11.02 |
| Germany | 18.25 | 32.58 | 46.02 | 35.74 | 63.81 | 90.13 | 1.27 | 2.27 | 3.21 | 0.73 | 1.32 | 1.86 |
| Spain | 18.98 | 33.90 | 47.88 | 11.79 | 21.06 | 29.75 | 3.67 | 6.55 | 9.25 | 21.55 | 38.47 | 54.35 |
| France | 1.29 | 2.30 | 3.25 | 46.03 | 82.19 | 116.10 | 1.47 | 2.62 | 3.71 | 7.20 | 12.86 | 18.17 |
| Italy | 7.84 | 13.99 | 19.77 | 33.38 | 59.6 | 84.18 | 2.56 | 4.58 | 6.47 | 12.22 | 21.81 | 30.81 |
| **Series: Isno_m** | | | | | | | | | | | | |
| EA | 15.25 | 27.23 | 38.47 | 4.63 | 8.26 | 11.67 | 8.73 | 15.58 | 22.01 | 27.39 | 48.90 | 69.07 |
| Germany | 13.98 | 24.96 | 35.26 | 5.18 | 9.25 | 13.07 | 8.62 | 15.39 | 21.74 | 28.21 | 50.38 | 71.16 |
| Spain | 0.87 | 1.56 | 2.2 | 5.8 | 10.35 | 14.63 | 10.35 | 18.49 | 26.11 | 38.97 | 69.58 | 98.28 |
| France | 1.77 | 3.17 | 4.48 | 0.21 | 0.38 | 0.53 | 0.34 | 0.60 | 0.85 | 53.67 | 95.83 | 135.36 |
| Italy | 2.76 | 4.92 | 6.96 | 5.19 | 9.26 | 13.09 | 4.42 | 7.90 | 11.15 | 43.63 | 77.90 | 110.03 |
| **Series: bssi** | | | | | | | | | | | | |
| EA | 23.40 | 41.79 | 59.02 | 0.05 | 0.09 | 0.13 | 31.99 | 57.12 | 80.69 | 0.55 | 0.98 | 1.39 |
| Germany | 22.48 | 40.14 | 56.69 | 0.00 | 0.00 | 0.00 | 31.37 | 56.01 | 79.11 | 2.15 | 3.84 | 5.43 |
| Spain | 22.89 | 40.87 | 57.73 | 0.08 | 0.15 | 0.22 | 31.08 | 55.49 | 78.38 | 1.94 | 3.46 | 4.89 |
| France | 20.14 | 35.97 | 50.81 | 0.08 | 0.15 | 0.21 | 34.55 | 61.69 | 87.13 | 1.22 | 2.18 | 3.08 |
| Italy | 23.88 | 42.63 | 60.22 | 0.01 | 0.02 | 0.03 | 31.47 | 56.20 | 79.38 | 0.63 | 1.13 | 1.60 |
| **Series: bssiF** | | | | | | | | | | | | |
| EA | 55.42 | 98.95 | 139.76 | 0.01 | 0.03 | 0.04 | 0.45 | 0.81 | 1.15 | 0.11 | 0.19 | 0.28 |
| Germany | 51.57 | 92.09 | 130.07 | 0.00 | 0.00 | 0.01 | 0.53 | 0.95 | 1.35 | 3.88 | 6.93 | 9.80 |
| Spain | 53.84 | 96.13 | 135.79 | 0.02 | 0.04 | 0.05 | 0.57 | 1.03 | 1.46 | 1.56 | 2.78 | 3.93 |
| France | 54.52 | 97.35 | 137.50 | 0.00 | 0.00 | 0.00 | 0.43 | 0.77 | 1.09 | 1.04 | 1.87 | 2.64 |
| Italy | 55.33 | 98.80 | 139.55 | 0.01 | 0.03 | 0.04 | 0.45 | 0.81 | 1.14 | 0.19 | 0.35 | 0.50 |
| **Series: sts_cobp_m** | | | | | | | | | | | | |
| Germany | 1.44 | 2.58 | 3.64 | 0.00 | 0.00 | 0.01 | 0.30 | 0.55 | 0.78 | 54.24 | 96.85 | 136.80 |
| Spain | 0.00 | 0.00 | 0.00 | 0.49 | 0.88 | 1.24 | 0.62 | 1.12 | 1.58 | 54.88 | 97.99 | 138.41 |
| **Series: sts_colb_m** | | | | | | | | | | | | |
| Germany | 0.05 | 0.09 | 0.13 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 55.95 | 99.90 | 141.11 |
| **Series: sts_cohb_m** | | | | | | | | | | | | |
| Germany | 0.10 | 0.18 | 0.26 | 0.03 | 0.05 | 0.08 | 0.77 | 1.37 | 1.94 | 55.09 | 98.37 | 138.94 |
| **Series: sts_copr_m** | | | | | | | | | | | | |
| EA | 6.72 | 12.00 | 16.95 | 3.26 | 5.83 | 8.23 | 28.04 | 50.07 | 70.72 | 17.97 | 32.09 | 45.33 |
| Germany | 3.59 | 6.42 | 9.06 | 3.00 | 5.36 | 7.57 | 20.54 | 36.68 | 51.81 | 28.85 | 51.52 | 72.77 |
| Spain | 0.73 | 1.31 | 1.86 | 4.69 | 8.37 | 11.83 | 0.87 | 1.55 | 2.20 | 49.70 | 88.74 | 125.35 |
| France | 3.52 | 6.28 | 8.87 | 1.84 | 3.29 | 4.66 | 15.05 | 26.87 | 37.96 | 35.58 | 63.53 | 89.74 |
| **Series: ext_st_27msbec** | | | | | | | | | | | | |
| Germany | 3.89 | 6.94 | 9.81 | 0.49 | 0.87 | 1.23 | 1.54 | 2.75 | 3.88 | 50.08 | 89.42 | 126.30 |
| Spain | 0.16 | 0.29 | 0.42 | 3.75 | 6.70 | 9.47 | 2.65 | 4.73 | 6.68 | 49.43 | 88.25 | 124.66 |
| France | 1.69 | 3.01 | 4.26 | 0.78 | 1.40 | 1.98 | 3.96 | 7.07 | 9.99 | 49.56 | 88.50 | 125.00 |
| Italy | 16.91 | 30.20 | 42.66 | 3.82 | 6.82 | 9.64 | 1.50 | 2.69 | 3.80 | 33.75 | 60.27 | 85.13 |
| **Series: imt_st_27msbec** | | | | | | | | | | | | |
| Germany | 4.34 | 7.75 | 10.95 | 0.38 | 0.69 | 0.98 | 0.00 | 0.00 | 0.00 | 51.27 | 91.54 | 129.30 |
| Spain | 0.15 | 0.27 | 0.38 | 0.00 | 0.00 | 0.00 | 0.53 | 0.95 | 1.35 | 55.32 | 98.77 | 139.51 |
| France | 9.92 | 17.72 | 25.03 | 0.82 | 1.46 | 2.07 | 0.49 | 0.87 | 1.24 | 44.76 | 79.92 | 112.89 |
| Italy | 0.14 | 0.26 | 0.37 | 4.95 | 8.84 | 12.49 | 1.49 | 2.66 | 3.77 | 49.41 | 88.22 | 124.6 |
| **Series: mny_stk_spy_m** | | | | | | | | | | | | |
| EA | 27.25 | 48.66 | 68.73 | 0.05 | 0.10 | 0.14 | 28.45 | 50.80 | 71.75 | 0.24 | 0.43 | 0.61 |
| Germany | 20.23 | 36.13 | 51.03 | 0.01 | 0.03 | 0.04 | 31.86 | 56.89 | 80.36 | 3.88 | 6.94 | 9.80 |
| Spain | 28.14 | 50.25 | 70.98 | 0.00 | 0.01 | 0.02 | 21.38 | 38.18 | 53.93 | 6.46 | 11.54 | 16.30 |
| France | 27.95 | 49.91 | 70.50 | 0.00 | 0.00 | 0.00 | 27.02 | 48.25 | 68.15 | 1.02 | 1.82 | 2.57 |
| Italy | 27.35 | 48.83 | 68.97 | 0.01 | 0.02 | 0.03 | 25.08 | 44.77 | 63.24 | 3.56 | 6.36 | 8.98 |
| **Series: mny_stk_mcp_m** | | | | | | | | | | | | |
| EA | 24.59 | 43.90 | 62.01 | 5.24 | 9.36 | 13.22 | 21.80 | 38.92 | 54.98 | 4.36 | 7.79 | 11.01 |
| Germany | 18.36 | 32.78 | 46.31 | 0.00 | 0.00 | 0.00 | 19.30 | 34.47 | 48.68 | 18.33 | 32.73 | 46.23 |
| Spain | 17.03 | 30.41 | 42.95 | 2.50 | 4.46 | 6.31 | 16.66 | 29.75 | 42.02 | 19.80 | 35.36 | 49.95 |
| France | 27.98 | 49.96 | 70.57 | 0.59 | 1.06 | 1.50 | 15.73 | 28.09 | 39.67 | 11.69 | 20.87 | 29.49 |
| Italy | 24.21 | 43.24 | 61.07 | 0.94 | 1.68 | 2.38 | 15.34 | 27.40 | 38.70 | 15.49 | 27.66 | 39.07 |
| **Series: mfma_m** | | | | | | | | | | | | |
| EA | 1.66 | 2.97 | 4.20 | 0.18 | 0.32 | 0.45 | 0.06 | 0.12 | 0.17 | 54.08 | 96.57 | 136.40 |
| Germany | 1.29 | 2.30 | 3.25 | 0.22 | 0.39 | 0.56 | 0.29 | 0.53 | 0.75 | 54.19 | 96.76 | 136.67 |
| Spain | 0.04 | 0.08 | 0.11 | 0.70 | 1.25 | 1.76 | 0.45 | 0.81 | 1.15 | 54.8 | 97.84 | 138.20 |
| France | 0.69 | 1.23 | 1.74 | 0.11 | 0.20 | 0.28 | 0.01 | 0.02 | 0.02 | 55.18 | 98.53 | 139.17 |
| Italy | 4.07 | 7.28 | 10.28 | 0.02 | 0.04 | 0.05 | 0.05 | 0.10 | 0.14 | 51.84 | 92.57 | 130.75 |

***Table 2: Continued**: Variance Decomposition for the each monthly indicator, see table 4 and 5 for the name and definition. The table reports the median value of the percentage share of variance explained by each factor, along with the 33th and 66th percentiles of the sampling distribution of the statistic.*

| | **Common European Factor** | | | **Country specific Factors** | | | **Component specific Factors** | | | **Idiosyncratic** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $1/3$ | Med | $2/3$ | $1/3$ | Med | $2/3$ | $1/3$ | Med | $2/3$ | $1/3$ | Med | $2/3$ |
| **Series: bssi_GCF** | | | | | | | | | | | | |
| EA | 27.64 | 49.35 | 69.71 | 0.19 | 0.34 | 0.48 | 0.00 | 0.00 | 0.00 | 28.16 | 50.29 | 71.03 |
| Germany | 24.12 | 43.06 | 60.83 | 0.07 | 0.13 | 0.18 | 0.02 | 0.04 | 0.06 | 31.78 | 56.75 | 80.15 |
| Spain | 25.76 | 46.00 | 64.97 | 0.77 | 1.39 | 1.96 | 0.18 | 0.33 | 0.47 | 29.27 | 52.26 | 73.82 |
| France | 12.07 | 21.56 | 30.46 | 0.73 | 1.31 | 1.86 | 0.28 | 0.51 | 0.73 | 42.90 | 76.59 | 108.18 |
| Italy | 22.65 | 40.44 | 57.13 | 0.17 | 0.30 | 0.43 | 0.17 | 0.31 | 0.44 | 33.00 | 58.92 | 83.22 |
| **Series: sts_intvd_m** | | | | | | | | | | | | |
| EA | 11.00 | 19.64 | 27.75 | 25.21 | 45.02 | 63.58 | 0.12 | 0.22 | 0.32 | 19.66 | 35.10 | 49.58 |
| Germany | 15.15 | 27.06 | 38.23 | 33.10 | 59.10 | 83.48 | 0.30 | 0.54 | 0.77 | 7.43 | 13.27 | 18.75 |
| Spain | 8.52 | 15.21 | 21.48 | 8.70 | 15.53 | 21.94 | 1.02 | 1.83 | 2.58 | 37.75 | 67.41 | 95.22 |
| France | 0.83 | 1.49 | 2.10 | 42.59 | 76.04 | 107.41 | 0.28 | 0.51 | 0.72 | 12.29 | 21.94 | 30.99 |
| Italy | 0.00 | 0.00 | 0.00 | 26.46 | 47.26 | 66.75 | 0.01 | 0.02 | 0.02 | 29.52 | 52.71 | 74.45 |
| **Series: sts_rt_careg** | | | | | | | | | | | | |
| EA | 0.59 | 1.05 | 1.48 | 2.28 | 4.08 | 5.76 | 51.06 | 91.17 | 128.77 | 2.06 | 3.69 | 5.21 |
| Germany | 0.92 | 1.65 | 2.34 | 0.22 | 0.40 | 0.56 | 13.35 | 23.84 | 33.68 | 41.49 | 74.09 | 104.65 |
| Spain | 0.18 | 0.33 | 0.47 | 2.52 | 4.50 | 6.36 | 9.24 | 16.50 | 23.31 | 44.05 | 78.65 | 111.09 |
| France | 1.61 | 2.87 | 4.06 | 0.07 | 0.13 | 0.18 | 12.94 | 23.11 | 32.64 | 41.37 | 73.88 | 104.35 |
| Italy | 7.31 | 13.05 | 18.43 | 0.74 | 1.32 | 1.87 | 13.13 | 23.44 | 33.11 | 34.82 | 62.17 | 87.81 |
| **Series: bssi_FCE** | | | | | | | | | | | | |
| EA | 15.38 | 27.47 | 38.80 | 1.07 | 1.91 | 2.70 | 1.76 | 3.14 | 4.44 | 37.78 | 67.46 | 95.29 |
| Germany | 8.64 | 15.42 | 21.79 | 0.10 | 0.19 | 0.27 | 0.22 | 0.40 | 0.56 | 47.03 | 83.97 | 118.61 |
| Spain | 6.90 | 12.33 | 17.42 | 0.08 | 0.14 | 0.20 | 4.32 | 7.71 | 10.90 | 44.69 | 79.80 | 112.71 |
| France | 3.24 | 5.79 | 8.18 | 0.46 | 0.82 | 1.17 | 0.55 | 0.99 | 1.40 | 51.74 | 92.38 | 130.48 |
| Italy | 8.78 | 15.68 | 22.15 | 0.26 | 0.46 | 0.66 | 0.24 | 0.44 | 0.62 | 46.71 | 83.40 | 117.80 |
| **Series: bsco_FCE** | | | | | | | | | | | | |
| EA | 7.47 | 13.33 | 18.84 | 1.67 | 2.99 | 4.22 | 45.49 | 81.23 | 114.74 | 1.36 | 2.43 | 3.43 |
| Germany | 2.26 | 4.03 | 5.70 | 1.32 | 2.37 | 3.35 | 21.00 | 37.49 | 52.96 | 31.41 | 56.09 | 79.22 |
| Spain | 5.99 | 10.70 | 15.11 | 0.42 | 0.76 | 1.08 | 11.07 | 19.77 | 27.92 | 38.51 | 68.75 | 97.11 |
| France | 2.03 | 3.63 | 5.13 | 0.00 | 0.00 | 0.01 | 11.97 | 21.38 | 30.20 | 41.98 | 74.97 | 105.89 |
| Italy | 0.12 | 0.22 | 0.31 | 0.66 | 1.18 | 1.67 | 8.66 | 15.47 | 21.85 | 46.55 | 83.11 | 117.40 |
| **Series: sts_trtu_tv** | | | | | | | | | | | | |
| EA | 9.59 | 17.13 | 24.20 | 2.30 | 4.10 | 5.80 | 39.79 | 71.04 | 100.35 | 4.31 | 7.70 | 10.88 |
| Germany | 1.47 | 2.62 | 3.70 | 0.35 | 0.63 | 0.89 | 24.02 | 42.88 | 60.57 | 30.16 | 53.85 | 76.06 |
| Spain | 14.84 | 26.50 | 37.43 | 4.62 | 8.25 | 11.66 | 3.77 | 6.73 | 9.51 | 32.76 | 58.50 | 82.63 |
| France | 4.89 | 8.74 | 12.35 | 4.12 | 7.36 | 10.39 | 15.93 | 28.44 | 40.18 | 31.05 | 55.44 | 78.31 |
| Italy | 5.30 | 9.47 | 13.37 | 0.00 | 0.01 | 0.02 | 18.95 | 33.84 | 47.81 | 31.73 | 56.66 | 80.03 |
| **Series: sts_trlb_m** | | | | | | | | | | | | |
| Germany | 0.24 | 0.43 | 0.61 | 4.10 | 7.32 | 10.34 | 0.51 | 0.91 | 1.29 | 51.14 | 91.32 | 128.98 |
| **Series: sbs_ciGHI** | | | | | | | | | | | | |
| EA | 0.92 | 1.64 | 2.32 | 1.57 | 2.81 | 3.97 | 4.51 | 8.06 | 11.39 | 48.99 | 87.47 | 123.54 |
| Germany | 1.60 | 2.87 | 4.05 | 0.19 | 0.35 | 0.49 | 4.41 | 7.87 | 11.12 | 49.78 | 88.89 | 125.56 |
| Spain | 1.63 | 2.91 | 4.11 | 0.03 | 0.05 | 0.08 | 1.41 | 2.53 | 3.57 | 52.92 | 94.49 | 133.47 |
| France | 0.95 | 1.69 | 2.39 | 2.14 | 3.82 | 5.40 | 0.89 | 1.60 | 2.26 | 52.01 | 92.86 | 131.17 |
| Italy | 1.03 | 1.84 | 2.60 | 0.06 | 0.12 | 0.17 | 0.54 | 0.96 | 1.36 | 54.36 | 97.06 | 137.10 |
| **Series: apro_mk_colm_milk** | | | | | | | | | | | | |
| EA | 4.15 | 7.41 | 10.47 | 0.00 | 0.00 | 0.00 | 4.06 | 7.26 | 10.25 | 47.78 | 85.32 | 120.51 |
| Germany | 9.94 | 17.75 | 25.07 | 0.07 | 0.12 | 0.18 | 0.86 | 1.54 | 2.18 | 45.12 | 80.57 | 113.80 |
| Spain | 0.00 | 0.01 | 0.01 | 0.34 | 0.61 | 0.86 | 2.71 | 4.84 | 6.84 | 52.94 | 94.53 | 133.52 |
| France | 1.27 | 2.27 | 3.20 | 0.20 | 0.36 | 0.51 | 2.36 | 4.22 | 5.96 | 52.16 | 93.14 | 131.56 |
| Italy | 3.03 | 5.41 | 7.64 | 0.08 | 0.15 | 0.21 | 0.93 | 1.67 | 2.35 | 51.95 | 92.76 | 131.02 |
| **Series: apro_ec_pwgtm_pigs** | | | | | | | | | | | | |
| Germany | 1.65 | 2.94 | 4.16 | 0.03 | 0.05 | 0.07 | 32.64 | 58.27 | 82.31 | 21.68 | 38.71 | 54.68 |
| Spain | 0.69 | 1.24 | 1.75 | 0.20 | 0.37 | 0.52 | 19.41 | 34.66 | 48.95 | 35.69 | 63.72 | 90.00 |
| France | 0.07 | 0.13 | 0.19 | 0.06 | 0.11 | 0.16 | 40.21 | 71.79 | 101.40 | 15.65 | 27.95 | 39.48 |
| Italy | 0.58 | 1.04 | 1.47 | 0.04 | 0.07 | 0.10 | 15.49 | 27.67 | 39.08 | 39.88 | 71.20 | 100.58 |
| **Series: apro_ec_pwgtm_bov** | | | | | | | | | | | | |
| EA | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 51.94 | 92.73 | 130.98 | 4.06 | 7.25 | 10.25 |
| Germany | 0.41 | 0.73 | 1.03 | 0.05 | 0.10 | 0.14 | 42.30 | 75.53 | 106.69 | 13.23 | 23.62 | 33.37 |
| Spain | 0.98 | 1.74 | 2.47 | 0.01 | 0.03 | 0.04 | 25.88 | 46.22 | 65.28 | 29.12 | 51.99 | 73.44 |
| France | 0.02 | 0.05 | 0.07 | 0.13 | 0.23 | 0.33 | 0.15 | 0.26 | 0.38 | 55.69 | 99.44 | 140.45 |
| Italy | 0.74 | 1.33 | 1.88 | 0.13 | 0.24 | 0.34 | 0.38 | 0.68 | 0.96 | 54.74 | 97.73 | 138.05 |
| **Series: lmhu_m** | | | | | | | | | | | | |
| EA | 1.68 | 3.00 | 4.24 | 1.62 | 2.89 | 4.08 | 0.39 | 0.70 | 0.99 | 52.30 | 93.39 | 131.91 |
| Germany | 0.05 | 0.10 | 0.14 | 0.00 | 0.01 | 0.02 | 0.00 | 0.00 | 0.00 | 55.93 | 99.87 | 141.06 |
| Spain | 0.00 | 0.00 | 0.00 | 0.03 | 0.07 | 0.10 | 0.05 | 0.10 | 0.14 | 55.91 | 99.82 | 140.99 |
| France | 0.71 | 1.28 | 1.81 | 0.09 | 0.16 | 0.23 | 0.04 | 0.08 | 0.12 | 55.14 | 98.45 | 139.06 |
| Italy | 2.92 | 5.21 | 7.36 | 5.44 | 9.72 | 13.73 | 0.04 | 0.08 | 0.12 | 47.59 | 84.96 | 120.01 |

error, (ME$_i$, the average of the 28 prediction errors), the prediction error variance, $V_i$, the mean square forecast error,

$$\text{MSFE}_i = \frac{1}{M}\sum_{\tau=1}^{M}\left(Y_{i\tau} - \hat{Y}_{i\tau}\right)^2 = \text{ME}_i^2 + V_i,$$

the mean absolute error (the average of $|Y_{i\tau} - \hat{Y}_{i\tau}|$) and the mean absolute percentage error,

$$\text{MAPE}_i = \frac{1}{M}\sum_{\tau=1}^{M}\frac{|Y_{i\tau} - \hat{Y}_{i\tau}|}{Y_{i\tau}}.$$

The values of the above statistics are reported in table 3, respectively for RW, EuroMInd and EuroMInd-C, for the 11 GDP components and for the combined estimate of GDP (last column). The table also reports the MSFE ratio using EuroMInd-C as the numéraire (values greater than 1 denoting that the forecasts are less efficient than EuroMInd-C), as well as the Diebold-Mariano-West test of equal forecasting accuracy. Denoting by $e_{i\tau}^{(k)} = Y_{i\tau} - \hat{Y}_{i\tau}^{(k)}$ the forecast error of predictor $(k)$, where $(k)$ is RW or EuroMInd, by $e_{i\tau}^{(Ec)}$ that of EuroMind-C, and $d_{i\tau} = e_{i\tau}^{2(k)} - e_{i\tau}^{2(Ec)}$ the quadratic loss differential, the Diebold-Mariano-West test of the null hypothesis of equal forecast accuracy, $H_0$ : E$(d_{i\tau}) = 0$, versus the one sided alternative $H_1$ : E$(d_{i\tau}) > 0$, is the test statistic

$$DMW_i = \frac{\bar{d}_i}{\sqrt{\sigma_i^2}}, \quad \bar{d}_i = \frac{1}{M}\sum_{\tau=1}^{M} d_{i\tau}, \quad \sigma_i^2 = \frac{1}{M}\left[c_0 + 2\sum_{k=1}^{q-1}\frac{q-k}{q}c_k\right],$$

where $c_k$ is the sample autocovariance of $d_{i\tau}$ at lag $k$ and $\sigma_i^2$ is a consistent estimate of the variance of the loss differential. See **?** and **?**. We set $q = 3$ and we adopt the DMW statistic with the small sample modification proposed by **?**, which corrects for the bias of $\sigma_i^2$ as an estimator of the variance of $d_j(h)$:

$$DMW_i^* = DMW_i\left[\frac{J + 1 - 2q + q(q+1)/J}{J}\right]^{1/2}.$$

The null distribution of the test is Student's $t$ with $M - 1$ degrees of freedom.

The results can be summarised as follows:

- The global evidence is that EuroMInd-C produces the most accurate forecasts of the components and GDP: in particular the forecasts are characterised by less variability and a lower mean absolute percentage error. The relative accuracy (as measured by the MSFE ratio) is greater than one for most components of GDP, with the exception of the value added of agriculture (A-B), for which the RW predictor is the most accurate, whereas EuroMInd produces more accurate forecasts of the value added of the sector G-H-I, and for Final Consumption Expenditures and Gross Capital Formation. For combined GDP at market prices EuroMInd is 7% less accurate than EuroMInd-C.

- The expenditure components, and in particular GCF, IMP and EXP, are much more difficult to predict; this is reflected in the forecast error variances which are very large for all the predictors.

***Table 3:*** *Comparative assessment of the predictive performance of EuroMInd-C and EuroMInd in real time. The RW entry refers to the random walk or "no change" predictor*

Mean Error

| *Models* | A-B | C-D-E | F | G-H-I | J-K | L-P | TlS | FCE | GCF | EXP | IMP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RW | 0.10 | 0.50 | -0.30 | 0.81 | 1.31 | 2.07 | 0.03 | 3.37 | -0.36 | 7.24 | 5.81 | 4.50 |
| EuroMInd | -0.18 | 0.20 | -0.33 | -0.55 | -0.02 | 0.05 | -0.50 | -2.25 | -2.02 | -0.15 | -1.58 | -1.56 |
| EuroMInd-C | -0.16 | -0.83 | -0.52 | -1.38 | -0.29 | -1.28 | -0.71 | -3.31 | -3.38 | -5.03 | -5.78 | -5.32 |

Variance

| *Models* | A-B | C-D-E | F | G-H-I | J-K | L-P | TlS | FCE | GCF | EXP | IMP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RW | 0.21 | 78.59 | 1.86 | 8.90 | 2.03 | 9.03 | 4.41 | 20.81 | 147.87 | 521.17 | 382.39 | 265.26 |
| EuroMInd | 0.55 | 40.12 | 1.77 | 5.77 | 1.89 | 9.51 | 4.52 | 16.54 | 109.56 | 308.84 | 264.15 | 129.23 |
| EuroMInd-C | 0.47 | 22.44 | 1.42 | 5.55 | 1.72 | 7.00 | 3.85 | 17.34 | 109.79 | 277.83 | 211.85 | 94.26 |

Mean Square Forecast Error

| *Models* | A-B | C-D-E | F | G-H-I | J-K | L-P | TlS | FCE | GCF | EXP | IMP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RW | 0.22 | 78.84 | 1.95 | 9.55 | 3.73 | 13.34 | 4.41 | 32.18 | 148.00 | 573.58 | 416.14 | 285.53 |
| EuroMInd | 0.58 | 40.16 | 1.88 | 6.07 | 1.89 | 9.52 | 4.77 | 21.62 | 113.64 | 308.86 | 266.63 | 131.65 |
| EuroMInd-C | 0.49 | 23.13 | 1.69 | 7.45 | 1.81 | 8.64 | 4.35 | 28.29 | 121.20 | 303.18 | 245.30 | 122.61 |

Mean Absolute Error

| *Models* | A-B | C-D-E | F | G-H-I | J-K | L-P | TlS | FCE | GCF | EXP | IMP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RW | 0.34 | 5.97 | 1.22 | 2.57 | 1.41 | 3.01 | 1.51 | 4.66 | 9.27 | 18.43 | 16.29 | 13.37 |
| EuroMInd | 0.46 | 3.77 | 1.14 | 1.69 | 1.10 | 2.35 | 1.62 | 3.94 | 8.42 | 12.02 | 11.58 | 7.50 |
| EuroMInd-C | 0.41 | 2.73 | 1.04 | 1.83 | 0.97 | 2.00 | 1.46 | 4.08 | 8.44 | 9.98 | 10.77 | 6.90 |

Mean Absolute Percentage Error

| *Models* | A-B | C-D-E | F | G-H-I | J-K | L-P | TlS | FCE | GCF | EXP | IMP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RW | 0.90 | 1.70 | 1.30 | 0.77 | 0.79 | 0.43 | 0.81 | 0.32 | 2.38 | 2.22 | 2.03 | 0.71 |
| EuroMInd | 1.25 | 1.07 | 1.23 | 0.50 | 0.61 | 0.33 | 0.87 | 0.27 | 2.17 | 1.44 | 1.45 | 0.40 |
| EuroMInd-C | 1.12 | 0.77 | 1.12 | 0.55 | 0.53 | 0.28 | 0.78 | 0.28 | 2.18 | 1.20 | 1.35 | 0.36 |

Relative MSFE (with respect to EuroMInd-C)

| *Models* | A-B | C-D-E | F | G-H-I | J-K | L-P | TlS | FCE | GCF | EXP | IMP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RW | 0.45 | 3.41 | 1.16 | 1.28 | 2.06 | 1.54 | 1.01 | 1.14 | 1.22 | 1.89 | 1.70 | 2.33 |
| EuroMInd | 1.18 | 1.74 | 1.11 | 0.82 | 1.04 | 1.10 | 1.10 | 0.76 | 0.94 | 1.02 | 1.09 | 1.07 |

Diebold-Mariano-West Statistic (versus EuroMInd-C)

| *Models* | A-B | C-D-E | F | G-H-I | J-K | L-P | TlS | FCE | GCF | EXP | IMP | GDP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RW | 0.62 | 1.48 | 0.91 | 1.16 | 1.82 | 1.29 | -0.20 | 0.44 | 1.33 | 2.80 | 2.88 | 2.38 |
| EuroMInd | 0.67 | 1.75 | 0.64 | -1.54 | 0.24 | 0.92 | 0.97 | -1.61 | -0.79 | 0.10 | 0.74 | 0.90 |

- The Diebold-Mariano-West tests of equal forecast accuracy are never significant as far as the comparison of EuroMInd and EuroMInd-C is concerned.

- EuroMInd-C significantly outperforms the RW predictor: the Diebold-Mariano-West test leads to a strong rejection for the combined GDP prediction and some GDP components.

## 10.1 Parameter instability

The recursive forecasting experiment enables us to assess the variability of the parameter estimates according to the sample period considered. Let $\tilde{\Xi}_\tau$ denote the parameter estimates obtained in quarter $\tau$, $\tau = 2005.q1, \dots, 2011.q4$: a measure of parameter instability with estimation period $\tau'$ should be based on the Mahalanobis distance between the vectors $\tilde{\Xi}_\tau$ and $\tilde{\Xi}_{\tau'}$, for $\tau' \neq \tau$. The latter is a weighted Euclidean distance, where the weighting matrix is the averagage precision matrix,

$$d(\tau, \tau') = \left(\tilde{\Xi}_\tau - \tilde{\Xi}_{\tau'}\right)' \left[\text{Var}(\tilde{\Xi})\right]^{-1} \left(\tilde{\Xi}_\tau - \tilde{\Xi}_{\tau'}\right).$$

This measure requires the precision matrix $\left[\text{Var}(\tilde{\Xi})\right]^{-1}$, which is not made available by the EM algorithm.

However, a descriptive assessment of parameter instability can be obtained from the spectral decomposition of the correlation matrix of the parameter estimates $\tilde{\Xi}_\tau, \tau = 2005.q1, \dots, 2011.q4$, by plotting the first two eigenvectors scaled by the square root of the corresponding eigenvalue. This amounts to performing a metric scaling analysis of the matrix stacking the vectors $\tilde{\Xi}_\tau'$ (see **?**).

This is done in figure 7, where each point represents the parameter estimate at a particular time point and the Euclidean distance between the points is the best rank 2 approximation to the Mahalanobis distance $d(\tau, \tau')$. The approximation is good as the first two eigenvalues account for $98\%$ of the total original variation.

The most prominent fact presented by the figure is the rather large shift in the parameter estimates occurring with the release $\tau = 2009.q2$, which uses the data up to May 2009. It ought to be mentioned the real effects of the global financial crisis hit the Euro area around the end of 2008; thus, the main factor driving the parameter change along the vertical dimension of the graph is the strong recessionary movement resulting from the global financial crisis. A second important fact is that a common factor (determining the horizontal dimension of the plot and accounting for $94\%$ of the total variation) drives the change in the parameters.

We did not address the important issue of assessing whether the estimation of structural change via the inclusion of suitable interventions helps improving the accuracy of EuroMInd-C in forecasting and signal extraction, along the lines indicated by **?**. This represents an interesting topic for further research in this area.

## 10.2 Comparison with other coincident indicators of economic growth

We have already compared EuroMInd-C with EuroMInd in terms of predictive accuracy. Another consideration is that EuroMInd-C provides monthly inidicators of the 11 GDP components for 4 countries along with the Euro area; furthermore, within the estimation sample the estimated components are more informative as they embody more information. This shows up in a higher

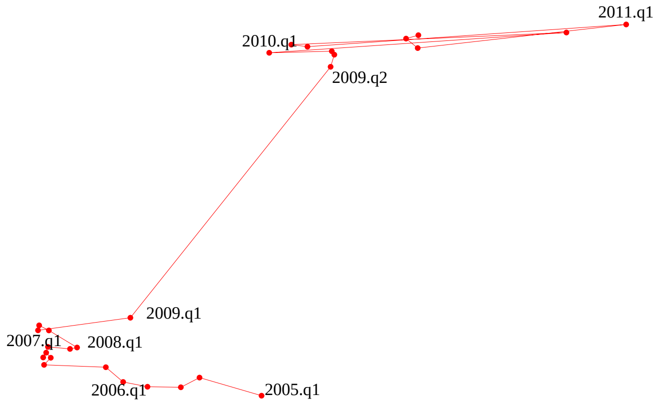


***Figure 7:*** *Metric scaling analysis of recursive parameter estimates. Each point represents the parameter estimate $\tilde{\Xi}_\tau$ obtained from the release available at quarter $\tau = 2005.q1, \ldots, 2011.q4$.*

reliability of the smoothed, or historical, estimates as reflected by a lower estimation standard error.

Several alternative monthly indicators of economic activity are constructed with similar objectives to EuroMInd-C; this subsection provides a qualitative assessment of their different information content. Among these indicators New Eurocoin (NE) and the Economic Sentiment Index (ESI) are prominent.

New Eurocoin is a monthly coincident indicator of economic growth for the Euro area (**?**) published monthly by CEPR (www.cepr.org) and the Bank of Italy. The indicator is a measure of underlying quarterly GDP growth, devoid of short run fluctuations (those with periodicity less than one year), based on its linear projection on the space of the relevant dynamic principal components, constructed on a set of about 150 transformed monthly time series.

The first panel of figure 8 compares NE with the quarterly logarithmic changes of EuroMInd-C. There are several differences:

1. NE is smoother than EuroMInd-C. This is so since EuroMInd-C does not impose smoothness priors.

2. NE is an estimate of GDP underlying growth, devoid of short-run fluctuations with periodicity less than one year (medium to long run growth) assuming a quarterly horizon for growth. EuroMInd-C does not assume a particular horizon and it can be used for the assessment of growth over any horizon.

3. The amplitude of the fluctuations and the characterisation of the quarterly growth rate cycle differ; NE ascribes a larger cyclical amplitude to the fluctuations before the global financial crisis, and the sharpness of the trough and the subsequent recovery during 2009 is very different.

4. NE does not enforce the temporal aggregation constraints and uses a different interpolation scheme. EuroMInd-C estimates the monthly level of GDP that is consistent with the quarterly totals.

5. There are other methodological differences and the elementary series considered differ in part. For instance, in the construction of EuroMInd-C the consumer and producer price series do not play a role.

The European Commission (Directorate-General for Economic and Financial Affairs) compiles the Economic Sentiment Indicator (ESI), a composite coincident indicator for the timely assessment of socio-economic situation in the Euro area. ESI is computed by aggregating the seasonally adjusted balances of opinions to a selection of qualitative survey questions from five sectors (industry, services, construction, retail trade and consumers) related to reference ”hard” variables, such as production in industry.

We have found that the most appropriate comparison is with the yearly changes of EuroMInd-C, which are displayed in the second panel of figure 8 and compared to the rescaled ESI. Despite the apparent similarity of the series, there are notable differences in the location of peaks and troughs of the annual growth rate cycle and in the speed of the recovery occurring around the years 2009-2010. In particular, ESI provided a much more optimistic assessment of the recovery than it shows in the monthly GDP estimates. While the comparison seems to establish that ESI tracks economic growth at a yearly horizon, we should mention that EuroMInd-C embodies a supervised approach that links the indicator to the national accounts series. On the contrary, the ESI methodology is ’unsupervised’, in that only the balances of opinions are used for the estimation of the indicator. As a result the interpretation of the scale of the indicator not related to GDP growth.

This qualitative assessment, coupled with the considerations arising from the rolling forecast evaluation, suggests that EuroMInd-C is a useful addition to the set of coincident indicators available for the Euro area and its main economies.

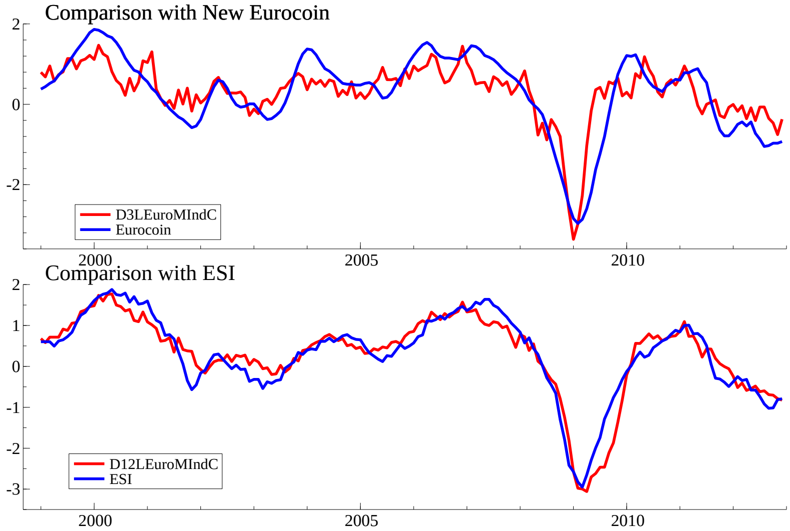


***Figure 8:*** *Comparison of EuroMInd with New Eurocoin and the Economic Sentiment Indicator.*

## 11 Conclusions

In this paper we have introduced EuroMInd-C, a set of monthly indicators of GDP for the Euro Area and its four largest economies that are relevant, representative and timely. The estimates can be used to assess the economic stance in a timely manner and allow to appreciate the broad picture (Euro Area) as well as the country specific and sector specific details.

The characteristic element of EuroMInd-C, which distinguishes it from other economic indicators based on dynamic factor models, is the centrality of the national accounts framework, which provides the backbone of the system, but also adds complexity to it.

We have illustrated the methodological issues that arise when estimating a large scale dynamic factor model subject to temporal aggregation constraints combined with generic patterns of missing observations, and we have documented the technical solutions that can be adopted in the state space framework.

For a real time implementation, batch estimation of the resulting mixed frequency factor model is computationally demanding, albeit feasible. Hence, we envisage a strategy that performs the estimation of the hyperparameters once every semester or every year, whereas filtering and smoothing can be performed in real time each time the dataset is updated.

A pseudo real-time recursive forecasting experiment has evidenced that EuroMInd-C can be more accurate for predicting GDP than our previous componentwise approach implemented in the EuroMInd indicator (see **?**). Moreover, the comparative assessment with other coincident indicators suggests that EuroMInd-C is a useful addition to the set of coincident indicators available for the Euro area and its main economies.

To conclude, we note that the methodology implemented in this paper can be also used in other contexts that require modelling and summarizing the information in multi-dimensional unbalanced datasets subject to sectional and temporal aggregation constraints.

## Appendix A: List of Series in the Dataset

The time series used for the estimation of EuroMInd-C were downloaded from the Europa dataset made available electronically by Eurostat. The series of Value added and GDP components are published with 65 days of delay respect to the end of each quarter accordingly to a release calendar available at the begin of each year on the web site of Eurostat. However a preliminary estimate of GDP at market prices, the so-called flash estimate, is released in advanced (45 days from the end of each quarter) with the aim of promptly reveal the tendency of the economic activity. Among monthly indicators the timeliest series are Business and Consumer survey data which are made available at the end of each month (around 28 days of delay) along with financial aggregates by ECB (around 30 days of delay) whereas data for Industry sector and retail (such as Car registration by OECD) are published about 45 days after the end of the reference period. Other indicators, such as indicators for the Construction (Index of production and Building permits), series for Agricultural sector and data on the labour market (Employment and Hours worked) show a publication delay of about 70 days. Finally, data on external trade, given the complexity of the collection data, arrives with a consistent delay, which amounts at about 90 days for the Index of Exports and Imports volumes by Eurostat.

Table 4 provides the complete list by GDP component, using the label of the Europa dataset. It also reports the time span of the series and the number of countries available for each series. Finally the Euro area is the one defined by 17 countries.

***Table 4:*** *National accounts series and monthly series by sector of activity and GDP component, output side. For each series are reported: the mnemonic name for the Eurostat database (`http://epp.eurostat.ec.europa.eu/`), the frequency, sample period and number of countries available. The available countries are: Euro area (EA), Germany (G), Spain (S), France (F) and Italy (I). In case of different sample period the name of the series and the sample period are also reported. If a dataset different from Eurostat is used it is reported between brackets.*

**Sector A-B: Agriculture**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_nace06_AB | Value added A-B | q | 1995.q1-2012.q2 | 5 |
| apro_mk_colm_sa_milk | Production of milk | m | 1995.m1-2012.m6 | 5 |
| apro_ec_pwgtm_sa_pigs | Pigs meat production in tons | m | 1995.m1-2011.m12 | 4:G, S, F, I |
| apro_ec_pwgtm_sa_bovines | Bovine meat production in tons | m | 1995.m1-2012.m6 | 5 |

**Sector C-D-E: Industry**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_nace06_k | Value added C-D-E | q | 1995.q1-2012.q2 | 5 |
| sts_inpr_msa | Index of industrial production | m | 1995.m1-2012.m6 | 5 |
| sts_intvd_m | Index of industrial turnover | m | 1996.m1-2012.m6 | 5 |
| | Spain | m | 2002.m1-2012.m6 | |
| | Italy | m | 2000.m1-2012.m6 | |
| isno_m | New Orders, Industry | m | 1995.m1-2012.m4 | 5 |
| | Spain | m | 2002.m1-2012.m4 | |
| | France | m | 2000.m1-2012.m4 | |
| bssi | Confidence Indicator, Industry | m | 1995.m1-2012.m6 | 5 |

**Sector F: Constructions**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_nace06_k | Value added F | q | 1995.q1-2012.q2 | 5 |
| sts_copr_m | Volume Index of production | m | 1995.m1-2012.m6 | 4:EA, G, S, F |
| sts_cohb_m | Hours worked | m | 1995.m1-2010.m6 | 1: G |
| sts_colb_m | Employment | m | 1995.m1-2010.m6 | 1: G |
| sts_cobp_m | Building permits | m | 1995.m1-2012.m6 | 2:G,S |

**Sector G-H-I: Trade, transport and communication services**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_nace06_k | Value added G-H-I | q | 1995.q1-2012.q2 | 5 |
| sts_trtu_tv | Index of turnover | m | 1995.m1-2012.m6 | 5 |
| | Italy | m | 2000.m1-2012.m6 | |
| sts_trtu_defltv | Index of deflated turnover | m | 1996.m1-2012.m6 | 5 |
| | Italy | m | 1999.m1-2011.m8 | |
| sts_trlb_m | Employment | m | 1995.m1-2012.m6 | 1: G |
| sts_rt_careg | Car registrations | m | 1995.m1-2011.m9 | 5 (OECD) |
| bssiF | Confidence Indicator, Construction | m | 1995.m1-2012.m6 | 5 |
| bs_ciGHI | Confidence Indicator, Retail | m | 1996.m1-2012.m6 | 5 |

**Sector J-K: Financial services and business activities**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_nace06_k | Value added J-K | q | 1995.q1-2012.q2 | 5 |
| sts_inpr_msa | Index of industrial production | m | 1995.m1-2012.m6 | 5 |
| mny_stk_spy_m | Share price indices | m | 1995.m1-2012.m6 | 5 |
| | Italy | m | 1998.m1-2012.m6 | |
| mny_stk_mcp_m | Euro area monetary aggregates | m | 1995.m1-2011.m9 | 5 |
| | France | m | 1995.m1-2006.m6 | |
| | Italy | m | 1995.m1-2009.m8 | |
| mfma_m | Tot. dep. of residents held at MFIs | m | 2000.m1-2011.m9 | 5 |

**Sector L-P: Other services**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_nace06_k | Value added LP | q | 1995.q1-2012.q2 | 5 |
| sts_inpr_msa | Index of industrial production | m | 1995.m1-2012.m6 | 5 |
| lmhu_m | Unemployment rate | m | 1995.m1-2012.m6 | 5 |
| sts_trtu_defltv | Index of deflated turnover | m | 1995.m1-2012.m6 | 5 |
| | Italy | m | 1999.m1-2011.m8 | |

**Sector TIS: Taxes less subsidies on products**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_gdp_k | Taxes less Subsidies | q | 1995.q1-2012.q2 | 5 |
| sts_inpr_msa | Index of industrial production | m | 1995.m1-2012.m6 | 5 |
| sts_trtu_defltv | Index of deflated turnover | m | 1995.m1-2011.m8 | 5 |
| | Italy | m | 1999.m1-2011.m8 | |

***Table 5:*** *National accounts series and monthly series by sector of activity and GDP component, expenditure side. For each series are reported: the mnemonic name for the Eurostat database (`http://epp.eurostat.ec.europa.eu/`), the frequency, sample period and number of countries available. The available countries are: Euro area (EA), Germany (G), Spain (S), France (F) and Italy (I). In case of different sample period the name of the series and the sample period are also reported. If a dataset different from Eurostat is used it is reported between brackets.*

**Sector G-H-I: Trade, transport and communication services**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_nace06_k | Value added G-H-I | q | 1995.q1-2012.q2 | 5 |
| sts_trtu_tv | Index of turnover | m | 1995.m1-2012.m6 | 5 |
| | Italy | m | 2000.m1-2012.m6 | |
| sts_trtu_defltv | Index of deflated turnover | m | 1995.m1-2012.m6 | 5 |
| | Italy | m | 1999.m1-2011.m8 | |
| sts_trlb_m | Employment | m | 1996.m1-2012.m6 | 1: G |
| sts_rt_careg | Car registrations | m | 1996.m1-2011.m9 | 5 (OECD) |
| bs_ciGHI | Confidence Indicator, Retail | m | 1996.m1-2012.m6 | 5 |

**Sector GCF: Gross Capital Formation**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_gdp_k | Gross Capital Formation | q | 1995.q1-2012.q2 | 5 |
| sts_inpr_msa | Index of industrial production | m | 1995.m1-2012.m6 | 5 |
| sts_intvd_m | Index of industrial turnover | m | 1995.m1-2012.m6 | 5 |
| | Spain | m | 2002.m1-2012.m6 | |
| | Italy | m | 2000.m1-2012.m6 | |
| bssi_GCF | Economic sentiment indicator | m | 1995.m1-2012.m6 | 5 |

**Sector IMP: Import of goods and services**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_gdp_k | Imp. of goods and services | q | 1995.q1-2012.q2 | 5 |
| sts_inpr_msa | Index of indus. prod. | m | 1995.m1-2012.m6 | 5 |
| imt_st_27msbec | Volume Index of Imports | m | 2000.m1-2012.m6 | 4:G,S,F,I |

**Sector FCE: Final Consumption Expenditures**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_gdp_k | Final Consumption Expenditures | q | 1995.q1-2012.q2 | 5 |
| sts_trtu_defltv | Index of deflated turnover | m | 1995.m1-2012.m6 | 5 |
| | Italy | m | 1999.m1-2011.m8 | |
| sts_ts_careg | Car registrations | m | 1995.m1-2011.m9 | 5 (OECD) |
| bssi_FCE | Consumer confidence Indicator | m | 1995.m1-2012.m6 | 5 |
| bsco_FCE | Major purchases at present | m | 1995.m1-2012.m6 | 5 |

**Sector EXP: Exports of goods and services**

| Name | Series | Frequency | Sample period | Number of series |
|---|---|---|---|---|
| namq_gdp_k | Exports of goods and services | q | 1995.q1-2012.q2 | 5 |
| sts_inpr_msa | Index of industrial production | m | 1995.m1-2012.m6 | 5 |
| ext_st_27msbec | Volume Index of Exports | m | 2000.m1-2012.m6 | 4:G,S,F,I |

## Appendix B: Initial conditions

As stated above, the state space formulation is modified when missing values are present and needs to be adjusted at the beginning of the sample. The initial two observations, $t = 0, 1$, have a special representation. At time $t = 0$ we have the following measurement equation:

$$\mathbf{Y}_{10}^{c} = \begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{\alpha}_0 \tag{20}$$

where the initial state vector is defined as

$$\boldsymbol{\alpha}_0 = \begin{bmatrix} \mathbf{Y}_{10}^{c\prime}, & \mathbf{y}_{10-1}^{\prime}, & \Delta\mathbf{y}_{10}^{\prime}, & \Delta\mathbf{y}_{1,-1}^{\prime}, & \mathbf{f}_t^{\prime}, & \mathbf{f}_{-1}^{\prime}, & \mathbf{u}_{10}^{\prime}, & \mathbf{u}_{20}^{\prime} \end{bmatrix}^{\prime}.$$

The specification of the model requires the specification of the distribution of the initial state vector $\boldsymbol{\alpha}_0$. For this purpose, the first block is rewritten $\mathbf{Y}_{1,0}^{c} = \exp(\mathbf{y}_{1,0})$, as $\rho_0 = 0$; its first order Taylor approximation around the trial value $\mathbf{y}_{1,0}^{*}$ is

$$\begin{aligned}\mathbf{Y}_{1,0}^{c} &= \exp(\mathbf{y}_{1,0}^{*}) + \mathbf{U}_{1,0}^{*}\mathbf{y}_{1,0} - \mathbf{U}_{1,0}^{*}\mathbf{y}_{1,0}^{*} \\ &= \left(\exp(\mathbf{y}_{1,0}^{*}) + \mathbf{U}_{1,0}^{*}\boldsymbol{\mu}_1 - \mathbf{U}_{1,0}^{*}\mathbf{y}_{1,0}^{*}\right) + \mathbf{U}_{1,0}^{*}\mathbf{y}_{1,-1} + \mathbf{U}_{1,0}^{*}\boldsymbol{\Theta}_1\mathbf{f}_0 + \mathbf{U}_{1,0}^{*}\mathbf{u}_{1,0}.\end{aligned}$$

The first two blocks of the state vector are nonstationary and are initialised by the vector $\boldsymbol{\beta} = \mathbf{y}_{1,-1}$, whereas the last four blocks have a stationary distribution, which depends on $\mathbf{f}_0 \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_f)$, where $\boldsymbol{\Sigma}_f$ solves the matrix equation $\boldsymbol{\Sigma}_f = \boldsymbol{\Phi}\boldsymbol{\Sigma}_f\boldsymbol{\Phi}^{\prime} + \boldsymbol{\Sigma}_\eta$, and $\mathbf{u}_0 = [\mathbf{u}_{1,0}^{\prime}, \mathbf{u}_{2,0}^{\prime}]^{\prime} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_u)$, where $\boldsymbol{\Sigma}_u$ solves the matrix equation $\boldsymbol{\Sigma}_u = \boldsymbol{\Psi}\boldsymbol{\Sigma}_\epsilon\boldsymbol{\Psi}^{\prime} + \boldsymbol{\Sigma}_\epsilon$.

Since $\mathbf{y}_{1,0} = \mathbf{y}_{1,-1} + \Delta\mathbf{y}_{1,0}$, the initial state vector is thus written as:

$$\boldsymbol{\alpha}_0 = \mathbf{A}_{0,0}\boldsymbol{\beta} + \mathbf{a}_{0,0} + \mathbf{H}_0\boldsymbol{\omega}_0, \quad \boldsymbol{\omega}_0 = \begin{bmatrix} \mathbf{f}_0 \\ \mathbf{f}_{-1} \\ \mathbf{u}_{10} \\ \mathbf{u}_{20} \\ \mathbf{u}_{1,-1} \end{bmatrix},$$

where

$$\mathbf{a}_{0,0} = \begin{bmatrix} \mathbf{Y}_{1,0}^{c*} - \mathbf{U}_{1,0}^{*}\mathbf{y}_{1,0}^{*} + \mathbf{U}_{1,0}^{*}\boldsymbol{\mu}_1 \\ \mathbf{0} \\ \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_1 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \mathbf{A}_{0,0} = \begin{bmatrix} \mathbf{U}_{1,0}^{*} \\ \mathbf{I}_{N1} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \mathbf{P}_{0,0} = \mathbf{H}_0\mathrm{Cov}(\boldsymbol{\omega}_0)\mathbf{H}_0^{\prime}.$$

$$\mathbf{H}_0 = \begin{bmatrix} \mathbf{U}_{1,0}^{*}\boldsymbol{\Theta}_1 & \mathbf{0} & \mathbf{U}_{1,0}^{*} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_1 & \mathbf{0} & \mathbf{I}_{N1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta}_1 & \mathbf{0} & \mathbf{0} & \mathbf{I}_{N1} \\ \mathbf{I}_K & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{N2} & \mathbf{0} \end{bmatrix}.$$

We assume that $\boldsymbol{\beta}$ is a diffuse random vector, i.e. it has an improper distribution with zero mean and an arbitrarily large variance matrix. The diffuse case captures the nonstationarity of a particular unobserved component and entails marginalising the inferences with respect to the parameter vector $\boldsymbol{\beta}$. As de Jong (1990) has shown, the posterior mean of $\boldsymbol{\beta}$ under the diffuse prior is coincident with the generalised least squares estimate of the parameter $\boldsymbol{\beta}$ considered as a fixed parameter vector in the classical sense.

At time $t = 1$ the measurement equation becomes:

$$\begin{bmatrix} \mathbf{Y}^c_{11} \\ \Delta \mathbf{y}_{21} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_2} \end{bmatrix} \boldsymbol{\alpha}_1 + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix} \tag{21}$$

where $\boldsymbol{\alpha}_1 = \begin{bmatrix} \mathbf{Y}^{c'}_{11}, & \mathbf{y}'_{10} & \Delta \mathbf{y}'_{11}, & \Delta \mathbf{y}'_{10}, & \mathbf{f}'_1, & \mathbf{f}'_0, & \mathbf{u}'_{11}, & \mathbf{u}'_{21} \end{bmatrix}'$.

The transition equation at time $t = 1$ is

$$\begin{bmatrix} \mathbf{Y}^c_{1t} \\ \mathbf{y}_{1t-1} \\ \Delta \mathbf{y}_{1t} \\ \Delta \mathbf{y}_{1t-1} \\ \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \mathbf{u}_{1t} \\ \mathbf{u}_{2t} \end{bmatrix} = \begin{bmatrix} \rho_1 \mathbf{I}_{N1} & \mathbf{U}^*_{11} & \mathbf{U}^*_{11}(\mathbf{I} + \boldsymbol{\Psi}_1) & \mathbf{0} & \mathbf{U}^*_{11}[\boldsymbol{\Theta}_1 \boldsymbol{\Phi} - \boldsymbol{\Psi}_1 \boldsymbol{\Theta}_1] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N1} & \mathbf{I}_{N1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{N_1} & \mathbf{0} & [\boldsymbol{\Theta}_1 \boldsymbol{\Phi} - \boldsymbol{\Psi}_1 \boldsymbol{\Theta}_1] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{N_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{N_2} \end{bmatrix} \begin{bmatrix} \mathbf{Y}^c_{10} \\ \mathbf{y}_{1-1} \\ \Delta \mathbf{y}_{10} \\ \Delta \mathbf{y}_{1-1} \\ \mathbf{f}_0 \\ \mathbf{f}_{-1} \\ \mathbf{u}_{10} \\ \mathbf{u}_{20} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{Y}^*_{1t} - \mathbf{U}^*_{1t} \mathbf{y}^*_{1t} + \mathbf{U}^*_{1t}(\mathbf{I} - \boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1 \\ \mathbf{0} \\ (\mathbf{I} - \boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{U}_{1t}\boldsymbol{\Theta}_1 & \mathbf{U}^*_{1t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_1 & \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}. \tag{22}$$

# Appedix C: The Treatment of Missing Values

## Case 1: Missing at time t but observed at time t-1

The state vector needs to be augmented by 2 components for each missing value affecting $\mathbf{y}_t$ at time $t$ becoming $\boldsymbol{\alpha}_t = [\mathbf{Y}^{c'}_{1t}, \mathbf{y}'_{1t-1}, \Delta \mathbf{y}'_{1t}, \Delta \mathbf{y}'_{1t-1}, \mathbf{f}'_t, \mathbf{f}'_{t-1}, \boldsymbol{\alpha}^{+'}_{1t}, \boldsymbol{\alpha}^{+'}_{2t}]'$, where the $\boldsymbol{\alpha}^{+'}_{1t}, \boldsymbol{\alpha}^{+'}_{2t}$ are latent states that are used to take into account the missing values. The measurement equation is modified as follows:

$$\begin{bmatrix} \mathbf{Y}^c_{1t} \\ \Delta \mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_2 & -\boldsymbol{\Psi}_2 \boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{\alpha}_t + \begin{bmatrix} \mathbf{0} \\ (\mathbf{I} - \boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}, \tag{23}$$

and the state equation becomes:

$$\begin{bmatrix} \mathbf{Y}^c_{1t} \\ \mathbf{y}_{1t-1} \\ \Delta\mathbf{y}_{1t} \\ \Delta\mathbf{y}_{1t-1} \\ \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \boldsymbol{\alpha}^+_{1t} \\ \boldsymbol{\alpha}^+_{2t} \end{bmatrix} = \begin{bmatrix} \rho_t\mathbf{I}_{N1} & \mathbf{U}^*_{1t} & \mathbf{U}^*_{1t}(\mathbf{I}+\boldsymbol{\Psi}_1) & \mathbf{0} & \mathbf{U}^*_{1t}[\boldsymbol{\Theta}_1\boldsymbol{\Phi}-\boldsymbol{\Psi}_1\boldsymbol{\Theta}_1] & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N1} & \mathbf{I}_{N1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{N_1} & \mathbf{0} & [\boldsymbol{\Theta}_1\boldsymbol{\Phi}-\boldsymbol{\Psi}_1\boldsymbol{\Theta}_1] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & [\boldsymbol{\Theta}_2\boldsymbol{\Phi}-\boldsymbol{\Psi}_2\boldsymbol{\Theta}_2] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & [\boldsymbol{\Theta}_2\boldsymbol{\Phi}-\boldsymbol{\Psi}_2\boldsymbol{\Theta}_2] & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Y}^c_{1t-1} \\ \mathbf{y}_{1t-2} \\ \Delta\mathbf{y}_{1t-1} \\ \Delta\mathbf{y}_{1t-2} \\ \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \end{bmatrix}$$
$$+ \begin{bmatrix} \mathbf{Y}^*_{1t} - \mathbf{U}^*_{1t}\mathbf{y}^*_{1t} + \mathbf{U}^*_{1t}(\mathbf{I}-\boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1 \\ \mathbf{0} \\ (\mathbf{I}-\boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{y}_{2t-1} + \boldsymbol{\Psi}_2\Delta\mathbf{y}_{2t-1} + (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \\ \boldsymbol{\Psi}_2\Delta\mathbf{y}_{2t-1} + (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{U}_{1t}\boldsymbol{\Theta}_1 & \mathbf{U}^*_{1t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_1 & \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{I}_{N_2} \\ \boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{I}_{N_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}. \tag{24}$$

The $\Delta\mathbf{y}_{2t}$ corresponds to the series that are available at time $t$ and $t-1$. The missing series simply do not enter in the calculation of the likelihood, but they are propagated in the latent states using $\boldsymbol{\alpha}^{+'}_{1t}, \boldsymbol{\alpha}^{+'}_{2t}$. The available observations for the missing series at time $t$ but observed at time $t-1$ are inserted in the latent states using the relation $\mathbf{y}_{2t-1} + \boldsymbol{\Psi}_2\Delta\mathbf{y}_{2t-1} + (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2$ where $\mathbf{y}_{2t-1}$ represent series in levels observed at time $t-1$ and $\Delta\mathbf{y}_{2t-1}$ represent series in first differences at time $t-1$.

### Case 2: Missing at time t and time t-1

As before $\Delta\mathbf{y}_{2t}$ corresponds to the available observations at time $t$ and $t-1$. The observations that are still missing are propagated using the $\boldsymbol{\alpha}^{+'}_{1t}, \boldsymbol{\alpha}^{+'}_{2t}$ created in Case 1. The measurement equation is modified as follows:

$$\begin{bmatrix} \mathbf{Y}^c_{1t} \\ \Delta\mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_2 & -\boldsymbol{\Psi}_2\boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{\alpha}_t + \begin{bmatrix} \mathbf{0} \\ (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}, \tag{25}$$

and the state equation becomes:

$$\begin{bmatrix} \mathbf{Y}_{1t}^c \\ \mathbf{y}_{1t-1} \\ \Delta\mathbf{y}_{1t} \\ \Delta\mathbf{y}_{1t-1} \\ \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \boldsymbol{\alpha}_{1t}^+ \\ \boldsymbol{\alpha}_{2t}^+ \end{bmatrix} = \begin{bmatrix} \rho_t\mathbf{I}_{N1} & \mathbf{U}_{1t}^* & \mathbf{U}_{1t}^*(\mathbf{I}+\boldsymbol{\Psi}_1) & \mathbf{0} & \mathbf{U}_{1t}^*[\boldsymbol{\Theta}_1\boldsymbol{\Phi}-\boldsymbol{\Psi}_1\boldsymbol{\Theta}_1] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N1} & \mathbf{I}_{N1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{N_1} & \mathbf{0} & [\boldsymbol{\Theta}_1\boldsymbol{\Phi}-\boldsymbol{\Psi}_1\boldsymbol{\Theta}_1] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & [\boldsymbol{\Theta}_2\boldsymbol{\Phi}-\boldsymbol{\Psi}_2\boldsymbol{\Theta}_2] & \mathbf{0} & \mathbf{I} & \boldsymbol{\Psi}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & [\boldsymbol{\Theta}_2\boldsymbol{\Phi}-\boldsymbol{\Psi}_2\boldsymbol{\Theta}_2] & \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{1t-1}^c \\ \mathbf{y}_{1t-2} \\ \Delta\mathbf{y}_{1t-1} \\ \Delta\mathbf{y}_{1t-2} \\ \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \\ \boldsymbol{\alpha}_{1t-1}^+ \\ \boldsymbol{\alpha}_{2t-1}^+ \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{Y}_{1t}^* - \mathbf{U}_{1t}^*\mathbf{y}_{1t}^* + \mathbf{U}_{1t}^*(\mathbf{I}-\boldsymbol{\Psi}_{N_1})\boldsymbol{\mu}_1 \\ \mathbf{0} \\ (\mathbf{I}-\boldsymbol{\Psi}_{N_1})\boldsymbol{\mu}_1 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \\ (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{U}_{1t}\boldsymbol{\Theta}_1 & \mathbf{U}_{1t}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_1 & \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{I}_{N_2} \\ \boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{I}_{N_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}. \tag{26}$$

## Case 3. Observed at time t but missing at time t-1

When the observations on $\mathbf{y}_{2t}$ are available, but the first differences are not, due to a missing value at time $t-1$, two additional latent states are introduced, $\boldsymbol{\alpha}_{1t}^{+'}, \boldsymbol{\alpha}_{2t}^{+'}$. The measurement equation is modified as follows:

$$\begin{bmatrix} \mathbf{Y}_{1t}^c \\ \Delta\mathbf{y}_{2t} \\ \mathbf{y}_{2t} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{N_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_2 & -\boldsymbol{\Psi}_2\boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \boldsymbol{\Psi}_2 \end{bmatrix} \boldsymbol{\alpha}_t + \begin{bmatrix} \mathbf{0} \\ (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}, \tag{27}$$

whereas the state equation becomes:

$$\begin{bmatrix} \mathbf{Y}_{1t}^c \\ \mathbf{y}_{1t-1} \\ \Delta\mathbf{y}_{1t} \\ \Delta\mathbf{y}_{1t-1} \\ \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \boldsymbol{\alpha}_{1t}^+ \\ \boldsymbol{\alpha}_{2t}^+ \end{bmatrix} = \begin{bmatrix} \rho_t \mathbf{I}_{N1} & \mathbf{U}_{1t}^* & \mathbf{U}_{1t}^*(\mathbf{I}+\boldsymbol{\Psi}_1) & \mathbf{0} & \mathbf{U}_{1t}^*[\boldsymbol{\Theta}_1\boldsymbol{\Phi} - \boldsymbol{\Psi}_1\boldsymbol{\Theta}_1] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N1} & \mathbf{I}_{N1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{N_1} & \mathbf{0} & [\boldsymbol{\Theta}_1\boldsymbol{\Phi} - \boldsymbol{\Psi}_1\boldsymbol{\Theta}_1] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_2 & -\boldsymbol{\Psi}_2\boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \boldsymbol{\Psi}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{1t-1}^c \\ \mathbf{y}_{1t-2} \\ \Delta\mathbf{y}_{1t-1} \\ \Delta\mathbf{y}_{1t-2} \\ \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \\ \boldsymbol{\alpha}_{1t-1}^+ \\ \boldsymbol{\alpha}_{2t-1}^+ \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{Y}_{1t}^* - \mathbf{U}_{1t}^*\mathbf{y}_{1t}^* + \mathbf{U}_{1t}^*(\mathbf{I}-\boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1 \\ \mathbf{0} \\ (\mathbf{I}-\boldsymbol{\Psi}_{N_1})\boldsymbol{\delta}_1 \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \\ (\mathbf{I}-\boldsymbol{\Psi}_2)\boldsymbol{\delta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{U}_{1t}\boldsymbol{\Theta}_1 & \mathbf{U}_{1t}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_1 & \mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{I}_{N_2} \\ \boldsymbol{\Theta}_2 & \mathbf{0} & \mathbf{I}_{N_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\epsilon}_{1t} \\ \boldsymbol{\epsilon}_{2t} \end{bmatrix}. \quad (28)$$

When there are no further missing values the components $\boldsymbol{\alpha}_{1t}^+$ and $\boldsymbol{\alpha}_{2t}^+$ are removed and the state space formulation collapses to (11)–(12).

# Appendix D: Chained GDP estimates

We start by indexing the month of the year by $j, j = 0, \dots, 11$ and the year by $m$, $m = 1, \dots, M = [(n+1)/12]$, so that the time index is written $t = j + 12m, t = 0, \dots, n$.

For a particular estimated monthly time series let us denote by $Y_{jm}$ the value at current prices of month $j$ in year $m$, $Y_{.m} = \sum_j Y_{jm}$ the annual total, $\bar{Y}_m = Y_{.m}/12$ the annual average (the annual and quarterly figures are available from the national accounts, compiled by Eurostat). The chain-linked volume estimate with reference year $b$ (the year 2000 in our case) will be denoted $\hat{Y}_{jm}^{(b)}$. The temporal disaggregation methods described in the paper are applied to the quarterly chained-linked volume series with reference year $b$ and yield estimates that add up to the quarterly and annual totals (temporal consistency), but are not additive in a horizontal (that is cross-sectional) sense.

The following multistep procedure enables the computation of volume measures expressed at the prices of the previous year that are additive, also horizontally.

1. *Dechaining*:

   (a) Transform the monthly estimates into Laspeyres-type quantity indices with reference year $b$ (volumes are evaluated at year $b$ average prices), by computing

   $$I_{jm}^{(b)} = \frac{\hat{Y}_{jm}^{(b)}}{\bar{Y}_{.b}}, j = 0, \dots, 11, m = 0, \dots, M,$$

   where the denominator is the annual average of year $b$ at current prices. In our case $b = 5$ (year 5 is the calendar year 2000).

(b) Change the reference year to $m = 1$, the second year of the series (1996 in our case), by computing:

$$I_{jm}^{(1)} = \frac{I_{jm}^{(b)}}{\bar{I}_1^{(b)}}, j = 1, \ldots, 11, m = 0, \ldots, M,$$

where $\bar{I}_1^{(b)} = \sum_j I_{j1}^{(b)}/12$ is the average quantity index for the second year of the the sample.

(c) Transform the quantity indices for year $m = 1, 2, \ldots, M$ into indices with reference year $m - 1$ (the previous year), by rescaling $I_{jm}^{(1)}$ as follows:

$$I_{jm}^{(m-1)} = \frac{I_{jm}^{(1)}}{\bar{I}_{m-1}^{(1)}}, j = 0, \ldots, 11, m = 1, \ldots, M,$$

where

$$\bar{I}_{m-1}^{(1)} = \frac{1}{12} \sum_j I_{j,m-1}^{(1)}, m = 1, \ldots, M$$

(d) Compute the series at the average prices of the previous year as:

$$\hat{Y}_{jm}^{(m-1)} = I_{jm}^{(m-1)} \bar{Y}_{m-1}, \quad j = 0, \ldots, 11, m = 1, \ldots, M,$$

2. *Aggregation step*: Let $\mathbf{Y}_t^{(m-1)}$ denote the disaggregate time series expressed at the average prices of the previous year. Using the original estimates and the dechaining procedure we can assume that, at least approximately,

$$\mathbf{Y}_t^{(m-1)} \sim \mathrm{N}\left(\hat{\mathbf{Y}}_t^{(m-1)}, \hat{\mathbf{V}}_t^{(m-1)}\right), t = 0, 1, \ldots, n,$$

where the first and second moments are given by the sequential constrained estimates produced by the Kalman filter and smoother outlined in the previous section, modified to take into account the dechaining procedure. If the $r$ cross-sectional constraints are expressed as

$$\mathbf{Q}\mathbf{Y}_t = \mathbf{q}$$

where $\mathbf{Q}$ is an $r \times N_1$ matrix, and $\mathbf{q}$ is $r \times 1$, the modified estimates that comply with those constraints and their MSE matrix are given respectively by

$$\begin{array}{rcl} \tilde{\mathbf{Y}}_t^{(m-1)} & = & \hat{\mathbf{Y}}_t^{(m-1)} + \hat{\mathbf{V}}_t^{(m-1)} \mathbf{Q}'(\mathbf{Q}\hat{\mathbf{V}}_t^{(m-1)}\mathbf{Q}'^{-1}(\mathbf{q} - \mathbf{Q}\hat{\mathbf{Y}}_t^{(m-1)}) \\ \tilde{\mathbf{V}}_t^{(m-1)} & = & \hat{\mathbf{V}}_t^{(m-1)} - \hat{\mathbf{V}}_t^{(m-1)} \mathbf{Q}'(\mathbf{Q}\hat{\mathbf{V}}_t^{(m-1)}\mathbf{Q}'^{-1}\mathbf{Q}\hat{\mathbf{V}}_t^{(m-1)} \end{array}$$

see, e.g. **?**. In our case, for each country and the Euro area, $r = 2$, $\mathbf{q} = \mathbf{0}$, and

$$\mathbf{Q} = \left[ \begin{array}{cccccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 \end{array} \right],$$

if the GDP components are arranged as in section 4 and the last element represents GDP at market prices. The new balanced estimates are now ready to be expressed at the average prices of reference year $b$.

3. *Chain-linking* (annual overlap):

   (a) Convert the aggregated volume measures into Laspeyres-type quantity indices with respect to the previous year:

$$\mathcal{I}_{jm}^{(m-1)} = \frac{\tilde{Y}_{jm}^{(m-1)}}{\bar{Y}_{m-1}}, \quad j = 0, \ldots, 11, m = 1, \ldots, M,$$

   where $\bar{Y}_{m-1} = \sum_j Y_{j,m-1}/12$ is the average of the previous year at current prices. The annual and quarterly totals is available from the national accounts compiled by Eurostat.

   (b) Chain-link the indices using the recursive formula (the first year is the reference year):

$$\mathcal{I}_{jm}^{(0)} = \mathcal{I}_{jm}^{(m-1)} \bar{\mathcal{I}}_{m-1}^{(0)}, \qquad j = 0, \ldots, 11, m = 1, \ldots, M,$$

   where $\bar{\mathcal{I}}_0^{(0)} = 1$ and

$$\bar{\mathcal{I}}_{m-1}^{(0)} = \frac{1}{12} \sum_j \mathcal{I}_{j,m-1}^{(0)}.$$

   (c) If $b > 0$ then change the reference year to year $b$:

$$\mathcal{I}_{jm}^{(b)} = \frac{\mathcal{I}_{jm}^{(0)}}{\bar{\mathcal{I}}_b^{(0)}} \quad j = 0, \ldots, 11, m = 1, \ldots, M.$$

   (d) Compute the chain-linked volume series with reference year $b$:

$$\tilde{Y}_{jm}^{(b)} = \mathcal{I}_{jm}^{(b)} \bar{Y}_b \quad j = 1, \ldots, 12, m = 2, \ldots, M,$$

   where $\bar{Y}_b = \frac{1}{12} \sum_j Y_{jb}$ is the value of GDP (at basic or market prices) at current prices of the reference year.

The multistep procedure just described enables to obtain monthly estimates in volume such that the values $\tilde{Y}_{jm}^{(m-1)}$ expressed at the average prices of the previous year add up to their quarterly and annual totals published by Eurostat and are consistent with the contemporaneous aggregation constraints. On the contrary, as a result of the chaining procedure, the chain-linked volumes $\tilde{Y}_{jm}^{(b)}$ expressed at the prices of the common reference year $b$ (2000) are consistent only with the temporal aggregation constraints; however, their estimates are more reliable since they have been combined with the estimates of other related variables.