Nowcasting with daily data

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

October 21, 2011

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

1 Introduction

Now-casting is defined as the prediction of the the present, the very near future and the very recent past. The term is a contraction for *now* and *forecasting* and has been used for a long-time in meteorology and recently also in economics (Giannone, Reichlin, and Small, 2008).

Now-casting is relevant in economics because key statistics on the present state of the economy are available with a significant delay. This is particularly true for those collected on a quarterly basis, with Gross Domestic Product (GDP) being a prominent example. For instance, the first official estimate of GDP in the United States (US) or in the United Kingdom is published approximately one month after the end of the reference quarter. In the euro area the corresponding publication lag is 2-3 weeks longer.

The basic principle of now-casting is to exploit the information which is published early and possibly at higher frequencies than the target variable of interest to obtain its 'early estimate' before the official figure becomes available. If the focus is on tracking GDP, one may look at its expenditure components, like for example personal consumption, which for the US is available at a monthly frequency, or variables related to the production side such as industry output. In addition, one may consider information contained in surveys or in forward looking indicators such as financial variables. The idea here is that both 'hard' information like industrial production and 'soft' information like surveys may provide an early indication of the current developments in economic activity. Surveys, being the first monthly variables to be released, are particularly valuable because of their timeliness. Financial variables, which are available at very high frequency and, in principle, carry information on expectations of future economic developments, may also be useful although there is less empirical work on this topic (on this see Andreou, Ghysels, and Kourtellos, 2008).

Starting with Giannone, Reichlin, and Small (2008) and Evans (2005), the literature has provided a formal statistical framework to embed the now-casting process. Key in this framework is to use a model with a state space representation. Such model can be written as a system with two types of equations: *measurement equations* linking observed series to a latent state process, and *transition equations* describing the state process dynamics. The latent state process is typically associated with the unobserved state of the economy or sometimes directly with the higher frequency counterpart of the target variable. State space representation allows the use of the Kalman filter to obtain an optimal projection for both, the observed and the state variables. Importantly, Kalman filter can easily cope with quintessential features of a now-casting

information set such as: missing data at the end of the sample due to the non-synchronicity of data releases ("ragged"/"jagged" edge problem), missing data in the beginning of the sample due to only a recent collection of some data sources, and the data at different frequencies.

Important feature of the framework proposed by Giannone, Reichlin, and Small (2008) is that it allows to formally interpret and comment various data releases in terms of the signal they provide on current economic conditions. This is possible because Kalman filter provides projections for all the variables in the the model and thus for each data release a model based surprise (the 'news') can be computed. Bańbura and Modugno (2010) have shown formally how to link such news to the resulting now-cast revision. This gives a transparent, model based approach to the interpretation of data releases and allows evaluating the role of different categories of data - surveys, financial, production or labor market - in signaling changes in economic activity.

Since the market as well as policy makers typically watch and comment many data, the now-cast model should ideally be able to handle efficiently a large information set. This is indeed one of the features of the econometric model proposed by Giannone, Reichlin, and Small (2008). The motivation behind a data-rich approach is not necessarily the improvement in forecasting accuracy, but rather the ability to evaluate and interpret any significant information that may affect the now-cast.

In Giannone, Reichlin, and Small (2008) the estimation procedure exploits the fact that relevant data series, although may be numerous, co-move quite strongly so that their behavior can be captured by few common factors. In other words, all the variables in the information set are assumed to be generated by a dynamic factor model which copes effectively with the so-called 'curse of dimensionality' (large number of parameters relative to the sample size). The estimation method in Giannone, Reichlin, and Small (2008) is the two-step procedure proposed by Doz, Giannone, and Reichlin (2011), which is based on principal components analysis. More recent works, as, for example Bańbura and Modugno (2010), apply a quasi maximum likelihood for which Doz, Giannone, and Reichlin (2006) have shown consistency and robustness properties when both the sample and the cross-section size are large.

The model of Giannone, Reichlin, and Small (2008) was first implemented to now-cast GDP at the Board of Governors of the Federal Reserve in a project which started in 2003. Since then various versions have been built for different economies and also implemented at other policy institutions; e.g. at the European Central Bank (Angelini, Camba-Méndez, Giannone, Reichlin, and Rünstler, 2011; Bańbura and Rünstler, 2011; Rünstler, Barhoumi, Benk, Cristadoro,

Reijer, Jakaitiene, Jelonek, Rua, Ruth, and Nieuwenhuyze, 2009), the International Monetary Fund (Matheson, 2011), central banks of Ireland (D'Agostino, McQuinn, and O'Brien, 2008), New Zealand (Matheson, 2010) or Norway (Aastveit and Trovik, 2008). Further work in this area, using the framework based on a factor model in the state space form, has been done by Camacho and Perez-Quiros (2010), Marcellino and Schumacher (2010) and Lahiri and Monokroussos (2011) for now-casting GDP in the euro area, Germany and the US, respectively. Giannone, Reichlin, and Simonelli (2009) propose a mixed frequency VAR for now-casting GDP in the euro area. A slightly different application of the framework is proposed by Aruoba, Diebold, and Scotti (2009) and Brave and Butters (2011) who construct, respectively, an index of aggregate economic activity and an index of financial conditions. Technically, the index is the estimated common factor; assumed to reflect an unobserved state of the economy or of financial conditions. Finally, Frale, Marcellino, Mazzi, and Proietti (2011) exploit the model for the construction of a monthly GDP for the euro area.

Empirical results in the literature on now-casting have pointed out several more general conclusions. First, gains of institutional and statistical forecasts of GDP relative to the naïve constant growth model are substantial only at very short horizons and in particular for the current quarter. This implies that the ability to forecast GDP growth mostly concerns the current (and previous) quarter. Second, the automatic statistical procedure performs as well as institutional forecasts which are the result of a process involving models and judgement. These two results suggest that now-casting has an important place in the broader forecasting literature. Further, it turns out that the progressive consideration of more information as data are published throughout the quarter is important since the estimates become gradually more accurate as the quarter comes to a close. Finally, the exploitation of early data releases leads to improvement in the now-cast accuracy. Surveys, for example, by providing the most timely information among monthly data, contribute to an improvement of the estimate early in the quarter. The relevance of different types of data for the now-cast therefore depends on where we are in the quarter. By the time hard information, such as industrial production, becomes available later in the quarter, the importance of surveys declines markedly. An extensive review of the literature, including empirical findings, is provided in the survey by Bańbura, Giannone, and Reichlin (2011).

Most of the previous now-casting applications combined only quarterly and monthly data. Higher frequency, e.g. daily, data such as asset prices were converted to monthly frequency. However, financial data might be valuable not only because they are forward looking but also because they are timely. This advantage can be partly lost when they are used at lower

frequency. In the present chapter we extend the framework to accommodate a general mixed frequency data set with both flow and stock variables. In the the empirical analysis of this chapter we include daily and weekly time series along the monthly and quarterly.

Our aim is twofold. First we want to establish whether high frequency information contributes to the precision of the now-cast of GDP. In the forecasting literature this problem has been studied extensively (see Stock and Watson, 2003; Forni, Hallin, Lippi, and Reichlin, 2003) but, with the exception of Andreou, Ghysels, and Kourtellos (2008), results are based on models which do not take into account the publication lags associated with different data series. The approach of Andreou, Ghysels, and Kourtellos (2008) is more similar to ours since it takes into account mixed frequencies and ragged edge data. However, since in their approach, macroeconomic variables are treated as quarterly, the contribution of financial variables to the forecast is over-emphasized by construction. Second, we want to study the interaction between financial and macro data by estimating the effect of news in macro fundamentals on stock prices. We can do this by exploiting the multivariate dynamic nature of our model which allows to produce forecasts and forecasts errors of both target variables and predictors and identify both the contribution to the GDP now-cast of the unexpected component of financial releases and that of macroeconomic news on the monthly growth rate of stock prices.

Market participants can be viewed as now-casters since they monitor macroeconomic data to get a view on current and future fundamentals. Most of the relevant information on the state of the economy is conveyed to markets through the release of macroeconomic reports. Market expectation for the headlines of these reports are collected up to the day before the actual release of the indicator and distributed by data providers (i.e. Bloomberg). When realizations are different than these expectations, that is when the news are sizeable, market's view of the world changes and this leads to changes in asset prices (for evidence on this point see (Boyd, Hu, and Jagannathan, 2005; Flannery and Protopapadakis, 2002)). Assuming that our model based news are correlated with surprises in the market, we should therefore expect macro-economic news to affect stock prices.

The chapter is organized as follows. The second section defines the problem of now-casting in general. In the third section, we review the estimation method and in the fourth we provide an empirical application. Section five concludes.

2 The problem

Before referring to a particular model, let us define formally the general problem of producing a nowcast and its updates, which arise as a result of an inflow of new information.[1]

To fix ideas we will illustrate the problem on an example of the GDP nowcast. As mentioned in the introduction, the first official estimate of GDP in the US is released around four weeks after the close of the reference quarter. In the meantime it can be estimated using higher-frequency data (daily, weekly and monthly) which 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 (v can be associated with a particular forecast update). Further let us denote GDP growth at time t as $\tilde{y}_t$. We define the problem of nowcasting of $\tilde{y}_t$ as the orthogonal linear projection of $\tilde{y}_t$ on the available information set Ω_v:

$$\mathbb{P}\Big[\tilde{y}_t|\Omega_v\Big] = \mathbb{E}\Big[\tilde{y}_t|\Omega_v\Big], \tag{1}$$

where $\mathbb{E}\Big[\,\cdot\,|\Omega_v\Big]$ refers to the 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 daily, weekly, monthly and quarterly. Hence we will have $\Omega_v = \Big\{y_{i,t}\,; i = 1, ..., n\,; t = t_1(f_i), t_2(f_i), \ldots, T_{i,v}\,; f_i \in \{D, W, M, Q\}\Big\}$ where $T_{i,v}$ corresponds to the last period for which in vintage v the series i has been observed. As the highest frequency in the data set is daily, t corresponds to daily (business) frequency and $t_1(Q), t_2(Q), \ldots$ denote the last business day of consecutive quarters in the sample (analogously $t_1(W), t_2(W), \ldots$ and $t_1(M), t_2(M), \ldots$ denote the last days of consecutive weeks and months, respectively).[2] Without a loss of generality we order the variables according to decreasing frequency and quarterly GDP is the last variable (i.e. $\tilde{y}_t = y_{n,t}, t = t_1(Q), t_2(Q), \ldots$). Because of the non-synchronicity of data releases and mixed frequency, $T_{i,v}$ is not the same across variables and therefore the data set exhibits the above mentioned jagged edge.

The problem of nowcasting needs to be analyzed in a framework which imposes a plausible probability structure on Ω_v and which can efficiently exploit all the relevant information from

[1] This and next section borrows heavily from Bańbura, Giannone, and Reichlin (2011)

[2] We adopt here the convention that quarterly (weekly/monthly) figures are assigned to the last day of a quarter (week/month).

such an information set, where, in particular, the number of potential monthly predictors, $y_{i,t}$, could be large.

One important feature of the 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 accrues. In other words we will, in general, perform a sequence of projections: $\mathbb{E}\left[\tilde{y}_t|\Omega_v\right]$, $\mathbb{E}\left[\tilde{y}_t|\Omega_{v+1}\right]$, ..., where v, $v+1$, ..., are associated with consecutive forecast updates. The intervals between two consecutive updates might short and change over time.

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

Let us first analyse the difference between the two information sets Ω_v and Ω_{v+1}. Between time $v+1$ and v we have a release of certain group of variables, $\{y_{j,T_{j,v+1}}, j \in \mathbb{J}_{v+1}\}$ and consequently the information set expands.[3] 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 $\Omega_{v+1} \setminus \Omega_v = \{y_{j,T_{j,v+1}}, j \in \mathbb{J}_{v+1}\}$.

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

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

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

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

and $I_{v+1} = (I_{v+1,1} \dots I_{v+1,J_{v+1}})'$, where J_{v+1} denotes the number of elements in $\mathbb{J}_{v+1}$. 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 the *news*. The concept of *news* is useful because what matters in understanding the updating process of the nowcast is not

[3]Note that for some variables, in particular the high frequency ones, there could be several releases between v and $v+1$. However, for the sake of simplicity of notation, we develop the formulas under the assumption that only one "additional" observation has been released. The extension to incorporate more releases of a given variable between forecast updates is straightforward.

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, we would intuitively expect that a negative *news*, for example a release of industrial production below expectations, should induce a downward revision of the GDP forecasts. 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* depends on the order in which new data are released.

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

$$\mathbb{E}\left[\tilde{y}_t|I_{v+1}\right] = \mathbb{E}\left[\tilde{y}_t I'_{v+1}\right] \mathbb{E}\Big[I_{v+1} I'_{v+1}\Big]^{-1} I_{v+1}\,. \tag{3}$$

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:

$$\underbrace{\mathbb{E}\left[\tilde{y}_t|\Omega_{v+1}\right]}_{\text{new forecast}} = \underbrace{\mathbb{E}\left[\tilde{y}_t|\Omega_v\right]}_{\text{old forecast}} + \sum_{j\in\mathbb{J}_{v+1}} b_{j,t,v+1} \underbrace{\Big(y_{j,T_{j,v+1}} - \mathbb{E}\left[y_{j,T_{j,v+1}}|\Omega_v\right]\Big)}_{\text{news}}\,. \tag{4}$$

In other words we can express the forecast revision as a weighted sum of *news* from the released variables. Hence, consistent 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. 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 therefore allowing commenting the revision of the target in relation to unexpected developments of the inputs.

I order to obtain (4) we need a forecast at any vintage for all the variables included in the information set. This requires a joint model for all the predictors.

3 The econometric framework

To compute nowcasts, *news* and their contributions to nowcast revisions all we need, in principle, is performing linear projections. In practice, we have to deal with several problems including mixed frequency, jagged edge and possibly other cases of missing data and the curse of dimensionality due to the richness of the available information which, if included, can lead to imprecise and volatile estimates.

In this paper we use the approach proposed by Giannone, Reichlin, and Small (2008) who offer a solution to these problems by modeling the monthly data as a parametric dynamic factor model cast in a state space representation. Here we extend that idea to accommodate for daily and weekly data. Once we obtain the state space representation, the Kalman filter techniques can be used to perform the projections as they automatically adapt to changing data availability. Importantly, the factor model representation allows inclusion of many variables, which is a desirable characteristic since many releases are commented in e.g. policy makers' briefings or monitored by the market.

As for estimation, we adopt the approach of Bańbura and Modugno (2010) who estimate the model by maximum likelihood. Doz, Giannone, and Reichlin (2006) have shown that the maximum likelihood approach is feasible and robust in the context of large scale factor models. It also allows us to take into account several important features of the nowcasting process as it is illustrated in the next section.

The next subsections describe the model and the estimation in detail.

3.1 Daily factor model

We start by specifying the dynamics for the daily data.

Let $y_t^D = (y_{1,t}, y_{2,t}, \ldots, y_{n_D,t})'$ denote the daily series, which have been transformed to satisfy the assumption of stationarity. We assume that y_t^D obey the following factor model representation:

$$y_t^D = \mu_D + \Lambda_D f_t + \varepsilon_t^D, \qquad \varepsilon_t^D \sim i.i.d.\ N\Big(0, \mathrm{diag}(\sigma_1^2, \cdots, \sigma_{n_D}^2)\Big) \tag{5}$$

where f_t is a $r \times 1$ vector of (unobserved) common factors and ε_t^D is a vector of idiosyncratic components. Λ_D denotes the factor loadings for the daily variables. The common factors and the idiosyncratic components are assumed to have mean zero and hence the constants $\mu_D = (\mu_1, \mu_2, \ldots, \mu_{n_D})'$ are the unconditional means. The projection of of a de-meaned series

$y_{i,t} - \mu_i$ on the factors, i.e. $\Lambda_{D,i\cdot} f_t$, is referred to as the common component of $y_{i,t}$.[4]

The factors are modelled as a VAR process of order p:

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

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

Taking explicitly into account the dynamics of the factors is particularly important in now-casting applications. The reason is that, due to publication delays, the information on the most recent periods can be scarce and exploiting the dynamics, in addition to contemporaneous relationships, can increase the precision of the factor estimates.

This model is obviously misspecified. It does not allow for serial and cross-sectional correlation among idiosyncratic components. The assumption that errors are normally distributed is particularly unrealistic at daily frequency. However, Doz, Giannone, and Reichlin (2006) have shown that, for large cross-sections, the maximum likelihood estimates of the model are robust to these forms of miss-specification.

In order to exploit information of data at different frequencies - daily, weekly, monthly and quarterly - we follow, as in our previous work, Mariano and Murasawa (2003). Roughly speaking, low frequency data are treated as high frequency series with missing observations and appropriate aggregators are derived to link the observed low frequency aggregates with the unobserved higher frequency component. In the Appendix we show the detailed derivation for mix of quarterly and daily data. The generalization to other frequencies is straightforward.

3.2 Estimation and forecasting

Let us define $y_t = (y_{1,t}, y_{2,t}, \ldots, y_{n,t})'$ and $\mu = (\mu_{1,t}, \mu_{2,t}, \ldots, \mu_{n,t})'$. The joint model specified for daily variables by the equations (5)-(6) and outlined for low frequency variables in the Appendix can be cast in a state space representation:

$$\begin{aligned} y_t &= \mu + Z(\theta)\alpha_t + \varepsilon_t\,, \qquad & \varepsilon_t &\sim\ i.i.d.\ N\Big(0, \Sigma_\varepsilon(\theta)\Big)\,, \\ \alpha_t &= T(\theta)\alpha_{t-1} + \eta_t\,, \qquad & \eta_t &\sim\ i.i.d.\ N\Big(0, \Sigma_\eta(\theta)\Big)\,, \end{aligned} \tag{7}$$

where the state vector α_t includes the common factor(s) and weekly, monthly and quarterly aggregators. All the parameters of the model are collected in θ. The details of the state space representation, and in particular how to derive the state vector α_t and the matrices, $Z(\theta)$, $T(\theta)$ and $\Sigma_\eta(\theta)$, are provided in the Appendix.

[4] $\Lambda_{D,i\cdot}$ denotes the i^{th} row of Λ_D.

In this paper, we estimate θ by maximum likelihood implemented by the Expectation Maximisation (EM) algorithm. This approach has been proposed for large data sets by Doz, Giannone, and Reichlin (2006) and extended by Bańbura and Modugno (2010) to deal with missing observations and idiosyncratic dynamics. Giannone, Reichlin, and Small (2008) used a different procedure involving two steps: first the parameters of the model are estimated using principal components as factor estimates; second, factors are re-estimated using the Kalman filter (see Doz, Giannone, and Reichlin, 2011). Roughly speaking, the maximum likelihood estimation using the EM algorithm consists in iterating the two-step approach: estimating the parameters conditional on the factor estimates from previous iteration and vice versa.

Maximum likelihood allows us to easily deal with substantial fraction of missing data and in addition, as our model is of moderate size (less than 30 variables), maximum likelihood approach should be more efficient.

Given an estimate of θ, the nowcasts as well as the estimates of the factors or of any missing observations in y_t, can be obtained from the Kalman filter or smoother. Precisely, under the assumption that the data generating process is given by (7) with θ equal to its QML estimate, the Kalman filter or smoother can be used to obtain, in an efficient and automatic manner, projection (1) for any pattern of data availability in Ω_v.[5] One way to understand how the Kalman filter and smoother deal with missing data is to imagine that they simply discard the rows in y_t and $Z(\theta)$ that correspond to the missing observations in the former vector, see e.g. Durbin and Koopman (2001).

In addition, the *news* I_{v+1} and the 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 (2010) for details. It is worth noting that for t large enough so that the Kalman filter has approached its steady state, the weights $b_{j,t,v+1}$ will not depend on a particular realisation of $\{y_{j,T_{j,v+1}}, j \in \mathbb{J}_{v+1}\}$ but only on θ and on the shape of the jagged edge in Ω_v and Ω_{v+1}.

The results presented in the next section have been obtained under the parametrization with one factor and one lag in the AR process that describes its dynamics. Given the composition of the data set, the first factor should capture the underlying developments in real activity (or the 'state' of *real* economy) and this is our principal focus in the empirical application. The parametrization of the dynamic behavior of the factor has been chosen looking at the in-sample performance of the model.

[5] Let $T_v = \max_i\{T_i \text{ s.t. } y_{i,T_i} \in \Omega_v\}$. The Kalman filter will be used in case the target period t in (1) is equal or larger than T_v. The Kalman smoother will be used otherwise.

4 Empirical application

4.1 Data

We are considering twenty-four series of which only GDP is quarterly. Among monthly data we include industrial production, labor market data, a variety of surveys but also price series, indicators of the housing market, trade and consumption statistics. The weekly series are initial jobless claims and the Bloomberg consumer comfort index while the five daily series refer to financial markets and oil prices. In general, the series we collected are marked on Bloomberg website as "Market Moving Indicators". Table 1 describes the series, the publication lag for a stylized calendar and the transformation we have adopted.

Table 1: Data

No	Name	Frequency	Publication delay (in days after reference period)	Transformation	
				log	diff
1	Real Gross Domestic Product	quarterly	28	x	x
2	Industrial Production Index	monthly	14	x	x
3	Purchasing Manager Index, Manufacturing	monthly	3		x
4	Real Disposable Personal Income	monthly	29	x	x
5	Unemployment Rate	monthly	7		x
6	Employment, Non-farm Payrolls	monthly	7	x	x
7	Personal Consumption Expenditure	monthly	29	x	x
8	Housing Starts	monthly	19	x	x
9	New Residential Sales	monthly	26	x	x
10	Manufacturers' New Orders, Durable Goods	monthly	27	x	x
11	Producer Price Index, Finished Goods	monthly	13	x	x
12	Consumer Price Index, All Urban Consumers	monthly	14	x	x
13	Imports	monthly	43	x	x
14	Exports	monthly	43	x	x
15	Philadelphia Fed Survey, General Business Conditions	monthly	-10		x
16	Retail and Food Services Sales	monthly	14	x	x
17	Conference Board Consumer Confidence	monthly	-5		x
18	Bloomberg Consumer Comfort Index	weekly	4		x
19	Initial Jobless Claims	weekly	4	x	x
20	S&P 500 Index	daily	1	x	x
21	Crude Oil, West Texas Intermediate (WTI)	daily	1	x	x
22	10-Year Treasury Constant Maturity Rate	daily	1		x
23	3-Month Treasury Bill, Secondary Market Rate	daily	1		x
24	Trade Weighted Exchange Index, Major Currencies	daily	1		x

Notes: The stylized calendar is based on the data releases in January 2011. Negative numbers for surveys mean that they are released before the reference month is over.

4.2 Nowcasting GDP

As mentioned in the introduction, previous research on now-casting GDP have shown the following empirical results. First, the root mean squared forecast error (RMSFE) of the early estimate declines as new information becomes available throughout the quarter. Similar effect can be observed for filter uncertainty. Second, survey variables, being the most timely, have a sizable impact on both the forecast and the filter uncertainty (for a survey see Bańbura,

Giannone, and Reichlin, 2011). Finally, in the case of the US, when the information set mirrors that of professional forecasters (around the middle of the quarter) the accuracy of the model based prediction is comparable to that of the mean SPF forecast.

Previous applications, while using similar framework to ours, disregarded higher frequency, i.e. daily or weekly, data. More precisely, higher frequency data such as financial indicators or initial jobless claims were included only after having been converted to monthly frequency. In this way their advantage in terms of timeliness with regard to e.g. survey data was lost. The present application evaluates the robustness of previous results once higher frequency information is included.

4.2.1 Forecast accuracy

In this section we look at the evolution of out-of-sample and in-sample measures of forecast precision as the information on the quarter of interest accrues.

The out-of-sample measure is the Root Mean Squared Forecast Error (RMSFE) from a simulated pseudo real time forecasting exercise over the period 1995-2010, see Figure 1a. For each quarter in the evaluation sample, we update the estimates three times per month. These dates correspond to the publication of the employment situation report, the release of industrial production and that of consumer confidence (according to our stylized calendar). For each reference quarter the first estimate is obtained in the first month of the preceding quarter while the last now-cast is based on the data available in the first month of the following quarter (few days before the official estimate is released). Depending on when in the quarter we perform the update, the availability of information differs; this is why we examine the average accuracy for each of them separately. On the horizontal axis we indicate the day, the month and the quarter of the update ('Q_1, Q_0, $Q+1$ refer to the 'preceding', 'current' and 'following' quarter, respectively). 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 according to the stylized calendar. Estimating the model recursively takes into account estimation uncertainty. The dot indicates the RMSFE of the survey of professional forecasters (SPF) which is conducted in the first half of the second month of each quarter.

Figure 1b reports the in-sample 'filter' uncertainty, i.e. uncertainty underlying the common component that is related to signal (or factor) extraction (see Giannone, Reichlin, and Small, 2008). This measure is evaluated using the Kalman filtering techniques based on the parameters estimated over the entire sample. It is computed for each business day from the beginning of

the preceding quarter through the first month of the following quarter. The horizontal axis is labeled by the dates corresponding to updates for 2008Q4. As in the case of forecast revisions, which can be expressed as weighted sums of news from particular releases (cf. Section 2), Kalman smoother output allows to decompose the declines in uncertainty into contributions from particular (groups of) variables.

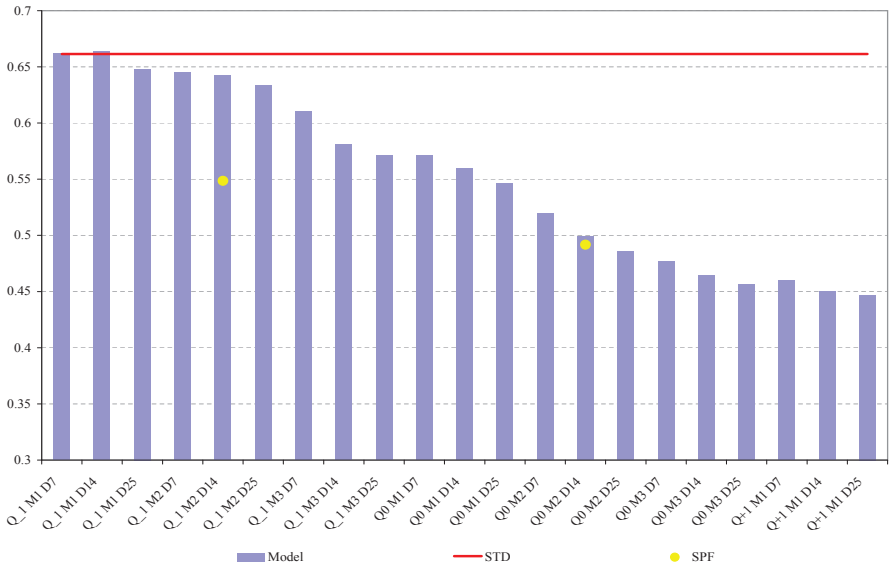
As found in earlier work (see for example Giannone, Reichlin, and Small, 2008), Figure 1a confirms that, as the information accumulates, the gains in forecast accuracy are substantial. In the next subsection, we show this point formally via a statistical test. Figure 1b also indicates increasing precision as more information becomes available.

Regarding the contribution of various groups of variables to the decline in uncertainty - results indicate that macroeconomic monthly releases have the largest effects. The big spikes are on the seventh of each month, corresponding to the release of the (monthly) employment situation report. Smaller spikes are on the fourteenth day of the month when industrial production is released. In contrast, daily or weekly data do not have much of an impact.

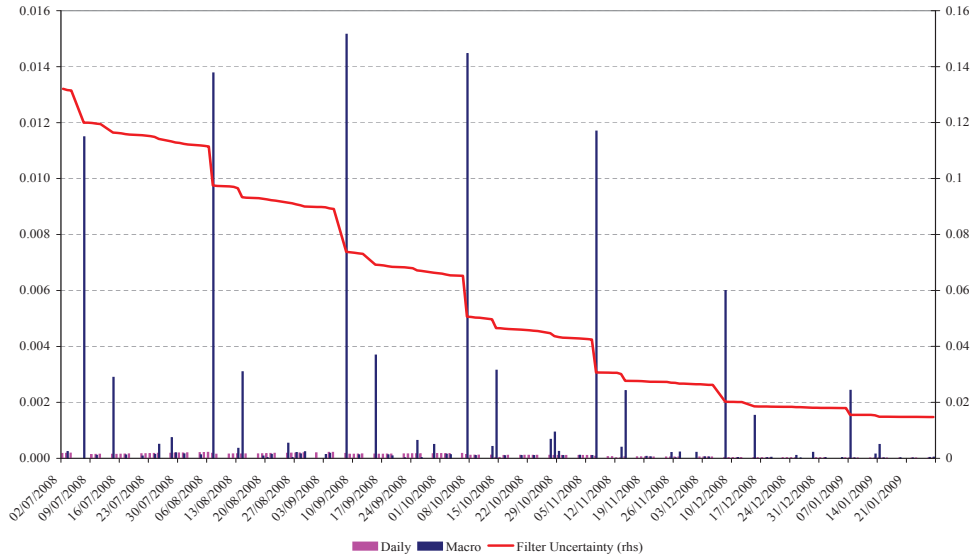
The fact that financial variables are not overly helpful in forecasting has already been highlighted by (see Stock and Watson, 2003; Forni, Hallin, Lippi, and Reichlin, 2003) for longer horizon forecasts. Our findings confirm these results in the case of now-casting with high frequency data - in the framework that allows to take into account the advantage of financial data in terms of timeliness. A different conclusion has been put forward by Andreou, Ghysels, and Kourtellos (2008) who find financial variables important for improving the now-cast accuracy of GDP. The difference might stem from the fact that Andreou, Ghysels, and Kourtellos (2008) convert all the macro variables to quarterly frequency and apply the same publication delays as for GDP. In that way financial variables might convey the information that would otherwise be already available from e.g. surveys.

Figure 1: Forecast accuracy, GDP

(a) Root Mean Squared Forecast Error (RMSFE)



(b) Filter uncertainty



Notes: The upper panel shows the Root Mean Squared Forecast Error (RMSFE) for our model over the sample period 1995 - 2010. The forecast accuracy of the model is evaluated three times per month (when the employment situation report becomes available, when industrial production is released and when the consumer confidence is released) for seven consecutive months (from the first month of the preceding quarter, 'Q_1 M_1' to the first month of the following quarter, '$Q+1$ M_1'). The dot corresponds to the RMSFE for the survey professional forecasters (SPF). The lower panel reports the evolution of filter uncertainty for the GDP now-cast for 2008Q4 corresponding to the updates from July 2008 through January 2009. The variables are grouped into 'Daily' (all the data available at daily frequency) and 'Macro' (remaining data).

4.2.2 Does information helps improving forecasting accuracy? Monotonicity tests

Figures 1a and 1b have shown heuristically that both out-of-sample and in-sample uncertainty decrease as more information becomes available. A natural way to formally test the decline in uncertainty as more data arrive is to apply the tests for forecast rationality proposed by Patton and Timmermann (2011) and based on the multivariate inequality tests in regression models of Wolak (1987, 1989). We rely on the first three tests of Patton and Timmermann (2011).[6]

Test 1: monotonicity of the forecast errors

Let us define $e_{t|\Omega_v} = \tilde{y}_t - \mathbb{E}\Big[\tilde{y}_t|\Omega_v\Big]$ as the forecast error obtained on the basis of the information set corresponding to the data vintage Ω_v and by $e_{t|\Omega_{v+1}}$ that obtained on the basis of a larger more recent vintage $v+1$ and $v = 1, \cdots, V$.

The Mean Squared Errors (MSE) differential is $\Delta_v^e = \mathbb{E}\Big[e^2_{t|\Omega_v}\Big] - \mathbb{E}\Big[e^2_{t|\Omega_{v+1}}\Big]$.

The test is defined as follows:

$$H_0 : \boldsymbol{\Delta}^e \geq 0 \quad vs \quad H_1 : \boldsymbol{\Delta}^e \ngeq 0\,,$$

where the $(V-1)\times 1$ vector of MSE-differentials is given by $\boldsymbol{\Delta}^e \equiv (\Delta^e_{V-1}, \Delta^e_{V-2}, ..., \Delta^e_1)'$.

Test 2: monotonicity of the mean squared forecast

Define the mean squared forecast (MSF) for a given vintage as $\mathbb{E}\Big[\tilde{y}^2_{t|\Omega_v}\Big] = \mathbb{E}\Big[\mathbb{E}\big[\tilde{y}^2_t|\Omega_v\big]\Big]$ and consider the difference $\Delta_v^f = \mathbb{E}\Big[\tilde{y}^2_{t|\Omega_v}\Big] - \mathbb{E}\Big[\tilde{y}^2_{t|\Omega_{v+1}}\Big]$ and its associated vector $\boldsymbol{\Delta}^f$.

The test is:

$$H_0 : \boldsymbol{\Delta}^f \leq 0 \quad vs \quad H_1 : \boldsymbol{\Delta}^f \nleq 0\,.$$

The idea behind this test is that the variance of each observation can be decomposed as follows:

$$V\Big(\tilde{y}_t\Big) = V\Big(\tilde{y}_{t|\Omega_v}\Big) + \mathbb{E}\Big[e^2_{t|\Omega_v}\Big]\,,$$

given that $\mathbb{E}\Big[\tilde{y}_{t|\Omega_v}\Big] = \mathbb{E}\Big[\tilde{y}_t\Big]$. Then a weakly increasing pattern in MSE (with increasing forecast horizon) directly implies a weakly decreasing pattern in the variance of the forecasts, i.e. $\boldsymbol{\Delta}^f \leq 0$.

[6] We thank Allan Timmermann for suggesting these tests in our context.

Test 3: monotonicity of covariance between the forecast and the target variable

Here we consider the covariance between the forecast and the target variable for different vintages v and the difference: $\Delta_v^c = \mathbb{E}\left[\tilde{y}_{t|\Omega_v}\tilde{y}_t\right] - \mathbb{E}\left[\tilde{y}_{t|\Omega_{v+1}}\tilde{y}_t\right]$. The associated vector is defined as $\mathbf{\Delta}^c$ and the test is:

$$H_0 : \mathbf{\Delta}^c \leq 0 \quad vs \quad H_1 : \mathbf{\Delta}^c \not\leq 0\,.$$

This test is closely related with previous one. Indeed the covariance between the target variable and the forecast can be written as:

$$Cov\left[\tilde{y}_{t|\Omega_v}, \tilde{y}_t\right] = Cov\left[\tilde{y}_{t|\Omega_v}, \tilde{y}_{t|\Omega_v} + e_{t|\Omega_v}\right] = V\left(\tilde{y}_{t|\Omega_v}\right)$$

Consequently a weakly increasing pattern in the variance of the forecasts implies a weakly increasing pattern in the covariances between the forecast and the target variable.

Results for the three tests are reported in Table 2. Monotonicity cannot be rejected in any of the three cases confirming the visual evidence of Figures 1a and 1b.

Table 2: *Monotonicity tests*

	$\Delta^e \geq 0$	$\Delta^f \leq 0$	$\Delta^c \leq 0$
p-value	0.49	0.50	0.50

Notes: Table reports the p-values of three of monotonicity tests for, respectively, the forecast errors, the mean squared forecast and covariance between the forecast and the target variable.

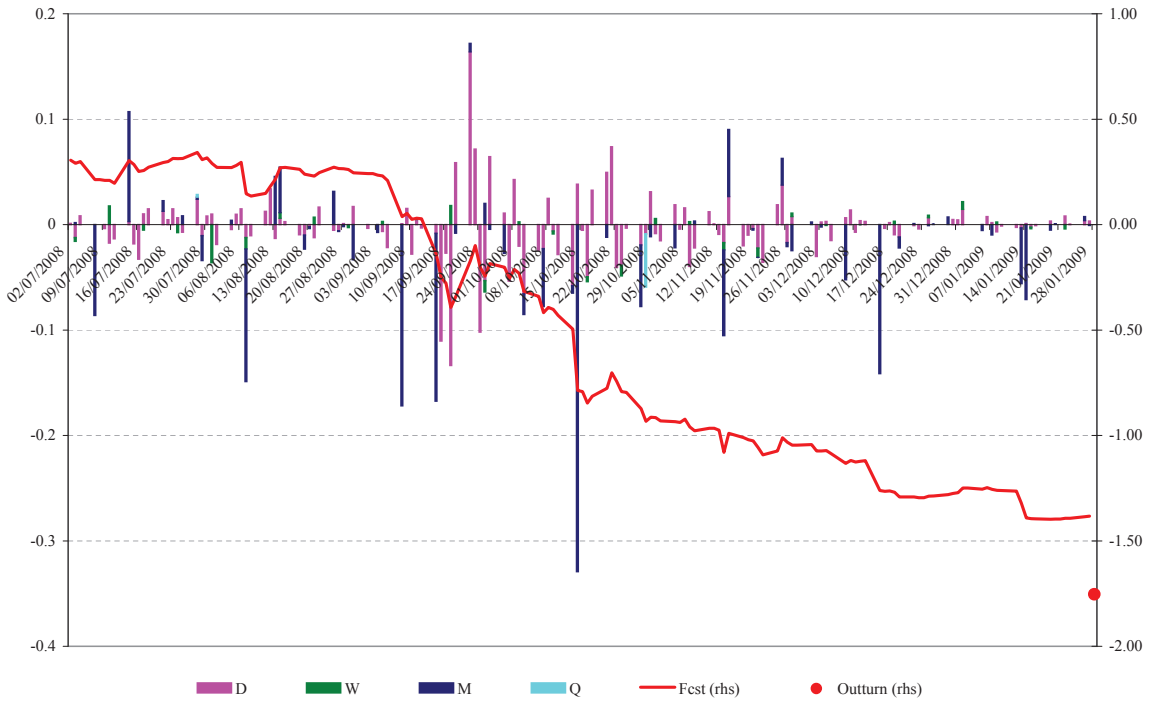
4.2.3 News

As an illustration of how the sequence of 'news' impacts the estimate of GDP, we show the forecast updates for the GDP growth rate in the fourth quarter of 2008 since the beginning of the third quarter of 2008 until the end of January 2009, when the first official estimate was released. This is an interesting episode since it corresponds to the onset of the financial crisis following the bankruptcy of Lehman Brothers. Specifically, we produce a first forecast with data available on first of July 2008 and we subsequently update it every day, each time incorporating new data releases. We use the same stylized calendar as that used to construct Figures 1a and 1b.

Figure 2 reports the evolution of the forecast and the contribution of the news component of

the various data groups to the forecast revision.[7] 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.

Figure 2: Contribution of news to forecast revisions



Notes: Figure shows how various data releases contribute to the GDP now-cast revisions for 2008Q4. As in Figure 1b horizontal axis provides the updates days. The data are grouped according to frequencies: daily ('D'), weekly ('W'), monthly ('M') and quarterly ('Q').

At the beginning of the forecasting period the forecast remains rather flat, corroborating the above mentioned difficulties in forecasting beyond the current quarter. The employment report in the beginning of September brings down the now-cast which then becomes negative with the release of industrial production for August (published mid September). Industrial production for September has the largest impact and leads to a substantial downward revision. This negative *news* in October is confirmed by subsequent data, both surveys and hard data. In

[7] In this exercise we abstract from the effect of parameter re-estimation. The parameters are estimated over the entire sample and kept constant for all the subsequent forecast updates.

fact, with all subsequent releases the tendency is for negative revisions.

The contribution from surveys is rather sizable at the beginning of the quarter but the larger role comes from the employment report.

The news from the weekly variables are not so prominent. Our conjecture for this finding is that initial jobless claims are rather noisy. This is line with the view of the NBER dating committee which does not use this series to determined the business cycle chronology (http://www.nber.org/cycles).

Finally, the behavior of the daily financial variables is different than that of weekly variables. The impact of financial news is sizable but also volatile and leads to revisions in different directions. A deeper analysis of the relation between financial daily information, with the focus on stock prices, and GDP and macro-economic releases is provided in Section 4.4.

4.3 A daily index of the state of the economy

To understand the working of the model, it is interesting to plot the estimated daily factor. Common factors extracted from a set of high frequency variables have become a popular tool to monitor business cycles conditions (see e.g. Aruoba, Diebold, and Scotti, 2009).[8] Our daily factor should be interpreted as a daily index of the underlying state of the economy, or rather its day-to-day change, which is to be distinguished from daily or intra-daily update of the estimate of quarterly GDP growth for the current quarter (the GDP now-cast).

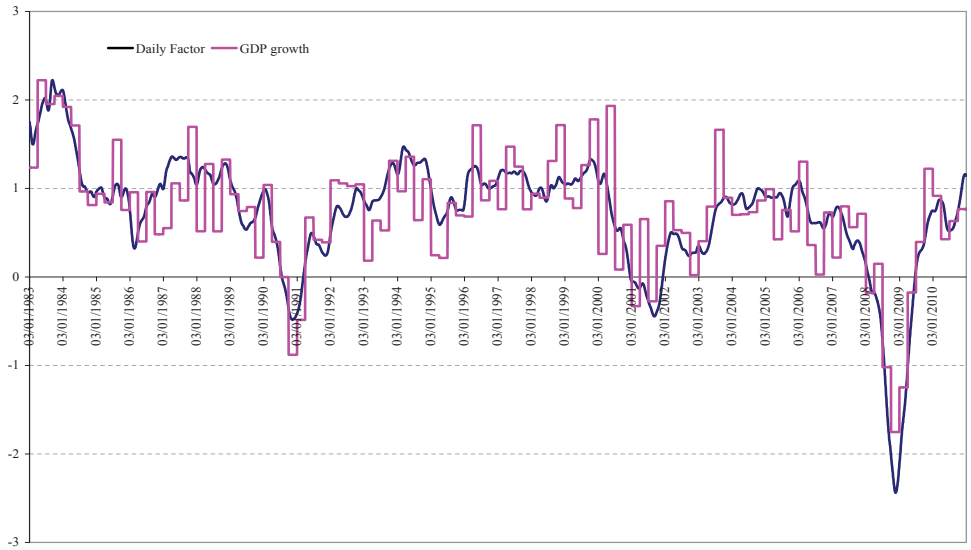
Figure 3a plots this daily index against GDP growth and shows that it tracks GDP quite well.

By appropriate filtering, this index can be aggregated to reflect quarter-on-quarter growth rate and we can then consider the projection of GDP on this quarterly aggregate (Figure 3b). This is the common component of GDP growth and captures that part of GDP dynamics which co-moves with the series included in the model (monthly, weekly and daily) while disregarding its idiosyncratic movements. The projection captures a large share of GDP dynamics suggesting that the common component, although it disregards the idiosyncratic residual, captures the bulk of GDP fluctuations.

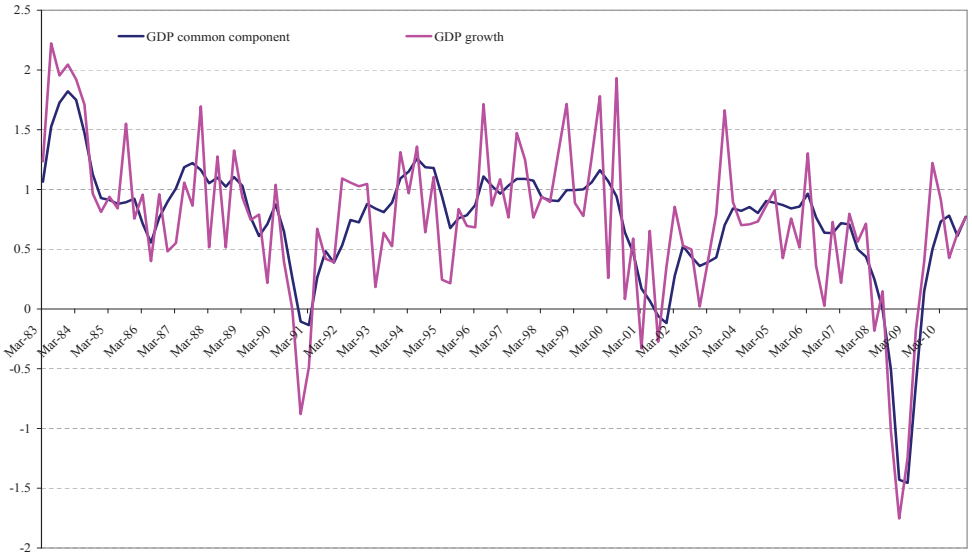
[8]The Philadelphia FED regularly publishes a daily index obtained by applying this framework to extract a daily common factor from the following indicators: weekly initial jobless claims; monthly payroll employment, industrial production, personal income less transfer payments, manufacturing and trade sales; and quarterly real GDP, see http://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/.

Figure 3: Daily factor, GDP and its common component

(a) Quarterly GDP growth and the daily factor



(b) Common component of GDP



Notes: The upper panel shows daily factor against quarterly GDP growth. The lower panel shows the GDP growth against its projection on the quarterly aggregation of the daily factor, i.e. against its common component.

4.4 Stock prices

Figure 4 reports the common component of the S&P 500 index for daily, monthly, quarterly and annual growth rates. We interpret the common component as the 'signal' in the stock prices as this is the part that is correlated with the underlying developments in the real economy.

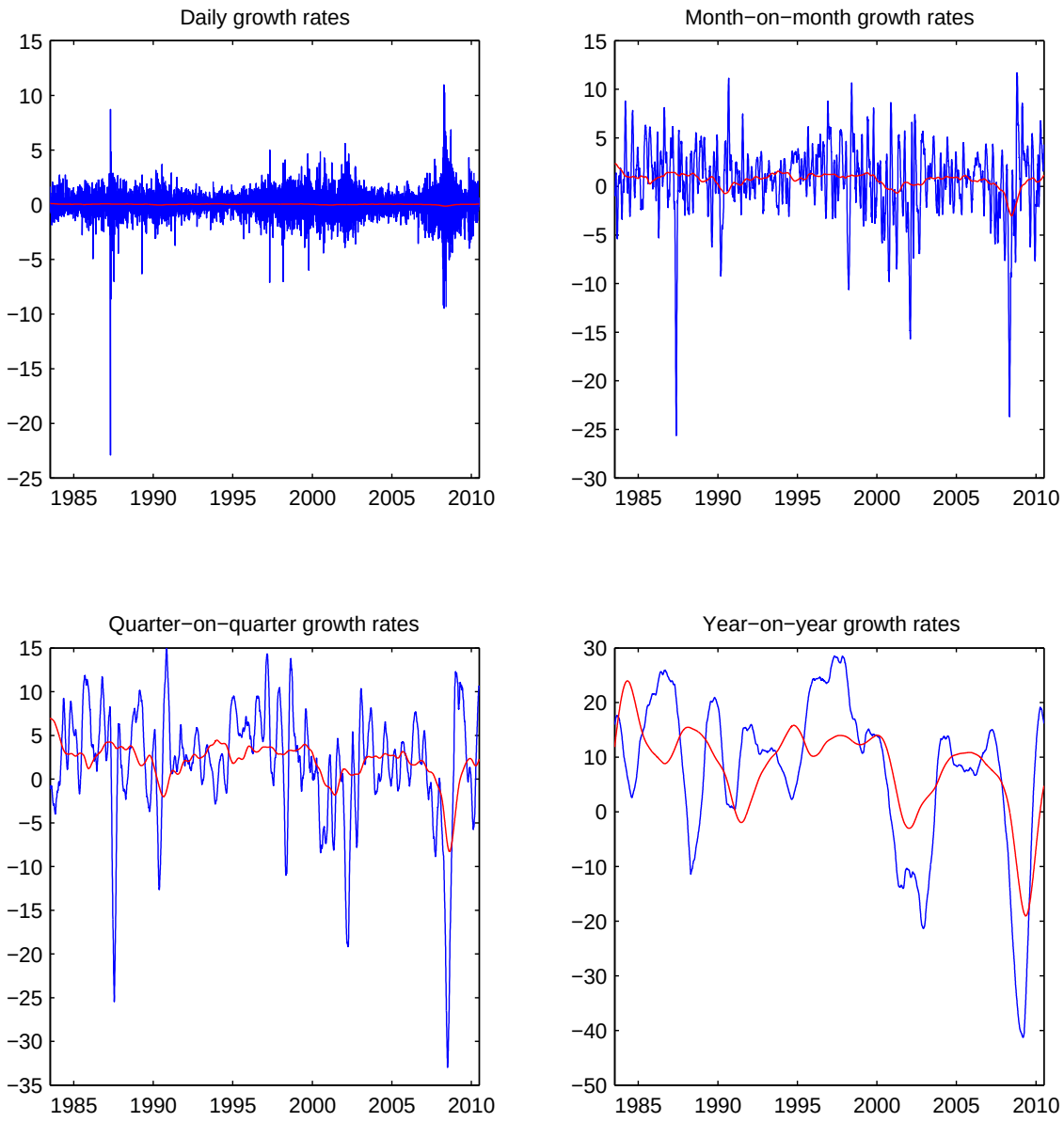
The different level of time aggregation is meant to show that, although the degree of commonality of the S&P 500 with the rest of the panel is less pronounced than seen for GDP (Figure 3b), it is more marked when we consider longer term fluctuations, annual in particular. Clearly, although stock prices provide some false signals, they do go down in recessions.

To further investigate this point we construct, from the estimated parameters of the model, the ratio of the spectral density of the common component and of the series itself. Results, reported in Figure 5, show that the bulk of commonality is at very low frequencies, i.e. the ratio of the variance of the common component relative to total is high at low frequencies. For a typical business cycle periodicity of eight years ($\omega = 0.003$) we have a quite sizable 'commonality', with the ratio of around 30%. This shows that low frequency components of stock prices are indeed related to macroeconomic fluctuations. Notice, however, that already for cycles with a yearly periodicity, corresponding to frequency $\omega = 0.0241$, the ratio is below 2%.

Figure 6 shows the evolution of in-sample filter uncertainty for the the monthly growth rate of S&P 500 index for a particular month (we use December 2008 as an example). Let us recall that this is the uncertainty related to the estimation of the common component. We show how the filter uncertainty declines along with the associated contributions of various groups of series, starting six months before the reference month, in this case in July 2008, up to the end of December when the target figure would be known.

The figure shows that there is a large drop in the uncertainty corresponding to macroeconomic data releases. This indicates that accounting for macroeconomic news helps forecasting the evolution of stock prices itself. This effect is particularly strong in case of the employment report. These drops in uncertainty become more and more pronounced as we shorten the forecast horizon, i.e. as we approach the reference month. However, the improvements are also substantial for the forecast at longer horizons. Our result confirms the evidence from event studies based on market measures of macroeconomic news (Boyd, Hu, and Jagannathan, 2005; Flannery and Protopapadakis, 2002). It is worth stressing, however, that we measure only the predictability of the common component. As we have seen in Figures 4 and 5, the latter

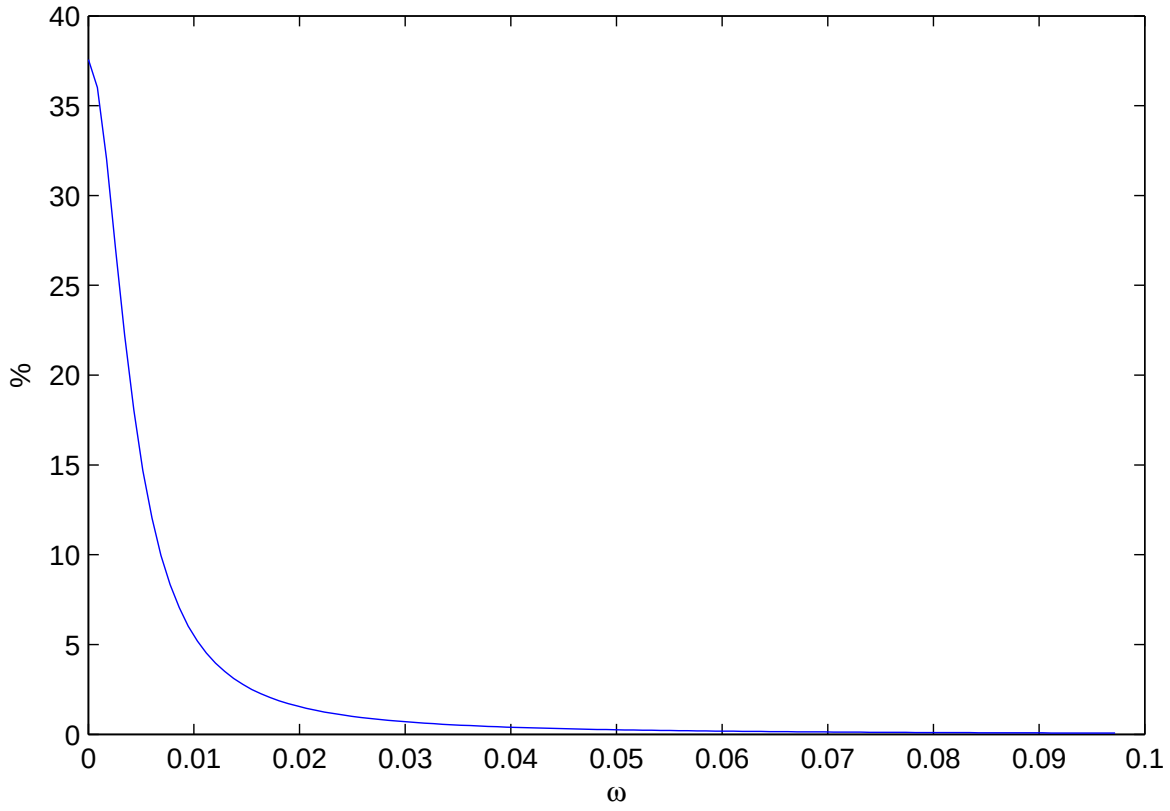
Figure 4: S&P 500 and its common component at different levels of time aggregation



Notes: Figure compares the S&P 500 and its common component for different transformations: daily growth rate (upper left panel), month-on-month growth rate (upper right panel), quarter-on-quarter growth rate (lower left panel) and year-on-year growth rate (lower right panel).

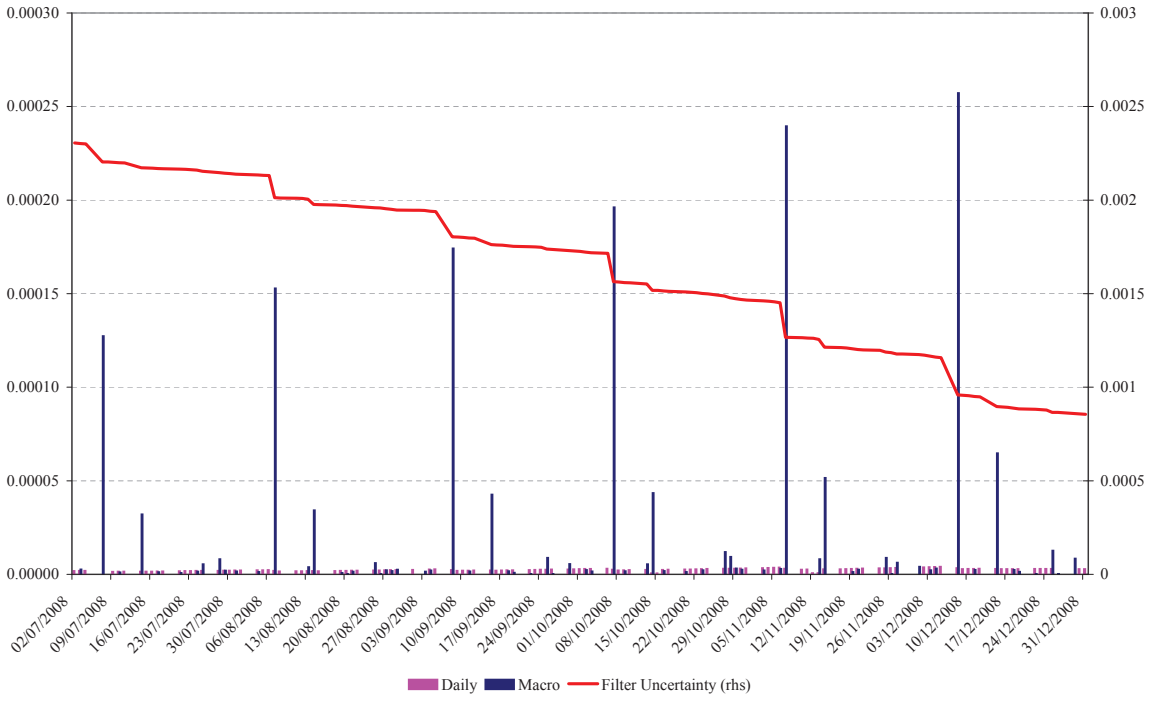
accounts only for a small percentage of monthly fluctuations of the stock prices. In addition the exercise considers only filter uncertainty, and hence it does not account for parameter uncertainty. As a consequence, it is not clear whether the predictability of stock returns based on macroeconomic data releases can be exploited in practice.

Figure 5: Spectral density ratio for S&P 500 and its common component (year- on-year growth rates)



Notes: Figure shows the spectral density ratio between the common component of the S&P500 and the series itself.

Figure 6: Filter uncertainty, S&P 500



5 Conclusions

This paper has reviewed the basic concepts and problems related to now-casting. Now-casting has been defined as forecasting at short (possible negative) horizons in the presence of real-time data flow. The discussion has been focused on now-casting quarterly GDP growth, but the methodology could be easily applied for other low frequency variables for which higher frequency information is available.

We have recalled the framework developed by Giannone, Reichlin, and Small (2008) along with subsequent extensions, which allows to produce now-casts in an automatic and efficient manner as well as to interpret data releases in terms of their implication for the now-cast revisions.

In the empirical application we extended on previous studies along the following lines. First, we considered daily and weekly data along monthly and quarterly ones. Second, we have confirmed previous heuristic results that accruing information improves now-cast precision by means of formal tests. Third, we established that daily financial variables do not help improving the precision of the GDP now-cast once the timeliness of monthly macro data is properly accounted for.

As a by-product of our analysis we have constructed a daily index of economic activity and considered the projection of both GDP and the S&P 500 index of stock prices on this index. Results show that while such projection explains the bulk of GDP dynamics, it accounts for much less of daily fluctuations in stock prices. On the other hand, the index explains low frequency movements of stock prices indicating that financial variables are linked to macroeconomic fundamentals. This conclusion is reinforced by the result that macroeconomic 'news' contribute to lowering filter uncertainty around the 'signal' in the stock price index.

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.

Andreou, 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, L. Reichlin, and Rünstler (2011): "Short-term forecasts of euro area GDP growth," *Econometrics Journal*, 14(1), C25–C44.

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 (2011): "Nowcasting," in *Oxford Handbook on Economic Forecasting*, ed. by M. P. Clements, and D. F. Hendry, pp. 63–90. Oxford University Press.

Bańbura, M., and M. Modugno (2010): "Maximum likelihood estimation of large factor model on datasets with arbitrary pattern of missing data.," Working Paper Series 1189, European Central Bank.

Bańbura, M., and G. Rünstler (2011): "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*, 27(2), 333–346.

Boyd, J. H., J. Hu, and R. Jagannathan (2005): "The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks," *Journal of Finance*, 60(2), 649–672.

Brave, S., and R. A. Butters (2011): "Monitoring financial stability: a financial conditions index approach," *Economic Perspectives*, (Q I), 22–43.

Camacho, M., and G. Perez-Quiros (2010): "Introducing the EURO-STING: Short Term INdicator of Euro Area Growth," *Journal of Applied Econometrics*, 25(4), 663–694.

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).

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

——— (2011): "A two-step estimator for large approximate dynamic factor models based on Kalman filtering," *Journal of Econometrics*, 164(1), 188–205.

Durbin, J., and S. J. Koopman (2001): *Time Series Analysis by State Space Methods*. Oxford University Press.

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

Flannery, J. M., and A. A. Protopapadakis (2002): "Macroeconomic Factors Do Influence Aggregated Stock Returns," *The Review of Financial Studies*, 15(3), 751–782.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2003): "Do financial variables help forecasting inflation and real activity in the euro area?," *Journal of Monetary Economics*, 50(6), 1243–1255.

Frale, C., M. Marcellino, G. L. Mazzi, and T. Proietti (2011): "EUROMIND: a monthly indicator of the euro area economic conditions," *Journal Of The Royal Statistical Society Series A*, 174(2), 439–470.

Giannone, D., L. Reichlin, and S. Simonelli (2009): "Nowcasting Euro Area Economic Activity in Real Time: The Role of Confidence Indicators," *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.

Harvey, A. (1989): *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

Lahiri, K., and G. Monokroussos (2011): "Nowcasting US GDP: The role of ISM Business Surveys," SUNY at Albany Discussion Papers 11-01, University at Albany, SUNY.

Marcellino, M., and C. Schumacher (2010): "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP," *Oxford Bulletin of Economics and Statistics*, 72(4), 518–550.

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. (2011): "New Indicators for Tracking Growth in Real Time," IMF Working Papers 11/43, International Monetary Fund.

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.

Modugno, M. (2011): "Maximum likelihood estimation of large factor model on datasets with arbitrary pattern of missing data.," Working Paper Series 1324, European Central Bank.

Patton, J. A., and A. Timmermann (2011): "Forecast Rationality Tests Based on Multi-Horizon Bounds," CEPR Discussion Papers 8194, C.E.P.R. Discussion Papers.

Rünstler, G., K. Barhoumi, S. Benk, R. Cristadoro, A. D. Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. V. Nieuwenhuyze (2009): "Short-term forecasting of GDP using large data sets: a pseudo real-time forecast evaluation exercise," *Journal of Forecasting*, 28, 595–611.

Stock, J. H., and M. W. Watson (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41(3), 788–829.

Wolak, F. A. (1987): "An Exact Test for Multiple Inequality and EqualityConstraints in the Linear Regression Model," *Journal of the American Statistical Association*, 82, 782–793.

——— (1989): "Testing Inequality Constraints in Linear Econometric Models," *Journal of Econometrics*, 31, 205–235.

Appendix: state space representation for mixed frequency data set

We first derive the representation for a mix of daily and of low frequency *flow* variables. To fix ideas we refer to the low frequency as "quarterly", however the derivation holds for any frequency lower than daily (with the appropriate adjustment of number of days per period).

Let Y_t^Q denote the vector of (log of) the quarterly flow series. Following Mariano and Murasawa (2003) we assume that Y_t^Q is the sum of daily contributions X_t (for the moment we assume that each quarter has the same number of days: k):

$$Y_t^Q = \sum_{s=t-k+1}^{t} X_s \,, \qquad t = k, 2k, \dots \,.$$

Hence we will have that the stationary series $y_t^Q = Y_t^Q - Y_{t-k}^Q$ can be written as:

$$y_t^Q = k \left(\sum_{s=t-k+1}^{t} \frac{t+1-s}{k} x_s + \sum_{s=t-2*(k-1)}^{t-k} \frac{s-t+2*k-1}{k} x_s \right) , \qquad t = k, 2k, \dots ,$$

where $x_s = X_s - X_{s-1}$ can be thought of as an unobserved daily growth rate (or difference).

To deal with different number of days per month or quarter, we make an approximation that

$$Y_t^Q = \frac{k}{k_t} \sum_{s=t-k_t+1}^{t} X_s \,, \qquad t = t_1(Q), t_2(Q), \dots ,$$

where k_t is the number of business days in the quarter ending on day t and k is the average number of business days per quarter over the sample. This can be justified by the fact that data are typically working day adjusted. Consequently, $y_t^Q = Y_t^Q - Y_{t-k}^Q$ becomes[9]

$$y_t^Q = k \left(\sum_{s=t-k_t+1}^{t} \frac{t+1-s}{k_t} x_s + \sum_{s=t-k_t-k_{t-k_t}+2}^{t-k_t} \frac{s-t+k_t+k_{t-k_t}-1}{k_{t-k_t}} x_s \right) , \quad t = t_1(Q), t_2(Q), \dots . \tag{8}$$

Formula (8) provides a link between the unobserved daily growth rates x_s and the observed quarterly aggregates y_t^Q. We assume that the former follow the same factor model as the daily variables:

$$x_s = \Lambda_Q f_s + \varepsilon_s \,, \qquad s = 1, 2, \dots \,. \tag{9}$$

[9] k_{t-k_t} refers to the number of days in the quarter preceding the one that ends on day t (the preceding quarter ends on day $t - k_t$).

To combine the representation for the daily variables given by (5)-(6) with relationships for the quarterly series expressed by (8) and (9), while keeping the size of the state vector modest, we will use the idea of aggregator, see Harvey (1989).

In what follows y_t^Q will be understood as vector of daily variables with missing observations for $t \neq t_1(Q), t_2(Q), \ldots$. For $p = 1$ the measurement equation for $(y_t^{D'}\ y_t^{Q'})'$ can be written as

$$\begin{pmatrix} y_t^D \\ y_t^Q \end{pmatrix} = \begin{pmatrix} \Lambda_D & 0 & 0 \\ 0 & \Lambda_Q & 0 \end{pmatrix} \begin{pmatrix} f_t \\ f_t^{QA} \\ f_t^{QP} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^D \\ \varepsilon_t^Q \end{pmatrix},$$

where f_t^{QA} aggregates the daily changes in the economy and for $t = t_1(Q), t_2(Q), \ldots$ it equals:

$$\begin{aligned} f_t^{QA} &= \sum_{s=t-k_t+1}^{t} k\frac{t+1-s}{k_t} f_s + \sum_{s=t-k_t-k_{t-k_t}+2}^{t-k_t} k\frac{s-t+k_t+k_{t-k_t}-1}{k_{t-k_t}} f_s = \\ &= \sum_{s=t-k_t+1}^{t} W_s^C f_s + \sum_{s=t-k_t-k_{t-k_t}+2}^{t-k_t} W_s^P f_s\,, \end{aligned} \tag{10}$$

where W_s^C and W_s^P denote the weights for the current and previous quarter daily increases respectively. This aggregation can be implemented with the "previous" quarter aggregator f_t^{QP} in the following manner:

1. When t corresponds to the first day of a quarter ($t = t_1(Q) + 1, t_2(Q) + 1, \ldots$) we have

$$\begin{aligned} f_t^{QA} &= f_{t-1}^{QP} + W_t^C f_t\,, \\ f_t^{QP} &= 0\,; \end{aligned}$$

2. When t corresponds to another day

$$\begin{aligned} f_t^{QA} &= f_{t-1}^{QA} + W_t^C f_t\,, \\ f_t^{QP} &= f_{t-1}^{QP} + W_t^P f_t\,. \end{aligned}$$

This could be implemented with the following time varying transition equation:

$$\begin{pmatrix} I_r & 0 & 0 \\ -W_t^C & I_r & 0 \\ 0 & 0 & I_r \end{pmatrix} \begin{pmatrix} f_t \\ f_t^{QA} \\ f_t^{QP} \end{pmatrix} = \begin{pmatrix} A_1 & 0 & 0 \\ 0 & 0 & I_r \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-1}^{QA} \\ f_{t-1}^{QP} \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \\ 0 \end{pmatrix}, \quad t = t_1(Q) + 1, t_2(Q) + 1, \ldots,$$

$$\begin{pmatrix} I_r & 0 & 0 \\ -W_t^C & I_r & 0 \\ -W_t^P & 0 & I_r \end{pmatrix} \begin{pmatrix} f_t \\ f_t^{QA} \\ f_t^{QP} \end{pmatrix} = \begin{pmatrix} A_1 & 0 & 0 \\ 0 & I_r & 0 \\ 0 & 0 & I_r \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_{t-1}^{QA} \\ f_{t-1}^{QP} \end{pmatrix} + \begin{pmatrix} u_t \\ 0 \\ 0 \end{pmatrix}, \quad \text{otherwise}\,,$$

where I_r denotes an $r \times r$ identity matrix. Adding flow variables of other frequencies to the observation vector requires augmenting the state vector by two aggregators per frequency (with the appropriate adaptation of the weights and of the range of summation in (10)).

Turning now to the *stock* variables, we have $Y_t^Q = X_t\,,\ t = t_1(Q), t_2(Q), \ldots$ and hence

$$y_t^Q = Y_t^Q - Y_{t-k_t}^Q = \Sigma_{s=t-k_t+1}^{t} x_s\,, \qquad t = t_1(Q), t_2(Q), \ldots\,.$$

Therefore only one aggregator variable:

$$\begin{aligned} f_t^{QA} &= 0\,, && t = t_1(Q), t_2(Q), \ldots\,, \\ f_t^{QA} &= f_{t-1}^{QA} + f_t\,, && \text{otherwise} \end{aligned}$$

(per frequency) is necessary, see Modugno (2011) for more details.

The resulting state vector will be of the size $r * \big(p + \sum_f (2 \cdot n_{fF} + n_{fS})\big)$, where n_{fF} and n_{fS} refer to the number of flow and stock variables, respectively, per frequency f.

Note that analogously to the common component, the quarterly idiosyncratic error ε_t^Q is a moving average of the daily ε. However in the estimation stage we will model it as a quarterly white noise.