Nowcasting

Marta Bańbura*, European Central Bank
Domenico Giannone, Université libre de Bruxelles, ECARES and CEPR
Lucrezia Reichlin, London Business School and CEPR

February 24, 2010

Very preliminary

* The opinions in this paper are those of the author and do not necessarily reflect the views of the European Central Bank.

1 Introduction

Economists have imperfect knowledge of the present state of the economy and even of the recent past. Many key statistics are released with a long delay and they are subsequently revised. As a consequence, unlike weather forecasters, who know what is the weather today and only have to predict the weather tomorrow, economists have to forecast the present and even the recent past. The problem of predicting the present, the very near future and the very recent past is labeled as *nowcasting* and this is what this paper is about.

Nowcasting is particularly relevant for key macro economic variables which are collected at low frequency, typically on a quarterly basis, and released with a substantial lag. To obtain "early estimates" of such key economic indicators, nowcasters use the information from data that are related to the target variable but are collected at higher frequency, typically monthly, and released in a more timely manner. One of the key features of a successful nowcasting tool is being able to incorporate the most up-to-date information in an environment in which data are released in a non-synchronous manner and with varying publication lags.

For example, euro area GDP is only available at quarterly frequency and is released six weeks after the close of the quarter. In March 2010 we only have information up to the last quarter of 2009 and we need to wait until mid-May to obtain a first estimate of the first quarter of 2010. However, there are several variables, available at monthly frequency and published with short delay, which can be used to construct early estimates of GDP. For example, in mid March comes a release of euro area industrial production for January. These series measure directly certain components of GDP and are considered to contain a strong signal on its short-term developments. Much more timely information, albeit potentially less precise, is provided by various surveys. They measure expectations of economic activity and are typically available shortly before the end of the month to which they refer. Beyond industrial production and surveys, many other data releases are likely to be informative on the state of the economy as it is revealed by the fact that many are closely watched by financial markets which react whenever there are surprises about the value of the new data (for evidence on this point, see Cutler, Poterba, and Summers, 1989).

While in this paper we concentrate the discussion around GDP, the ideas developed here could be applied to nowcasting any low frequency variable released with a substantial delay, for which we can exploit more timely, higher frequency information. The emphasis on GDP here is justified by the fact that this is a key statistics describing the state of the economy. In policy institutions, and in particular central banks, its nowcast is closely monitored and frequently updated to incorporate the information from latest data releases. Further, the nowcast is used as an input for the more general forecasting process which is concerned with longer horizon and is often conducted on the basis of large structural models.

Until recently, nowcasting had received very little attention by the academic literature, although it was routinely conducted in policy institutions either through a judgemental process or on the basis of simple models. Among these simple models are the so called bridge equations, which relate GDP to quarterly aggregates of one or few monthly series.

Although the bridge between monthly and quarterly variables is an essential component of nowcasting, as monthly data are more timely than quarterly and they are released more often, nowcasting ideally requires more complex modelling than what is offered by bridge equations. This is because it not only consists in updating the estimates of the target quarterly variable as new data become available throughout the quarter, but also in commenting and interpreting the sequence of revisions of those estimates. Not only do we want to know by how much GDP nowcast has been revised, but also what explains the revision. Typical questions for the briefer are: is an upward revision explained by higher than expected readings of industrial production or surveys or by both and what weighs the most? In other words, we are interested in relating the part of the monthly release that was previously unexpected (the *news*) to the revisions of GDP estimates. For this kind of analysis we need to model the joint dynamics of the monthly input data and the quarterly target variable in a unified framework.

Two seminal papers (Evans, 2005; Giannone, Reichlin, and Small, 2008) have formalized this process in statistical models. Both approaches allow to model, within the same statistical framework, the joint dynamics of GDP and the monthly data releases as well as handling data which have missing observations at the end of the sample, due to non synchronized publication lags (the so called jagged/ragged edge problem).

This paper is based on the factor model by Giannone, Reichlin, and Small (2008) but also discusses several improvements, due to Bańbura and Modugno (2009). The Giannone, Reichlin, and Small (2008)'s methodology has a number of desirable features and, in particular, it offers a parsimonious solution for the inclusion of a rich information set. Data which are typically watched and commented on throughout a quarter are at least a dozen, but this number can be higher. The model was first implemented at the Board of Governors of the Federal Reserve in a project which started in 2003 and then at the European Central Bank (Angelini, Camba-Méndez, Giannone, Rünstler, and Reichlin, 2008; Bańbura and Rünstler, 2010; Rünstler, Barhoumi, Cristadoro, Reijer, Jakaitiene, Jelonek, Rua, Ruth, Benk, and Nieuwenhuyze, 2008). The method has also been implemented in other central banks for other economies, including Ireland (D'Agostino, McQuinn, and O'Brien, 2008), New Zealand (Matheson, 2010) and Norway (Aastveit and Trovik, 2008).

Two results that have emerged from the empirical literature suggest that nowcasting has an important place in the broader forecasting literature. First, Giannone, Reichlin, and Small (2008) show that institutional and statistical forecasts of GDP are outperformed by the naïve constant growth model at horizons longer than the current quarter. This implies that our ability to forecast GDP growth is mostly concentrated in the current (and previous) quarter. Second, Giannone, Reichlin, and Sala (2004) show that the automatic statistical procedure in Giannone, Reichlin, and Small (2008) performs as well as the nowcast published in the Greenbooks, which is the result of a complex process involving models and judgement. For the euro area, similar results are obtained in Angelini, Camba-Méndez, Giannone, Rünstler, and Reichlin (2008).

Another robust empirical result coming from this work is that timeliness matters, that is the exploitation of early releases leads to improvement in the nowcast accuracy. In particular, the literature shows that surveys, which are the most timely information, contribute to an improvement of the estimate early in the quarter but by the time hard information, such as industrial production, becomes available later in the quarter their contribution vanishes (Angelini, Camba-Méndez, Giannone, Rünstler, and Reichlin, 2008; Bańbura and Rünstler, 2010; Giannone, Reichlin, and Small, 2008; Matheson, 2010).

The chapter is organized as follows. The second section defines the problem of nowcasting in general and relates it to the concept of *news* in macroeconomic data releases briefly described above. In the third section, we explain the details of our approach. In section four we discuss related literature while, in section five, we illustrate the characteristics of the model via an application to the nowcast of GDP and inflation in the euro area. The last section concludes.

2 The problem

Before referring to a particular model, let us define formally the general problem of nowcast and its updates, which arise as a result of an inflow of new information.

To fix ideas we will illustrate the problem on an example of the GDP nowcast. As mentioned in the introduction, GDP figure becomes available only 6 weeks after the close of the reference quarter. In the mean-time it can be estimated using higher-frequency, namely monthly, variables that are published in a more timely manner.

To describe the problem more formally, let us denote by Ω_v a vintage of data available at time v, where v refers to the date of a particular data release. Further let us denote GDP at time t as y_t^Q. We define the problem of nowcasting of y_t^Q as the orthogonal projection of y_t^Q on the available information set Ω_v:

$$\mathbb{P}\left[y_t^Q|\Omega_v\right] = \mathbb{E}\left[y_t^Q|\Omega_v\right], \tag{1}$$

where $\mathbb{E}\left[\,\cdot\,|\Omega_v\right]$ refers to a conditional expectation. One of the elements that distinguish nowcasting from other forecast applications is the structure of the information set Ω_v. One particular feature is typically referred to as its "ragged" or "jagged edge". It means that, since data are released in a non-synchronous manner and with different degrees of delay, the time of the last available observation differs from series to series. Another feature is that it contains mixed frequency series, in our case monthly and quarterly. Hence we will have $\Omega_v = \{x_{i,t_i}\,, t_i = 1, 2, ..., T_{i,v}\,, i = 1, ..., n\ ; y_{3k}^Q\,, 3k = 3, 6, ..., T_{Q,v}\}$ where $T_{i,v}$ corresponds to the last period for which in vintage v the series j has been observed.[1] Because of the non-synchronicity of data releases, $T_{i,v}$ is not the same across variables and therefore the data set exhibits the above mentioned "jagged edge".

Hence the problem of nowcasting consist in designing a framework which first, imposes a plausible probability structure on Ω_v and second, can efficiently exploit all the relevant information from such Ω_v, where the number of potential monthly predictors, $x_{i,t}$, could be large.

Important feature of nowcasting process is that one rarely performs a single projection for a quarter of interest but rather a sequence of nowcasts, which are updated as new data arrive. The first nowcasts are usually made with very little or no information on the reference quarter. With subsequent data releases they are revised, leading to more precise projections as the information on the period of interest accumulates. In other words we will, in general, perform a sequence of projections: $\mathbb{E}\left[y_t^Q|\Omega_v\right]$, $\mathbb{E}\left[y_t^Q|\Omega_{v+1}\right]$, ..., where v, $v+1$, ..., refer to dates of consecutive data releases. Typically the intervals between two consecutive data releases are short (possible couple of days or less) and change over time. Consequently, v has high frequency and it is irregularly spaced.

We now explain why and how the nowcast is updated and introduce the concept of *news* which is central to understanding the nowcast revisions.

[1] Given our definition of nowcast as prediction of the present, the very near future and the very recent past, the difference between $T_{Q,v}$ and $\max_i T_{i,v}$ is usually small and can be negative. Ω_v could possibly include more quarterly variables, we limit this set to GDP for the sake of simplicity.

Let us first analyse the difference between the two information sets Ω_v and Ω_{v+1}. At time $v+1$ we have a release of certain group of variables, $\{x_{j,T_{j,v+1}}, j \in J_{v+1}\}$ and consequently the information set expands.[2] The new information set differs from the preceding one for two reasons. First, it contains new, more recent figures. Second, old data might get revised. In what follows we will abstract from the problem of data revisions. Therefore, we have $\Omega_v \subseteq \Omega_{v+1}$ and $\Omega_{v+1} \setminus \Omega_v = \{x_{j,T_{j,v+1}}, j \in J_{v+1}\}$.

Given the "expanding" character of the information and the properties of orthogonal projections we can decompose the new forecast as:

$$\underbrace{\mathbb{E}\left[y_t^Q|\Omega_{v+1}\right]}_{\text{new forecast}} = \underbrace{\mathbb{E}\left[y_t^Q|\Omega_v\right]}_{\text{old forecast}} + \underbrace{\mathbb{E}\left[y_t^Q|I_{v+1}\right]}_{\text{revision}}, \tag{2}$$

where I_{v+1} is the subset of the information set Ω_{v+1} whose elements are orthogonal to all the elements of Ω_v. Given the difference between Ω_v and Ω_{v+1} specified above, we have that

$$I_{v+1,j} = x_{j,T_{j,v+1}} - \mathbb{E}\left[x_{j,T_{j,v+1}}|\Omega_v\right]$$

and $I_{v+1} = (I_{v+1,1} \dots I_{v+1,J})'$. Hence, the only element that leads to a change in the nowcast is the "unexpected" (with respect to the model) part of the data release, I_{v+1}, which we label as *news*. The *news* concept is useful because what matters in understanding the updating process of the nowcast is not the release itself but the difference between that release and what had been forecast before it. In particular, in an unlikely case that the released numbers are exactly as predicted by the model, the nowcast will not be revised. On the other hand, intuitively we would expect that a negative *news* in industrial production should revise the GDP forecasts downwards. Below we show how this can be quantified.

It is worth noting that the *news* is not a standard Wold forecast error. First of all, the pattern of data availability changes with time. Second, the *news* depend on the order in which new data are released.

From the properties of conditional expectation, we can further develop (2) as:

$$\mathbb{E}\left[y_t^Q|I_{v+1}\right] = \mathbb{E}\left[y_t^Q I_{v+1}'\right] \mathbb{E}\left[I_{v+1}I_{v+1}'\right]^{-1} I_{v+1}\,. \tag{3}$$

In order to expand (3) further and to extract a meaningful model-based *news* component, one needs to have a model which can reliably account for the joint dynamic relationships among the data. Given such model and assuming that the data are Gaussian, it turns out that we can find coefficients $b_{j,t,v+1}$ such that:

[2] Typically one "additional" observation is released and we have $T_{j,v+1} = T_{j,v} + 1$ for all $j \in J_{v+1}$. GDP could be also included in a release, we abstract from this case in order not to complicate the notation.

$$\underbrace{\mathbb{E}\left[y_t^Q|\Omega_{v+1}\right]}_{\text{new forecast}} = \underbrace{\mathbb{E}\left[y_t^Q|\Omega_v\right]}_{\text{old forecast}} + \sum_{j\in J_{v+1}} b_{j,t,v+1} \underbrace{\left(x_{j,T_{j,v+1}} - \mathbb{E}\left[x_{j,T_{j,v+1}}|\Omega_v\right]\right)}_{\text{news}} .$$

In other words we can express the forecast revision as a weighted sum of *news* from the released variables:

$$\underbrace{\mathbb{E}\left[y_t^Q|\Omega_{v+1}\right] - \mathbb{E}\left[y_t^Q|\Omega_v\right]}_{\text{forecast revision}} = \sum_{j\in J_{v+1}} b_{j,t,v+1} \underbrace{\left(x_{j,T_{j,v+1}} - \mathbb{E}\left[x_{j,T_{j,v+1}}|\Omega_v\right]\right)}_{\text{news}} . \tag{4}$$

Hence, consistently with the intuition, the magnitude of the forecast revision depends, on one hand, on the size of the *news* and, on the other hand, on its relevance for the target variable as quantified by the associated weight $b_{j,t,v+1}$.

Decomposition (4) enables us to trace the sources of forecast revisions back to individual predictors. More precisely, in the case of a simultaneous release of several (groups of) variables it is possible to decompose the resulting forecast revision into contributions from the *news* in individual (groups of) series. Note that if the release concerns only one group or one series, the contribution of its *news* is simply equal to the change in the forecast. In addition, we can produce statements like e.g. "after the release of industrial production, the forecast of GDP went down because the indicators turned out to be (on average) lower than expected".

3 The econometric framework

To compute nowcasts, *news* and their contributions to nowcast revisions all we need in practice is to perform linear projections which can be obtained efficiently by using Kalman filter techniques if the model can be cast in state space form. This is the basic idea of the approach proposed by Giannone, Reichlin, and Small (2008).

The approach allows the inclusion of many variables and this is a desirable characteristics since many releases are commented in the briefing process and monitored by the market. Including many variables in a statistical model easily leads to a curse of dimensionality problem with the consequence of imprecise and volatile estimates. To solve this problem, Giannone, Reichlin, and Small (2008) model the monthly data as a parametric dynamic factor model and handle the jagged edge data problem due to publication lags with the Kalman filter. The factors are estimated via a two step procedure where, in the first step, they are estimated by principal components and, in the second, by the Kalman filter. The asymptotic justification for this procedure is given in Doz, Giannone, and Reichlin (2006b).

Bańbura and Modugno (2009) have developed the model to introduce autoregressive dynamics and block specific factors which allows efficiency improvements. The model is estimated by Quasi Maximum Likelihood (QML) as proposed by Doz, Giannone, and Reichlin (2006a). In this approach the problem of jagged edged data is treated as a missing data problem whereby the lower frequency variable is considered to be a periodical missing aggregate of an unobserved higher frequency concept.

The next subsections describe the methodology in detail.

3.1 Monthly factor model

We start by specifying the dynamics for the monthly data. How to include quarterly variables within this framework is discussed in the next subsection.

Let $x_t = (x_{1,t}, x_{2,t}, \ldots, x_{n,t})'$ denote the monthly series, which have been transformed to satisfy the assumption of stationarity. More precisely, x_t are month-on-month growth rates (or differences) of the original variables, see the Appendix for details on the transformations applied. We assume that x_t obey the following factor model representation:

$$x_t = \mu + \Lambda f_t + \varepsilon_t \,, \tag{5}$$

where f_t is a $r \times 1$ vector of (unobserved) common factors and ε_t is a vector of idiosyncratic components. Λ denotes the factor loadings for the monthly variables. The common factors and the idiosyncratic components are assumed to have mean zero and hence the constants $\mu = (\mu_1, \mu_2, \ldots, \mu_n)'$ are the unconditional means. Further, the factors are modelled as a VAR process of order p:

$$f_t = A_1 f_{t-1} + \cdots + A_p f_{t-p} + u_t \,, \qquad u_t \sim i.i.d.\ N(0, Q) \,, \tag{6}$$

where $A_1, \ldots, A_p$ are $r \times r$ matrices of autoregressive coefficients.

Finally, we assume that the idiosyncratic component of the monthly variables follows an AR(1) process:

$$\varepsilon_{i,t} = \alpha_i \varepsilon_{i,t-1} + e_{i,t} \,, \qquad e_{i,t} \sim i.i.d.\ N(0, \sigma_i^2) \,, \tag{7}$$

with $\mathbb{E}\left[e_{i,t} e_{j,s}\right] = 0$ for $i \neq j$.

Taking explicitly into account the dynamics of the factors and of the idiosyncratic component is particularly important in nowcasting applications since it allows to fully exploit past information that is particularly useful at the beginning of the quarter when only few variables are available for the most recent period.

In contrast to models typically used in the context of nowcasting, we further restrict Λ, A_1, ..., A_p and Q. Specifically, we partition f_t into mutually independent global, real and nominal factors. We assume that global factor is loaded by all the variables while real and nominal factors are specific to real and nominal variables, respectively. Precisely, assuming (without loss of generality) that all the nominal variables are ordered before the real, we have:

$$\Lambda = \begin{pmatrix} \Lambda_{G,N} & \Lambda_{N,N} & 0 \\ \Lambda_{G,R} & 0 & \Lambda_{R,R} \end{pmatrix},$$

$$f_t = \begin{pmatrix} f_t^G \\ f_t^N \\ f_t^R \end{pmatrix}, \qquad A_i = \begin{pmatrix} A_{i,G} & 0 & 0 \\ 0 & A_{i,N} & 0 \\ 0 & 0 & A_{i,R} \end{pmatrix}, \qquad Q = \begin{pmatrix} Q_G & 0 & 0 \\ 0 & Q_N & 0 \\ 0 & 0 & Q_R \end{pmatrix}.$$

This framework is used to model local cross-sectional correlation within the real and nominal blocks, which is helpful for a more efficient extraction of the global factor. This type of restriction is easily accommodated within maximum likelihood approach to estimation as discussed below. Of course, this approach also allows implementation of other structures, e.g. more local factors for a finer grouping of the variables.

It has been shown, that for large cross-sections, the model given by (5) can be estimated by maximum likelihood under the assumption of lack of serial and cross-sectional correlation in the idiosyncratic component even if this condition is not satisfied by the data. However, this mis-specification can cause problems in small samples and consequently in nowcasting because of the incomplete cross-sections at the end of the sample. Explicit modelling of serial correlation of the idiosyncratic component and including local factors aims at mitigating this problem.[3]

3.2 Modelling quarterly variables

We follow Mariano and Murasawa (2003) and incorporate quarterly variables into the framework by constructing for each of them a partially observed monthly counterpart.

Lets us explain it on the example of GDP. In what follows we adopt the convention in which the value of the quarterly variable is “assigned” to the third month of the respective quarter. Accordingly, quarterly level of GDP, which we denote by GDP_t^Q, $t = 3, 6, 9, ...$, can be expressed as the sum of its unobserved monthly contributions, GDP_t^M:

$$GDP_t^Q = GDP_t^M + GDP_{t-1}^M + GDP_{t-2}^M \qquad t = 3, 6, 9, ...$$

[3]Explicit modelling of the dynamics of idiosyncratic component can be also useful to forecast variables with strong non-common dynamics.

Let us define $Y_t^Q = 100 \times \log GDP_t^Q$ and $Y_t^M = 100 \times \log GDP_t^M$. We assume that the unobserved monthly growth rate of GDP, $y_t = \Delta Y_t^M$, admits the same factor model representation as the monthly real variables:

$$
\begin{aligned}
y_t &= \mu_Q + \Lambda_Q f_t + \varepsilon_t^Q \,, && (8)\\
\varepsilon_t^Q &= \alpha_Q \varepsilon_{t-1}^Q + e_t^Q \,, \qquad e_t^Q \sim i.i.d.\ N(0, \sigma_Q^2)\,, && (9)
\end{aligned}
$$

with $\Lambda_Q = (\Lambda_{Q,G} \quad 0 \quad \Lambda_{Q,R})$.

To link y_t with the observed GDP data we construct a partially observed monthly series:

$$
y_t^Q = \begin{cases} Y_t^Q - Y_{t-3}^Q, & t = 3, 6, 9, ... \\ \text{unobserved}, & \text{otherwise} \end{cases}
$$

and use the approximation of Mariano and Murasawa (2003):

$$
\begin{aligned}
y_t^Q &= Y_t^Q - Y_{t-3}^Q \approx (Y_t^M + Y_{t-1}^M + Y_{t-2}^M) - (Y_{t-3}^M + Y_{t-5}^M + Y_{t-5}^M) \\
&= y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4}, \qquad t = 3, 6, 9, ... && (10)
\end{aligned}
$$

3.3 Estimation and forecasting

Let us define $\bar{x}_t = (x_t', y_t^Q)'$ and $\bar{\mu} = (\mu', \mu_Q)'$. The joint model specified by the equations (5)-(10) can be cast in a state space representation:

$$
\begin{aligned}
&\bar{x}_t = \bar{\mu} + Z(\theta)\alpha_t \,, \\
&\alpha_t = T(\theta)\alpha_{t-1} + \eta_t \,, \qquad \eta_t \sim \ i.i.d.\ N(0, \Sigma_\eta(\theta))\,, && (11)
\end{aligned}
$$

where the vector of states includes the common factors and the idiosyncratic components. In the case $p \leq 5$, we have

$$
\alpha_t = (f_t', f_{t-1}', f_{t-2}', f_{t-3}', f_{t-4}', \varepsilon_{1t}, \ldots, \varepsilon_{nt}, \varepsilon_t^Q, \varepsilon_{t-1}^Q, \varepsilon_{t-2}^Q, \varepsilon_{t-3}^Q, \varepsilon_{t-4}^Q)'.
$$

All the parameters of the model, $\bar{\mu}$, Λ, Λ_Q, A_1, Q, $\alpha_1, \ldots, \alpha_n$, α_Q, $\sigma_1, \ldots, \sigma_n$, σ_Q, are collected in θ. The details of the state space representation, and in particular the structure of the matrices, $Z(\theta)$, $T(\theta)$ and $\Sigma_\eta(\theta)$, are provided in the Appendix.[4]

In this paper, we estimate θ by maximum likelihood implemented by Expectation Maximisation (EM) algorithm. This approach has been proposed for large datasets by Doz, Giannone, and Reichlin (2006a) and extended by Bańbura and Modugno (2009) to deal with missing

[4]For the sake of simplicity in the presentation we assume that there is only 1 quarterly variable, GDP. However, it is straightforward to incorporate more quarterly variables, see the Appendix.

observations and idiosyncratic dynamics. Roughly speaking, the maximum likelihood estimation using the EM algorithm consists in iterating the two-step approach of Doz, Giannone, and Reichlin (2006b) and Giannone, Reichlin, and Small (2008)- estimating the parameters conditional on the factor estimates from previous iteration and vice versa.

Contrary to popular non-parametric method based on principal components,[5] maximum likelihood allows us to easily deal with such features of the model as substantial fraction of missing data, restrictions on the parameters or serial correlation of the idiosyncratic component. In addition, as we also study models of moderate sizes (less than 30 variables), maximum likelihood approach should be more efficient. Finally, in this framework, it is straightforward to introduce factors that are specific to a subgroup of variables, see above. The details of the EM iterations, following Bańbura and Modugno (2009), are given in the Appendix.

Given an estimate of θ, the nowcasts as well as the estimates of the factors or of any missing observations in $\bar{x}_t$, can be obtained from Kalman filter and smoother. In other words, under the assumption that the data generating process is given by (11) with θ equal to its QML estimate, Kalman filter and smoother can be used to obtain, in an efficient and automatic manner, projection (1) for any pattern of data availability in Ω_v. One way to understand how Kalman filter deals with missing data is to imagine that it simply discards the rows in $\bar{x}_t$ and $Z(\theta)$ that correspond to the missing observations in the former vector.

In addition, the *news* I_{v+1} and expectations needed to compute $b_{j,t,v+1}$ in (4) can be also easily retrieved from the Kalman smoother output, see Bańbura and Modugno (2009) for details. It is worth noting that for t large enough so that Kalman filter has approached its steady state, the weights $b_{j,t,v+1}$ will not depend on particular realisation of $\{x_{j,T_{j,v+1}}, j \in J_{v+1}\}$ but only on θ and on the shape of ragged edge in Ω_v and Ω_{v+1}.

4 Related Literature

Our approach, described in the previous section, is based on a factor model cast in state space and a particular solution to the problem of missing data and mixed frequency. Recent applications of the factor model approach are Angelini, Camba-Méndez, Giannone, Rünstler, and Reichlin (2008), Rünstler, Barhoumi, Cristadoro, Reijer, Jakaitiene, Jelonek, Rua, Ruth, Benk, and Nieuwenhuyze (2008), Bańbura and Rünstler (2010), Camacho and Perez-Quiros (2008), Marcellino and Schumacher (2008) amongst others. The model by Evans (2005) is

[5] See e.g. Bańbura and Rünstler (2010); Giannone, Reichlin, and Small (2008); Stock and Watson (2002)

similar in spirit and, roughly speaking, consists in assuming that GDP is the only unobservable factor.

In this section we briefly review alternative solutions to the problems of jagged edged data and mixed frequency which have been offered by the literature and used for the nowcasting problem.

Let us start with the mixed frequency problem. The traditional approach in nowcasting is the bridge equations solution, which consists in estimating the model at the lowest frequency after having aggregated monthly data to quarterly and then plugging them into a bridge equation to obtain an estimate of the target quarterly variable. Early applications of bridge equations are in Trehan (1989), Parigi and Schlitzer (1995) and examples of more recent applications are Parigi and Golinelli (2007) Rünstler and Sédillot (2003) and Diron (2006) amongst others. More recently, in a literature which does not focus on nowcasting in particular, Clements and Galvao (2009) and Marcellino and Schumacher (2008), use the MIDAS approach proposed by Ghysels, Santa-Clara, and Valkanov (2004). The idea consists in regressing quarterly GDP on distributed lags of variables sampled at higher frequency and using a weighting function for the lags to insure parsimony. Both these methods do not fully account for dynamic correlations between the regressors and hence the outcome cannot be interpreted in terms of updates and *news*. A third solution, proposed in Giannone, Reichlin, and Small (2008), consists in bridging quarterly GDP with unobserved factors estimated from a panel of monthly data and then aggregated to quarterly frequency. In this paper,as we have seen, we have used a different approach, first proposed by Mariano and Murasawa (2003) in a context of a model aimed at constructing a coincident index of aggregate economic activity rather than at nowcasting. The solution consists in considering lower frequency variables as being periodically missing and using a linear projection to obtain an approximation of the observed quarterly growth rate. In the context of nowcasting, the Mariano and Murasawa (2003)'s approach has been used by Angelini, Camba-Méndez, Giannone, Rünstler, and Reichlin (2008) and Bańbura and Modugno (2009) amongst others. As explained in the previous section, once the model is cast in state space form, the Kalman filter can be used to compute the projection. A recent paper, Ghysels and Wright, 2009, compare the Kalman filter approach with the MIDAS approach[6].

For what concerns the problem of jagged edged data, two solutions have been offered. A first approach consists in using auxiliary forecasting models for each of the monthly series, or

[6]Mitchell, Smith, Weale, Wright, and Salazar (2005) and Proietti (2008) propose alternative approaches which do not use approximation (10) and ensure that the sum of estimates of the monthly levels of GDP is consistent with the observed quarterly figure.

for their subgroups. The predictions produced by these models are used to "fill in" missing information. This approach is typically used in bridge equations in order to close the quarter of interest. The same approach is also used by the Chicago FED for the construction of a coincident index of economic activity (ChicagoFED (2001)). Alternatively, one can shift the data in order obtain a dataset that is artificially balanced at the end of the sample. For example, if there is one more month available for surveys than for industrial production, we can realign the two series by dating industrial production at time t rather than $t-1$. Such approach has been used for the construction of the coincident indicators EUROCOIN (Altissimo, Bassanetti, Cristadoro, Forni, Hallin, Lippi, Reichlin, and Veronese (2001), Altissimo, Cristadoro, Forni, Lippi, and Veronese (2006)). In this case, however, the model used for the projection is not time invariant since it changes with the pattern of data availability. For this reason the nowcast cannot be expressed as a function of well defined and model consistent *news*.[7]

The solution for missing data proposed by Bańbura and Modugno (2009) and used in this paper is similar to what proposed by Stock and Watson (2002), who developed an algorithm for the extraction of principal components from panels with missing data and mixed frequency. Their approach, however, is not well suited for the understanding of different updates as new data become available since, when filling the data, one only considers cross-sectional dependence, but not the time dependence. Schumacher and Breitung (2008) apply Stock and Watson (2002)'s approach to nowcast German GDP from monthly data and forecast the missing data for the end of the quarter via auxiliary forecasting model (VAR) for the factors.

Finally, let us stress that our solution to the problem of missing data and mixed frequency can be handled by any model that can be cast in state space form. We have chosen the factor model for reasons of parsimony, but the Vector Auto Regression (VAR) is an obvious alternative. For example, the state space form of the VAR has been used for nowcasting in (see Zadrozny (1990) and Giannone, Reichlin, and Simonelli (2009)). This approach, however, can only handle few series.

5 Empirical results

In this section we illustrate the ideas developed above by employing the model described in the previous section to forecasting of quarter-on-quarter GDP growth and of year-on-year inflation.

[7]These methods have been compared empirically with the Kalman filter solution proposed by Giannone, Reichlin, and Small (2008) and used in this paper by Marcellino and Schumacher (2008) and Rünstler, Barhoumi, Cristadoro, Reijer, Jakaitiene, Jelonek, Rua, Ruth, Benk, and Nieuwenhuyze (2008).

The purpose is to illustrate how the real time data flow shapes the evolution of consecutive forecast updates. More precisely, we examine how releases of different groups of data revise the forecast and affect the associated forecast uncertainty.

For each target variable and each reference period, we consider a sequence of forecast updates. These are produced twice a month at dates which correspond approximately, to releases of major groups of hard and soft data (in the middle and at the end of each month, respectively).

We are also interested in the role of more disaggregated sectoral data. To this end we compare the performance of a benchmark model that contains mainly aggregated data with the results from a richer dataset including sectorial information. Such disaggregated data are routinely monitored by sectorial experts and can be important not only to eventually improve forecast accuracy but also for understanding and interpreting the forecasts. Most of the factor models used in central banks for nowcasting are based on large disaggregated data sets. However, sectorial information can lead to model mis-specification in small samples since it introduces idiosyncratic cross-correlation. Hence, the comparison is interesting to understand the robustness of the model with respect to the inclusion of many variables.

In all the exercises we assume 1 global, 1 real and 1 nominal factor and number of lags in the factor VAR $p = 1$.

5.1 Data set

Let us first comment on the data set for our benchmark model. It contains twenty-six major indicators on the euro area economy. The series are presented in Table 1. As mentioned above, most of the series relate to the total economy. The only exception are surveys which are disaggregated into major sectors. This can be important as surveys are the only monthly source of information on services.

The data set contains mainly monthly series and such is the frequency of our model. Data with native frequency higher than monthly are aggregated as monthly averages. The exception are commodity prices, which enter as 15-day averages and hence, for a given month, are available already in its middle.[8]

In the table we also report respective publication delays (in days). There are substantial differences between the series in terms of their timeliness. For example survey and financial

[8] Monthly averages would have been smoother but also less timely. Empirical results indicate that considering more timely information on commodity prices is more optimal for inflation. More systematic analysis on inclusion of higher frequency is left for future research.

series, which are sometimes labelled as soft data, are already available at the end of the respective reference period (or even couple of days before). In contrast, hard data on real activity are released with 2-3 months delay. However, they typically carry a more precise signal for GDP developments. Since there is likely to be a tradeoff between timeliness and precision, the data set is constructed to contain both "timely" soft data and "precise" hard data. The last two columns of Table 1 report the (stylised) data availability patterns, or the "shape" of the ragged edge", that we apply for the bi-monthly forecast updates. It should be noted that our exercises are pseudo real time, that is, while we observe the real time publication delays, we do not take into account the real time data revisions.

The disaggregated data set contains a sectoral split for industrial production, more detailed labor market information as well as few more quarterly series. A detailed list is provided in the Appendix.

5.2 Forecast updates and news

As an illustration, we first produce a sequence of forecast updates for GDP growth rate in the fourth quarter of 2008 and for the yearly inflation in 2008. Since inflation is available at monthly frequency and with short publication lags, it is not our focus. However, having a model that can consider jointly prices and quantities is potentially useful for interpreting results.

For the GDP we consider bi-monthly updates of next, current and previous quarter forecasts. Specifically, we produce a first forecast with data available in mid July 2008 and we subsequently update it at two-week intervals, each time incorporating new data releases. The resulting six updates performed from July till September target the next-quarter GDP growth. With the update from mid-October till end-December we effectively project current quarter GDP growth. The last two updates are performed in January 2009 and they refer to the previous quarter (the flash estimate for 2008 Q4 GDP was released in mid February). In some applications next, current and previous quarter forecasts are labelled as "forecasts", "nowcasts" and "backcasts", respectively.

Concerning HICP, we proceed in a similar manner. We produce the first forecast in mid-July and we update it twice a month up to end of December 2008 (HICP is typically released around two weeks after the end of the reference period).[9]

[9]Using the logarithmic approximation of a growth rate, yearly inflation can be expressed as a sum of 12 month-on-month growth rates of prices. Since prices enter the dataset as month-on-month growth rates, the forecast for yearly inflation is obtained as sum of partially observed and partially forecast month-on-month

The evolution of the forecast for both variables as produced by our model is depicted in Figure 1. In the same chart, we report the contribution of the *news* component of the various data groups to the forecast revision.[10] As explained in Section 2, the difference between two consecutive forecasts, i.e. the forecast revision, is the sum over all the released variables of the product of the *news* related to a particular variable and the associated weight in the GDP estimate (see equation (4)). The contribution of the *news* from a block of variables is the sum of contributions of the series belonging to this block. The composition of different blocks is indicated in the second column of Table 1. To make the graphs easier to read, certain groups have been merged. In the case of GDP forecast, e.g. all nominal variables constitute a single group.

Let us comment on the evolution of the GDP forecast. At the beginning of the forecasting period the forecast remains rather flat, corroborating the above mentioned difficulties in forecasting beyond the current quarter. The first substantial downward revision (pointing to a negative GDP growth) comes with the release of surveys for October, which is the first block of real data referring to the current quarter. This negative *news* in October is confirmed by subsequent data, both surveys and hard data. In fact, with all subsequent releases the forecasts are revised downwards. In addition, later in the reference quarter, the *news* from the hard data block become more sizeable. This is in line with the results of Giannone, Reichlin, and Small (2008) and Bańbura and Rünstler (2010) who show that less timely hard data become important only later in the forecast sequence. The contribution of the nominal block is rather limited throughout the whole forecast cycle.

Concerning HICP inflation, the largest revisions are caused by the releases of HICP itself and of commodity prices. These seem to be the most informative data sources on the short-term developments in inflation. In contrast, the contribution of the *news* from the surveys on prices and from the real block is relatively small. The same is true for *news* on other nominal variables such as money, exchange rate or interest rates.

5.3 Forecast uncertainty

Uncertainty around the nowcast related to signal extraction at any point in time can be easily evaluated using Kalman filtering techniques (see Giannone, Reichlin, and Small, 2008). How-

growth rates. For example in mid-July we already observe the monthly growth rates for the first half of the year and need to forecast only the remaining 6 months.

[10] In this exercise we abstract from the effect of parameter re-estimation. For each forecast sequence the parameters are estimated only once before the first forecast in the sequence is made and kept constant for all the subsequent forecast updates.

ever, these estimates only hold under the the assumption that errors are Gaussian and that the model is well specified. To overcome these limitations we will assess forecast uncertainty by evaluating the average historical performances of the model.

To this end, we perform a simulated pseudo real time forecasting exercise. This means that at each point in time we estimate the parameters of the model and produce forecasts using the data that replicates the pattern of data availability at the time. Estimating the model recursively takes into account estimation uncertainty.

We are, in particular, interested in how uncertainty evolves as the information related to the target period accumulates. Since the bi-monthly updates described in the previous section differ in terms of available information, we examine the average accuracy for each of them separately. As the measure of uncertainty we choose the Root Mean Squared Forecast Error (RMSFE) and we evaluate it over the period 2000-2007. The resulting uncertainty for our benchmark model is depicted in Figure 2. On the x-axis we use the same labels as in Figure 1 to indicate that the average uncertainty was computed with the same data availability assumptions, relative to the target period. There is a slight difference in the chart for inflation as for RMSFE we also consider longer forecast horizons.

For comparison we plot the same average uncertainty measure for forecasts produced by univariate naïve models. For GDP it is random walk with drift for levels of logged GDP. For HICP it is driftless random walk for year-on-year inflation.

We can observe that, as the information accumulates, the gains in forecast accuracy are substantial. For GDP the RMSFE is reduced by 50% as we move from the first to the last forecast in the sequence. For "earlier" forecasts larger gains are obtained when surveys are released. When hard data for the reference quarter become available, surveys lose their importance. This suggests that soft data are relevant due to their timeliness but, conditionally on the availability of hard data for the same reference period, they are uninformative. This confirms the results in Giannone, Reichlin, and Small (2008), Bańbura and Rünstler (2010) and Matheson (2010).

We also note that the uncertainty measures associated with next forecasts for the benchmark and naïve model are comparable, confirming earlier results about the difficulties of forecasting beyond the current quarter. This also applies to institutional forecasts (see Giannone, Reichlin, and Small, 2008).

Decreasing uncertainty corresponding to the inclusion of the newly published data as we proceed throughout the quarter is also true for HICP inflation. We gain in forecast accuracy mostly due to mid-month releases, corresponding to the release of the HICP itself and of commodity prices.

Finally let us compare the results with forecast accuracy of the model including more disaggregated data. Table 2 reports the corresponding RMSFE based uncertainty. We also recall the results for the benchmark and random walk models and in addition consider autoregressive univariate models.

The exercise based on disaggregated data shows that including more variables does not improve the accuracy of the forecast but does not affect its stability. Since, in e.g. the preparation of policy briefings, it might be necessary to comment on many releases including disaggregated data, this is good news. Our framework is robust to the inclusion of a rich data set.

6 Conclusions

In this paper we define *nowcasting* as the prediction of the present, the very near future and the very recent past.

Key in this process is to use timely monthly information in order to nowcast quarterly variables that are published with long delays. We have argued that the nowcasting process goes beyond the simple production of an early estimate and it consists in the analysis of the link between the *news* in consecutive data releases and the resulting forecast revisions for the target variable. We have described an econometric framework which is designed for this analysis. In this framework all variables are considered endogenous and hence a meaningful model based *news* can be extracted and the revisions of GDP estimates can be expressed as a function of these *news*.

The framework is based on the factor model proposed and analysed by Doz, Giannone, and Reichlin (2006b) and Doz, Giannone, and Reichlin (2006a), adapted for nowcasting by Giannone, Reichlin, and Small (2008) and by Bańbura and Modugno (2009). This model has been successfully used in many applications both in central banks and in the academic literature since it provides a parsimonious solution to the problem of modeling many data releases and therefore allows to mimic, via a coherent statistical model, the judgemental process of nowcasting traditionally conducted in policy institutions. To illustrate these ideas, we provide an application for the nowcast of euro area GDP in the fourth quarter of 2008 and we also present results for annual inflation in 2008.

Factor models are not the only solution to the problem of nowcasting. In principle, any dynamic model that can handle mixed frequencies and missing observations and that can capture the joint dynamic of the target and the data releases can be used. Different examples in the literature are Evans (2005) or the VAR proposed by Zadrozny (1990) and Giannone, Reichlin, and Simonelli (2009). VAR models, however, are not suitable to handle more than few series. A promising line for future research is to build on ideas in Bańbura, Giannone, and Reichlin (2010) to develop nowcasting tools based on VARs where Bayesian shrinkage is used to cope with the curse of dimensionality problem.

Other ideas for further research consist in linking the high frequency nowcasting with a quarterly structural model in a model coherent way. Giannone, Monti, and Reichlin (2009) have suggested a solution and other developments are in progress. A byproduct of this analysis is that one can obtain real time estimates of variables that can only be defined theoretically such as the output gap or the natural rate of interest.

We should stress that, related to nowcasting, is a literature on coincident indicators of economic activity where, rather than focusing on an early estimate of GDP, an unobserved state of the economy is estimated from a multivariate model. A classic paper in this field is Stock and Watson (1989). More recently, new ideas on how to construct these indicators have led to the Eurocoin index for the euro area (Altissimo, Bassanetti, Cristadoro, Forni, Hallin, Lippi, Reichlin, and Veronese, 2001; Altissimo, Cristadoro, Forni, Lippi, and Veronese, 2006) and the Chicago Fed index for the US (ChicagoFED, 2001). Aruoba, Diebold, and Scotti (2009) are posting a similar index which also contains high frequency financial data in the Philadelphia Fed website. Although some of the problems in this literature are related to those described above for nowcasting, in this chapter we do not review this literature and limit the discussion to pure nowcasting defined as timely estimation and its analysis of a particular target variable such as GDP.

Finally, this chapter does not contain a discussion on how to extend the model to incorporate data at frequencies higher than monthly. The model we base our discussion on can be updated at any frequency (minute, day, week, ...) as data are released but includes only monthly and quarterly variables. Financial variables, for example, are aggregated to monthly and treated as being released only when information on the entire month is available. Although the model can be adapted to properly take into account high frequency data, this is still unfinished work. Aruoba, Diebold, and Scotti (2009) is a first attempt to deal with this problem. They use a small factor model and apply it to the construction of a coincident indicator of the state of the economy rather than to the nowcasting problem. Andreu, Ghysels, and Kourtellos (2008) propose an alternative approach based on MIDAS but treat the predictors as exogenous and focus on forecasting. The challenge is to model higher frequency within a joint model in order to maintain the ability of understanding the nowcast updates.

References

Aastveit, K. A., and T. G. Trovik (2008): "Nowcasting Norwegian GDP: The role of asset prices in a small open economy," Working Paper 2007/09, Norges Bank.

Altissimo, F., A. Bassanetti, R. Cristadoro, M. Forni, M. Hallin, M. Lippi, L. Reichlin, and G. Veronese (2001): "EuroCOIN: A Real Time Coincident Indicator of the Euro Area Business Cycle," CEPR Discussion Papers 3108, C.E.P.R. Discussion Papers.

Altissimo, F., R. Cristadoro, M. Forni, M. Lippi, and G. Veronese (2006): "New EuroCOIN: Tracking Economic Growth in Real Time," CEPR Discussion Papers 5633.

Andreu, E., E. Ghysels, and A. Kourtellos (2008): "Should macroeconomic forecasters look at daily financial data?," Manuscript, University of Cyprus.

Angelini, E., G. Camba-Méndez, D. Giannone, G. Rünstler, and L. Reichlin (2008): "Short-term forecasts of euro area GDP growth," Working Paper Series 949, European Central Bank.

Aruoba, S., F. X. Diebold, and C. Scotti (2009): "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics*, 27(4), 417–27.

Bańbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian VARs," *Journal of Applied Econometrics*, 25(1), 71–92.

Bańbura, M., and M. Modugno (2009): "Maximum likelihood estimation of large factor model on datasets with arbitrary pattern of missing data.," Working paper series, European Central Bank, forthcoming.

Bańbura, M., and G. Rünstler (2010): "A look into the factor model black box. Publication lags and the role of hard and soft data in forecasting GDP.," *International Journal of Forecasting*, forthcoming.

Camacho, M., and G. Perez-Quiros (2008): "Introducing the EURO-STING: Short Term INdicator of Euro Area Growth," Banco de España Working Papers 0807, Banco de España.

ChicagoFED (2001): "CFNAI Background Release," Discussion paper, http://www.chicagofed.org/economicresearchanddata/national/pdffiles/CFNAI bga.pdf.

Clements, M. P., and A. B. Galvao (2009): "Forecasting US output growth using leading indicators: an appraisal using MIDAS models," *Journal of Applied Econometrics*, 24(7), 1187–1206.

Cutler, D. M., J. M. Poterba, and L. H. Summers (1989): "What Moves Stock Prices?," *Journal of Portfolio Management*, 15, 4–12.

D'Agostino, A., K. McQuinn, and D. O'Brien (2008): "Now-casting Irish GDP," Research Technical Papers 9/RT/08, Central Bank & Financial Services Authority of Ireland (CBFSAI).

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum Likelihood Estimation From Incomplete Data," 14, 1–38.

Diron, M. (2006): "Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data," Working Paper Series 622, European Central Bank.

Doz, C., D. Giannone, and L. Reichlin (2006a): "A Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," Working Paper Series 674, European Central Bank.

——— (2006b): "A two-step estimator for large approximate dynamic factor models based on Kalman filtering," Unpublished manuscript, Université Libre de Bruxelles.

Evans, M. D. D. (2005): "Where Are We Now? Real-Time Estimates of the Macroeconomy," *International Journal of Central Banking*, 1(2).

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004): "The MIDAStouch: MIxed Data Sampling Regression Models," mimeo, Chapel Hill, N.C.

Giannone, D., F. Monti, and L. Reichlin (2009): "Incorporating Conjunctural Analysis in Structural Models," in *The Science and Practice of Monetary Policy Today*, ed. by V. Wieland, pp. 41–57. Springer, Berlin.

Giannone, D., L. Reichlin, and L. Sala (2004): "Monetary Policy in Real Time," in *NBER Macroeconomics Annual*, ed. by M. Gertler, and K. Rogoff, pp. 161–200. MIT Press.

Giannone, D., L. Reichlin, and S. Simonelli (2009): "Nowcasting Euro Area Economic Activity in Real-Time: The Role of Confidence Indicator," *National Institute Economic Review*, 210, 90–97.

Giannone, D., L. Reichlin, and D. Small (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55(4), 665–676.

Marcellino, M., and C. Schumacher (2008): "Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP," CEPR Discussion Papers 6708, C.E.P.R. Discussion Papers.

Mariano, R., and Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18, 427–443.

Matheson, T. D. (2010): "An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys," *Economic Modelling*, 27(1), 304–314.

Mitchell, J., R. J. Smith, M. R. Weale, S. Wright, and E. L. Salazar (2005): "An Indicator of Monthly GDP and an Early Estimate of Quarterly GDP Growth," *Economic Journal*, 115(501), F108–F129.

Parigi, G., and R. Golinelli (2007): "The use of monthly indicators to forecast quarterly GDP in the short run: an application to the G7 countries," *Journal of Forecasting*, 26(2), 77–94.

Parigi, G., and G. Schlitzer (1995): "Quarterly forecasts of the italian business cycle by means of monthly indicators," *Journal of Forecasting*, 14(2), 117–141.

Proietti, T. (2008): "Estimation of Common Factors under Cross-Sectional and Temporal Aggregation Constraints: Nowcasting Monthly GDP and its Main Components," Manuscript.

Rünstler, G., K. Barhoumi, R. Cristadoro, A. D. Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, S. Benk, and C. V. Nieuwenhuyze (2008): "Short-term forecasting of GDP using large monthly data sets: a pseudo real-time forecast evaluation exercise," Occasional Paper Series No 84, European Central Bank, forthcoming on the International Journal of Forecasting.

Rünstler, G., and F. Sédillot (2003): "Short-Term Estimates Of Euro Area Real Gdp By Means Of Monthly Data," Working Paper Series 276, European Central Bank.

Schumacher, C., and J. Breitung (2008): "Real-time forecasting of German GDP based on a large factor model with monthly and quarterly Data," *International Journal of Forecasting*, 24, 386–398.

Shumway, R., and D. Stoffer (1982): "An approach to time series smoothing and forecasting using the EM algorithm," *Journal of Time Series Analysis*, 3, 253–264.

Stock, J. H., and M. W. Watson (1989): “New Indexes of Coincident and Leading Economic Indicators,” in *NBER Macroeconomics Annual*, ed. by O. J. Blanchard, and S. Fischer, pp. 351–393. MIT Press.

——— (2002): “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economics Statistics*, 20(2), 147–162.

Trehan, B. (1989): “Forecasting growth in current quarter real GNP,” *Economic Review*, (Win), 39–52.

Watson, M. W., and R. F. Engle (1983): “Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models,” *Journal of Econometrics*, 23, 385–400.

Zadrozny, P. (1990): “Estimating a multivariate ARMA model with mixed-frequency data: an application to forecating U.S. GNP at monthly intervals,” Working Paper Series 90-6, Federal Reserve Bank of Atlanta.

Table 1: Data set

No	Group	Series	Frequency	Publication delay (in days after reference period)	No of missing observations mid month	end month
1	Real, Hard data	IP, total industry	Monthly	37-40	2	2
2	Real, Hard data	IP, manufacturing	Monthly	32-35	2	2
3	Real, Hard data	Retail trade, except for motor vehicles and motorcycels	Monthly	33-39	2	2
4	Real, Hard data	New passenger car registrations	Monthly	15-17	1	1
5	Real, Hard data	New orders, manufacturing working on orders	Monthly	42-45	3	2
6	Real, Hard data	Extra euro area trade, export, value	Monthly	44-48	3	2
7	Real, Hard data	Unemployment rate, total	Monthly	29-32	2	2
8	Real, Hard data	Index of employment, total industry	Monthly	80-140	4,5,3	4,3,3
9	Real, Surveys	Purchasing manager index, manufacturing	Monthly	-	1	0
10	Real, Surveys	Purchasing managers survey, services, business activity	Monthly	-	1	0
11	Real, Surveys	Consumer survey, consumer confidence indicator	Monthly	-	1	0
12	Real, Surveys	Industry survey, industrial confidence indicator	Monthly	-	1	0
13	Real, Surveys	Retail trade survey, retail confidence indicator	Monthly	-	1	0
14	Real, Surveys	Services survey, services confidence indicator	Monthly	-	1	0
15	Nominal, HICP	HICP, overall index	Monthly	15-18	1	1
16	Nominal, PPI	PPI, total industry excluding construction	Monthly	32-35	2	2
17	Nominal, Surveys	Consumer survey, price trends over next 12 months	Monthly	-	1	0
18	Nominal, Surveys	Industry survey, selling price expectations for the months ahead	Monthly	-	1	0
19	Nominal, Money	M3, index of notional stocks	Monthly	25-29	2	1
20	Nominal, Money	Index of loans	Monthly	25-29	2	1
21	Real, Financial	Dow Jones Euro Stoxx, broad stock exchange index	Daily	-	1	0
22	Nominal, Financial	Euribor 3 months	Daily	-	1	0
23	Nominal, Financial	Nominal effective exch. rate, core group of currencies against euro	Daily	-	1	0
24	Nominal, Comm prices	Raw materials excl. energy, market prices	Daily	-	1	0
25	Nominal, Comm prices	Raw materials, crude oil, market prices	Daily	-	1	0
26	Real, GDP	Gross domestic product, chain linked	Quarterly	42-44	4,2,3	4,2,3

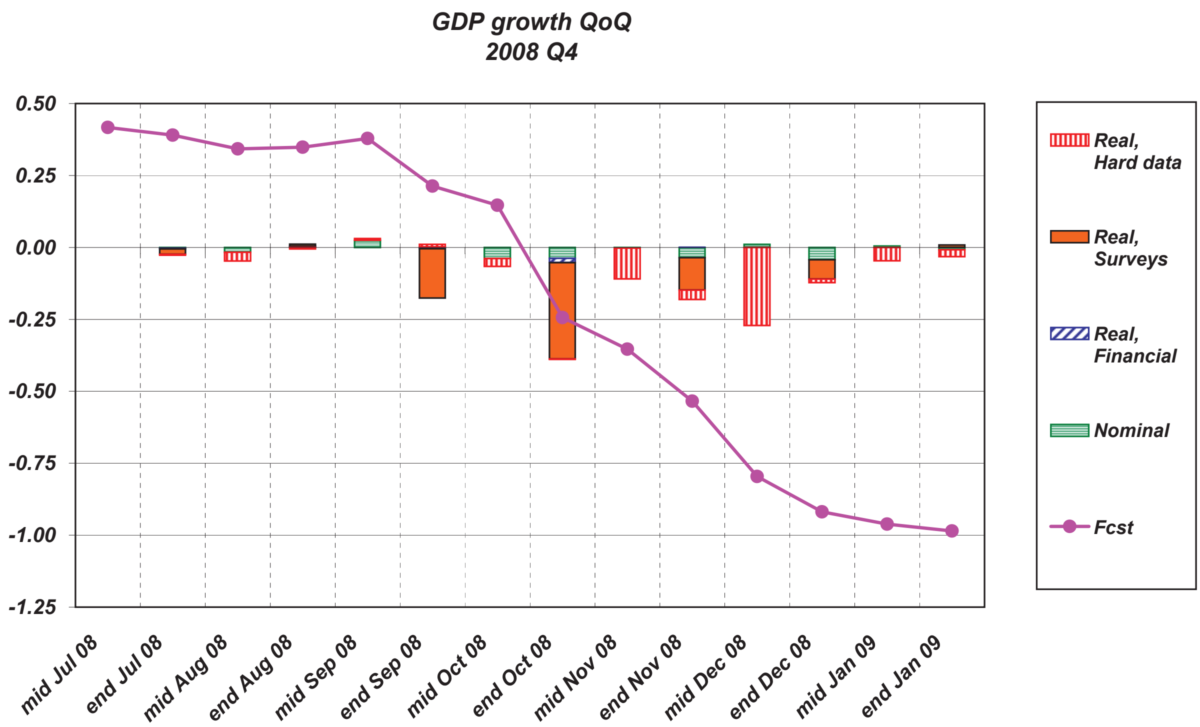
Notes: Fifth column of the table indicates typical publication delay (in days) for each series. It may vary from month on month, depending e.g. on configuration of business days. "-" means "no publication delay", such series are available directly at the end of the reference period (or couple of days before). Based on a typical publication delay, we report in sixth and seventh column a (stylised) number of missing observations at the end of the sample in the middle and at the end of a month. For employment index and GDP we report 3 numbers (corresponding to first, second and third month of each quarter) as these series are not released each month. We use these data availability patterns in the forecast evaluation exercises.

Table 2: Forecast uncertainty

	GDP				HICP				
	Ben	Disagg	RW	AR		Ben	Disagg	RW	AR
mid Jul 08	0.28	0.29	0.32	0.33	mid Dec 07	0.58	0.59	0.72	0.67
end Jul 08	0.27	0.27	0.32	0.33	end Dec 07	0.58	0.59	0.72	0.67
mid Aug 08	0.27	0.27	0.32	0.32	mid Mar 08	0.51	0.52	0.63	0.60
end Aug 08	0.26	0.25	0.32	0.32	end Mar 08	0.51	0.52	0.63	0.60
mid Sep 08	0.25	0.26	0.32	0.32	mid Jun 08	0.43	0.46	0.55	0.52
end Sep 08	0.25	0.25	0.32	0.32	end Jun 08	0.43	0.46	0.55	0.52
mid Oct 08	0.24	0.24	0.32	0.32	mid Sep 08	0.33	0.39	0.52	0.47
end Oct 08	0.23	0.23	0.32	0.32	end Sep 08	0.34	0.38	0.52	0.47
mid Nov 08	0.21	0.23	0.31	0.27	mid Oct 08	0.29	0.33	0.48	0.40
end Nov 08	0.21	0.21	0.31	0.27	end Oct 08	0.29	0.33	0.48	0.40
mid Dec 08	0.20	0.22	0.31	0.27	mid Nov 08	0.22	0.23	0.39	0.31
end Dec 08	0.20	0.21	0.31	0.27	end Nov 08	0.22	0.23	0.39	0.31
mid Jan 09	0.18	0.20	0.31	0.27	mid Dec 08	0.13	0.13	0.24	0.19
end Jan 09	0.18	0.20	0.31	0.27	end Dec 08	0.13	0.13	0.24	0.19

Notes: Table provides forecast uncertainty for quarter-on-quarter GDP and year-on-year HICP for different models. *Ben* refers to the benchmark model with 26 variables, see Table 1; *Disagg* refers to the specification with more disaggregated data, see the Appendix. *RW* denotes random walk with drift model for levels of logged GDP and random walk without drift for year-on-year inflation. *AR* refers to an autoregressive model for quarterly growth rates of GDP and monthly growth rates of HICP. Uncertainty is given by the Root Mean Squared Forecast Error evaluated in the period 2000-2007. Dates in the first columns indicate data availability patterns with respect to the reference period of 2008 Q4 for GDP and 2008 for yearly inflation. These availability patterns were applied recursively in the forecast evaluation.

Figure 1: Contribution of news to forecast revisions



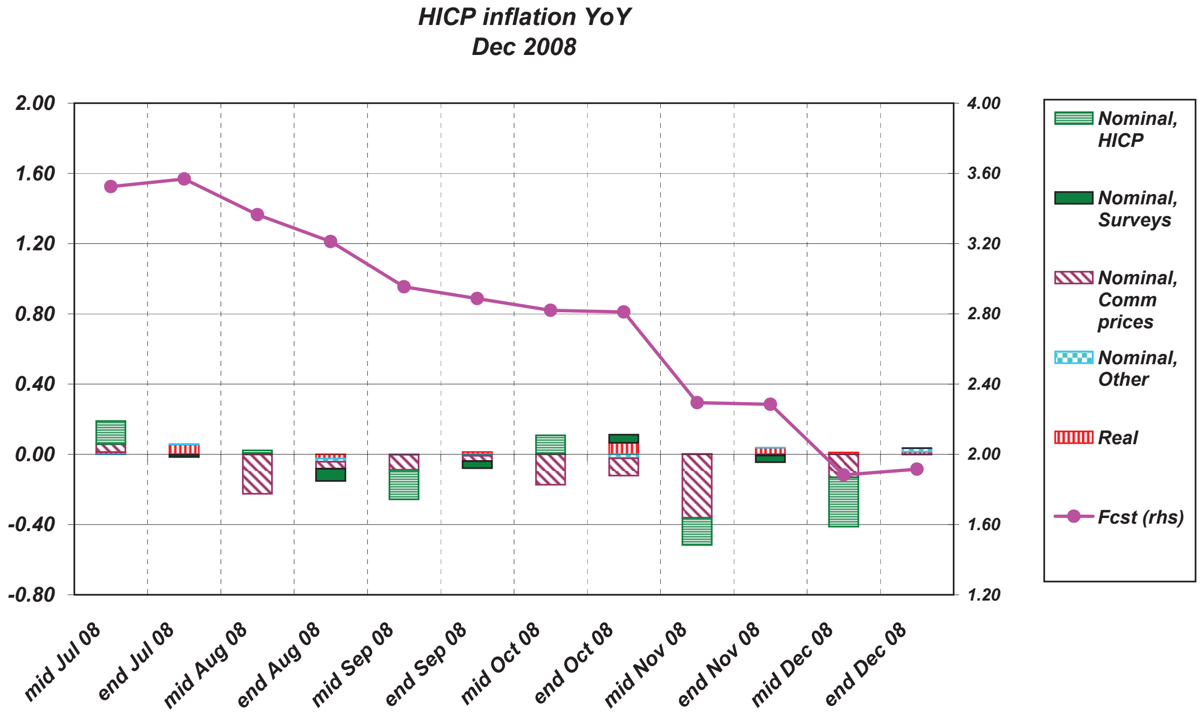
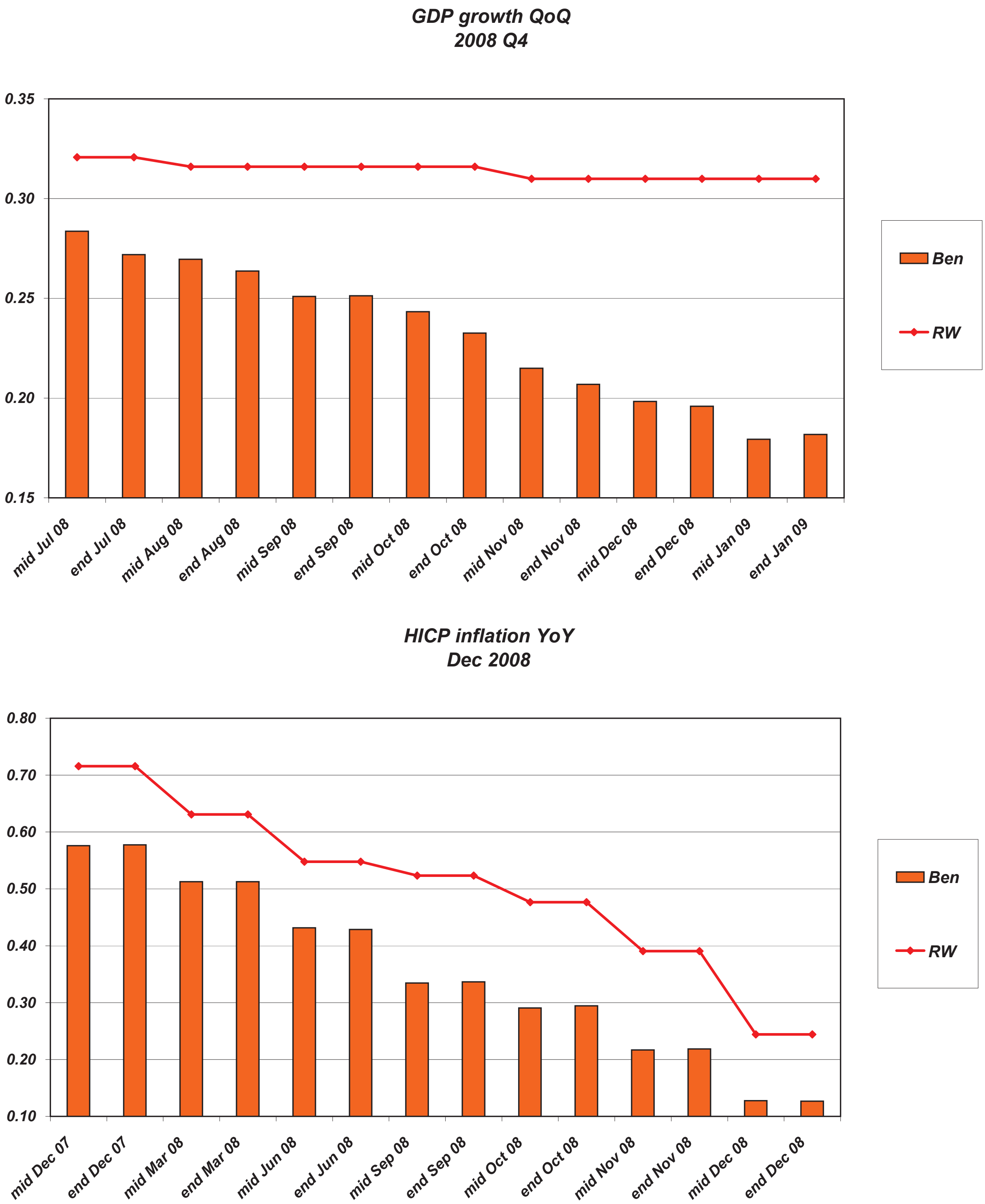


Figure 2: Unconditional uncertainty around the forecast



A Description of the data set

No	Group	Series	Frequency	No of missing observations		Specification		Transf	
				mid month	end month	Ben	Disagg	log	diff
1	Real, Hard data	IP, Total industry	Monthly	2	2	x		x	x
2	Real, Hard data	IP, Manufacturing	Monthly	2	2	x		x	x
3	Real, Hard data	IP, Construction	Monthly	2	2		x	x	x
4	Real, Hard data	IP, Energy	Monthly	2	2		x	x	x
5	Real, Hard data	IP, Intermediate goods industry	Monthly	2	2		x	x	x
6	Real, Hard data	IP, Capital goods industry	Monthly	2	2		x	x	x
7	Real, Hard data	IP, Durable consumer goods industry	Monthly	2	2		x	x	x
8	Real, Hard data	IP, Non-durable consumer goods industry	Monthly	2	2		x	x	x
9	Real, Hard data	Retail trade, except of motor vehicles and motorcycles	Monthly	2	2	x	x	x	x
10	Real, Hard data	New passenger car registraions	Monthly	1	1	x	x	x	x
11	Real, Hard data	New orders, manufacturing working on orders	Monthly	3	2	x	x	x	x
12	Real, Hard data	Extra euro area trade, export value	Monthly	3	2	x	x	x	x
13	Real, Hard data	Intra euro area trade, export, value	Monthly	3	2		x	x	x
14	Real, Hard data	Unemployment rate, total	Monthly	2	2	x			x
15	Real, Hard data	Unemployment rate, 25 years and over	Monthly	2	2		x		x
16	Real, Hard data	Unemployment rate, under 25 years	Monthly	2	2		x		x
17	Real, Hard data	Index of Employment, total industry	Monthly	4,5,3	4,3,3	x		x	x
18	Real, Hard data	Index of Employment, construction	Monthly	4,5,3	4,3,3		x	x	x
19	Real, Hard data	Index of Employment, retail trade, exc.motor vehicles/motorcycles	Monthly	4,5,3	4,2,3		x	x	x
20	Real, Hard data	Index of Employment, manufacturing	Monthly	4,5,3	4,2,3		x	x	x
21	Real, Surveys	Purchasing manager index, manufacturing	Monthly	1	0	x	x		x
22	Real, Surveys	Purchasing managers survey, services, business activity	Monthly	1	0	x	x		x
23	Real, Surveys	Purchasing managers survey, manufacturing, employment	Monthly	1	0		x		x
24	Real, Surveys	Purchasing managers survey, services,employment	Monthly	1	0		x		x
25	Real, Surveys	Eur. Com. survey, industry confidence indicator	Monthly	1	0	x	x		x
26	Real, Surveys	Eur. Com. survey, consumer confidence indicator	Monthly	1	0	x	x		x
27	Real, Surveys	Eur. Com. survey, services confidence indicator	Monthly	1	0	x	x		x
28	Real, Surveys	Eur. Com. survey, retail confidence indicator	Monthly	1	0	x	x		x
29	Real, Surveys	Eur. Com. survey, construction confidence indicator	Monthly	1	0	x	x		x
30	Real, Surveys	Eur. Com. survey, industry employment expectations	Monthly	1	0		x		x
31	Real, Surveys	Eur. Com. survey, consumer unemployment expectations	Monthly	1	0		x		x
32	Real, Surveys	Eur. Com. survey, services employment expectations	Monthly	1	0		x		x
33	Real, Surveys	Eur. Com. survey, retail employment expectations	Monthly	1	0		x		x
34	Real, Surveys	Eur. Com. survey, construction employment expectations	Monthly	1	0		x		x

35	Nominal, HICP	HICP, overall index	Monthly	1	1	x	x	x	x
36	Nominal, HICP	HICP, energy	Monthly	1	1		x	x	x
37	Nominal, HICP	HICP, processed food incl. alcohol and tobacco	Monthly	1	1		x	x	x
38	Nominal, HICP	HICP, unprocessed food	Monthly	1	1		x	x	x
39	Nominal, HICP	HICP, industrial goods excluding energy	Monthly	1	1		x	x	x
40	Nominal, HICP	HICP, services	Monthly	1	1		x	x	x
41	Nominal, PPI	PPI, total industry (excluding construction)	Monthly	2	2	x	x	x	x
42	Nominal, PPI	PPI, manufacture of food products and beverages	Monthly	2	2		x	x	x
43	Nominal, PPI	PPI, MIG energy	Monthly	2	2		x	x	x
44	Nominal, PPI	PPI, consumer goods industry	Monthly	2	2		x	x	x
45	Nominal, Surveys	Eur. Com. survey, industry selling price expectations	Monthly	1	0	x			x
46	Nominal, Surveys	Eur. Com. survey, industry consumer goods selling price expectations	Monthly	1	0		x		x
47	Nominal, Surveys	Eur. Com. survey, industry intemediate goods selling price expectations	Monthly	1	0		x		x
48	Nominal, Surveys	Eur. Com. survey, consumer price trends over last 12 months	Monthly	1	0		x		x
49	Nominal, Surveys	Eur. Com. survey, consumer price trends over next12 months	Monthly	1	0	x	x		x
50	Nominal, Money	M3, index of notional stocks	Monthly	2	1	x	x	x	x
51	Nominal, Money	Index of loans	Monthly	2	1	x	x	x	x
52	Real, Financial	Dow Jones Euro Stoxx, broad stock exchange index	Daily	1	0	x	x	x	x
53	Nominal, Financial	Euribor, 3 months	Daily	1	0	x	x		x
54	Nominal, Financial	10-year govt. bonds	Daily	1	0		x		x
55	Nominal, Financial	Nominal effective exch. rate, core group of currencies against euro	Daily	1	0	x	x	x	x
56	Nominal, Comm prices	Raw materials excl. energy, mkt prices	Daily	1	0	x	x	x	x
57	Nominal, Comm prices	Raw materials, foodstuff and beverages , mkt prices	Daily	1	0		x	x	x
58	Nominal, Comm prices	Raw materials , crude oil, mkt prices	Daily	0	0	x	x	x	x
59	Real, Hard data	Gross domestic product, chain linked	Quarterly	4,2,3	4,2,3	x	x	x	x
60	Real, Hard data	Employment, domestic	Quarterly	4,5,3	4,5,3		x	x	x
61	Real, Surveys	Eur. Com. survey, industry current level of capacity utilization	Quarterly	1,-1,0	-2,-1,0		x		x
62	Real, Hard data	GDP, United States	Quarterly	4,2,3	1,2,3		x	x	x
63	Nominal, Labour costs	Unit labour cost	Quarterly	4,5,6	4,5,3		x	x	x
64	Nominal, Labour costs	Compensation per employee	Quarterly	4,5,6	4,5,3		x	x	x

Notes: Columns 5 and 6 provide the number of missing observations at the end of the sample caused by the publication delays. The number of missing observations depends if the data is captured in the middle or at the end of the month and for the quarterly variables also on the month within a quarter, cf. Notes for Table 1. Negative numbers for capacity utilisation reflect the fact that the figure on the current quarter is released before the end of this quarter (in its second month). Columns under *Specification* indicate which series were included in the benchmark model (*Ben*) and model with disaggregated data (*Disagg*), respectively. Columns under *Trans* specify whether logarithm and/or differencing was applied to the initial series.

B State space representation of the model

Below are the details for the state space representation (11) as specified by the equations (5)-(10), for $p = 1$ (which was chosen in the empirical application):

$$\underbrace{\begin{pmatrix} x_t \\ y_t^Q \end{pmatrix}}_{\bar{x}_t} = \underbrace{\begin{pmatrix} \mu \\ \mu_Q \end{pmatrix}}_{\bar{\mu}} + \underbrace{\begin{pmatrix} \Lambda & 0 & 0 & 0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 \\ \Lambda_Q & 2\Lambda_Q & 3\Lambda_Q & 2\Lambda_Q & \Lambda_Q & 0 & 1 & 2 & 3 & 2 & 1 \end{pmatrix}}_{Z(\theta)} \underbrace{\begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{pmatrix}}_{\alpha_t} \tag{12}$$

$$\begin{pmatrix} f_t \\ f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ \varepsilon_t \\ \varepsilon_t^Q \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \end{pmatrix} = \underbrace{\begin{pmatrix} A_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_r & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathrm{diag}(\alpha_1, \ldots, \alpha_n) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \alpha_Q & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}}_{T(\theta)} \begin{pmatrix} f_{t-1} \\ f_{t-2} \\ f_{t-3} \\ f_{t-4} \\ f_{t-5} \\ \varepsilon_{t-1} \\ \varepsilon_{t-1}^Q \\ \varepsilon_{t-2}^Q \\ \varepsilon_{t-3}^Q \\ \varepsilon_{t-4}^Q \\ \varepsilon_{t-5}^Q \end{pmatrix} + \underbrace{\begin{pmatrix} u_t \\ 0 \\ 0 \\ 0 \\ 0 \\ e_t \\ e_t^Q \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{\eta_t}$$

where $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{n,t})'$ and $e_t = (e_{1,t}, e_{2,t}, \ldots, e_{n,t})'$.

The block specific factor structure further implies that

$$\Lambda = \begin{pmatrix} \Lambda_{G,N} & \Lambda_{N,N} & 0 \\ \Lambda_{G,R} & 0 & \Lambda_{R,R} \end{pmatrix}, \quad \Lambda_Q = (\Lambda_{Q,G} \quad 0 \quad \Lambda_{Q,R}),$$

$$f_t = \begin{pmatrix} f_t^G \\ f_t^N \\ f_t^R \end{pmatrix}, \qquad A_1 = \begin{pmatrix} A_{1,G} & 0 & 0 \\ 0 & A_{1,N} & 0 \\ 0 & 0 & A_{1,R} \end{pmatrix}, \qquad Q = \begin{pmatrix} Q_G & 0 & 0 \\ 0 & Q_N & 0 \\ 0 & 0 & Q_R \end{pmatrix}.$$

Hence, the parameters of the model are:

$$\theta = (\mu, \mu_Q, \mathrm{vec}(\Lambda_{G,N})', \mathrm{vec}(\Lambda_{N,N})', \mathrm{vec}(\Lambda_{G,R})', \mathrm{vec}(\Lambda_{R,R})', \mathrm{vec}(\Lambda_{Q,Q})', \mathrm{vec}(\Lambda_{Q,R})', \mathrm{vec}(A_{1,G})', \mathrm{vec}(A_{1,N})',$$
$$\mathrm{vec}(A_{1,R})', \mathrm{vec}(Q_G)', \mathrm{vec}(Q_N)', \mathrm{vec}(Q_R)', \alpha_1, \ldots, \alpha_n, \alpha_Q, \sigma_1, \ldots, \sigma_n, \sigma_Q)'.$$

For the sake of simplicity we presented a model with a single quarterly variable (y_t^Q is of dimension 1). However, the state space representation can be easily modified to include an arbitrary number of quarterly variables n_Q (for example, the model with disaggregated data contains 6 quarterly variables). In that case y_t^Q, μ_Q, ε_t^Q and e_t^Q will be vectors of length n_Q. Λ_Q will be a matrix of size $n_Q \times r$ and α_Q will be a $n_Q \times n_Q$ diagonal matrix. Finally, the scalars in the lines of $Z(\theta)$ and $T(\theta)$ corresponding to y_t^Q and ε_t^Q need to be replaced by $n_Q \times n_Q$ identity or zero matrices.

C EM algorithm

The parameters θ of the state space form (12) are estimated by the Expectation Maximisation (EM) algorithm. EM is a popular solution to problems, for which latent or missing data yield the direct maximisation of the likelihood function intractable or computationally difficult.[11] The basic principle behind EM is to write the likelihood in terms of observable as well as latent variables (in our case in terms of $\bar{x}_t$ and α_t, $t = 1, \dots, T_v = \max_i T_{i,v}$) and given the available data Ω_v,[12] obtain the maximum likelihood estimates in a sequence of two alternating steps:

- E-step - the expectation of the log-likelihood conditional on the data is calculated using the estimates from the previous iteration $\theta(r)$,
- M-step - the parameters are re-estimated through the maximisation of the expected log-likelihood with respect to θ.

Below we provide the details of the implementation of the EM algorithm for the state space form (12) (based on the results in Bańbura and Modugno, 2009).

We first estimate μ and μ_Q by sample means and use the de-meaned data throughout the EM steps.

To deal with missing observations in $\bar{x}_t$ we follow Bańbura and Modugno (2009) and introduce the selection matrices W_t and W_t^Q. They are diagonal matrices of size n and 1, respectively, with ones corresponding to the non-missing values in x_t and y_t^Q, respectively.

For the sake of simplicity, we first consider the case without restrictions on Λ, Λ_Q, A_1 and Q implied by block specific factors.

[11] See Dempster, Laird, and Rubin (1977) for a general EM algorithm and Shumway and Stoffer (1982) or Watson and Engle (1983) for application to state space representations.

[12] $\Omega_v \subseteq \{y_1, \dots, y_{T_v}\}$ because some observations in y_t are missing.

The maximisation of the expected likelihood (M-step) with respect to θ in the $(r+1)$-iteration would yield the following expressions:

- The matrix of loadings for the monthly variables:

$$\begin{aligned} \text{vec}\big(\Lambda(r+1)\big) &= \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(r)}\left[f_t f_t' | \Omega_v\right] \otimes W_t\right)^{-1} \\ &\quad \text{vec}\left(\sum_{t=1}^{T} W_t x_t \mathbb{E}_{\theta(r)}\left[f_t' | \Omega_v\right] + W_t \mathbb{E}_{\theta(r)}\left[\varepsilon_t f_t' | \Omega_v\right]\right) \end{aligned} \tag{13}$$

- The matrix of loadings for the quarterly variables:

  Let$f_t^{(p)} = [f_t', \ldots, f_{t-p+1}']'$ and $D = \sum_{t=1}^{T} \mathbb{E}_{\theta(r)}\big[f_t^{(5)} f_t^{(5)'} | \Omega_v\big] W_t^Q$. The unrestricted $\bar{\Lambda}_Q = (\Lambda_Q \;\; 2\Lambda_Q \;\; 3\Lambda_Q \;\; 2\Lambda_Q \;\; \Lambda_Q)$ is given by

$$\text{vec}\big(\bar{\Lambda}_Q^{ur}(r+1)\big) = D^{-1}\left(\sum_{t=1}^{T} W_t^Q y_t^Q \mathbb{E}_{\theta(r)}\big[f_t^{(5)'} | \Omega_v\big]\right)$$

  For the restricted $\bar{\Lambda}_Q$ it holds $C\bar{\Lambda}_Q = 0$ with

$$C = \begin{bmatrix} I_r & -2I_r & 0 & 0 & 0 \\ I_r & 0 & -3I_r & 0 & 0 \\ I_r & 0 & 0 & -2I_r & 0 \\ I_r & 0 & 0 & 0 & -I_r \end{bmatrix}$$

  Consequently the restricted $\bar{\Lambda}_Q$ is given by:

$$\bar{\Lambda}_Q(r+1) = \bar{\Lambda}_Q^{ur}(r+1) - D^{-1}C'(CDC')^{-1}C\bar{\Lambda}_Q^{ur}(r+1)$$

- The autoregressive coefficients in the factor VAR:

$$A_1(r+1) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(r)}\left[f_t f_{t-1}' | \Omega_v\right]\right)\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(r)}\left[f_{t-1} f_{t-1}' | \Omega_v\right]\right)^{-1} \tag{14}$$

- The covariance matrix in the factor VAR:

$$Q(r+1) = \frac{1}{T}\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(r)}\left[f_t f_t' | \Omega_v\right] - A_1(r+1)\sum_{t=1}^{T} \mathbb{E}_{\theta(r)}\left[f_{t-1} f_t' | \Omega_v\right]\right) \tag{15}$$

- The autoregressive coefficients in the AR representation for the idiosyncratic component of the monthly variables:

$$\alpha_i(r+1) = \left(\sum_{t=1}^{T} \mathbb{E}_{\theta(r)}\left[\varepsilon_{i,t}\varepsilon_{i,t-1} | \Omega_v\right]\right)\left(\sum_{t=1}^{T} \mathbb{E}_{\theta(r)}\left[\left(\varepsilon_{i,t-1}\right)^2 | \Omega_v\right]\right)^{-1} \qquad i = 1, \ldots, n, Q$$

- The variance in the AR representation for the idiosyncratic component of the monthly variables:

$$\sigma_i^2(r+1) = \frac{1}{T}\left(\sum_{t=1}^{T}\mathbb{E}_{\theta(r)}\left[\left(\varepsilon_{i,t}\right)^2|\Omega_v\right] - \alpha_i(r+1)\sum_{t=1}^{T}\mathbb{E}_{\theta(r)}\left[\varepsilon_{i,t-1}\varepsilon_{i,t}|\Omega_v\right]\right) \qquad i = 1,\ldots,n,Q\,.$$

The conditional expectations (the E-step) in the expressions above are computed using the Kalman smoother on the state space representation (12) with the previous iteration parameters $\theta(r)$. The initial parameters $\theta(0)$ are obtained on the basis of principal components analysis (in the spirit of the 2-step method of Doz, Giannone, and Reichlin, 2006b).

To account for the restrictions imposed by group specific factors, we would split the parameters in Λ into $\Lambda_N = (\Lambda_{G,N} \quad \Lambda_{N,N})$ and $\Lambda_R = (\Lambda_{G,R} \quad \Lambda_{R,R})$ and obtain the $r+1$-iteration of Λ_N by modifying formula (13) as

$$\begin{aligned}\mathrm{vec}\big(\Lambda_N(r+1)\big) \quad &= \quad \left(\sum_{t=1}^{T}\mathbb{E}_{\theta(r)}\left[f_t^{G,N} f_t^{G,N'}|\Omega_v\right]\otimes W_t^N\right)^{-1} \\ &\quad \mathrm{vec}\left(\sum_{t=1}^{T} W_t^N x_t^N \mathbb{E}_{\theta(r)}\left[f_t^{G,N'}|\Omega_v\right] + W_t^N \mathbb{E}_{\theta(r)}\left[\varepsilon_t^N f_t^{G,N'}|\Omega_v\right]\right)\end{aligned}$$

where $f_t^{G,N} = (f_t^{G'}\ f_t^{N'})'$, x_t^N and ε_t^N are the subvectors of x_t and ε_t containing only nominal variables and idiosyncratic components, respectively. W_t^N can be obtained from W_t by discarding all the rows and columns corresponding to the real data. The updating formulas for Λ_R can be obtained in an analogous fashion. To obtain restricted versions of A_1 and Q we can use the formulas (14) and (15) for each of the factors, f_t^G, f_t^N, f_t^R, separately.