CHAPTER 4

Now-Casting and the Real-Time Data Flow

Marta Bańbura[*,††], Domenico Giannone[†], Michele Modugno[‡] and Lucrezia Reichlin[**]

[*]European Central Bank, Kaiserstrasse 29 60311 Frankfurt am Main, Germany
[†]Universite' libre de Bruxelles, ECARES, Av. Roosevelt 50, cp 114 1050 Bruxelles, Belgium
[‡]Universite' libre de Bruxelles, ECARES, Av. Roosevelt 50, cp 114 1050 Bruxelles, Belgium
[**]London Business School, Regent's Park, London NW1 4SA, UK

Contents

Abstract

The term now-casting is a contraction for now and forecasting and has been used for a long time in meteorology and recently also in economics. In this chapter we survey recent developments in economic now-casting with special focus on those models that formalize key features of how market participants and policymakers read macroeconomic data releases in real-time, which involves

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

Handbook of Economic Forecasting, Volume 2A
ISSN 1574-0706, http://dx.doi.org/10.1016/B978-0-444-53683-9.00004-9

monitoring many data, forming expectations about them and revising the assessment on the state of the economy whenever realizations diverge sizeably from those expectations.

Keywords

Macroeconomic news, Macroeconomic forecasting, High-dimensional data, Real-time data, Mixed frequency, Dynamic factor model, State-space models

1. INTRODUCTION

Now-casting is defined as the prediction of 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 et al., 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. Now-casting can also be meaningfully applied to other target variables revealing particular aspects of the state of the economy and thereby followed closely by markets. An example is inflation, like in Modugno (2011).

The basic principle of now-casting is the exploitation of the information which is published early and possibly at higher frequencies than the target variable of interest in order to obtain an "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 are particularly valuable because of their timeliness: they are the first monthly releases relating to the current quarter. 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 et al., 2013) and the present chapter.

Until recently, the approach used in, for example, policy institutions to obtain an early estimate of GDP was based on judgment combined with simple models often called "bridge equations" (see Baffigi et al., 2004). Bridge equations are essentially regressions relating quarterly GDP growth to one or a few monthly variables (such as industrial production or surveys) aggregated to quarterly frequency. Since, typically, only partial monthly information is available for the target quarter, the monthly variables are forecasted using auxiliary models such as ARIMA. In order to exploit information from several monthly predictors bridge equations are sometime pooled (see, for example, Kitchen and Monaco, 2003).

Although part of this survey describes and evaluates this traditional approach to short-term forecasting, our focus is on models which provide a comprehensive solution to the problem of *now-casting*. In our definition *now-casting* is the exercise of reading, through the lenses of a model, the flow of data releases in real time. Ideally, a now-casting model should formalize key features of how market participants and policy makers read data in real time, which involves: monitoring many data releases, forming expectations about them and revising the assessment on the state of the economy whenever realizations diverge sizeably from those expectations.

Starting with Giannone et al. (2008) and Evans (2005), the literature has provided a formal statistical framework to embed the now-casting process defined in this broader sense. 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 process is typically associated with the unobserved state of the economy or sometimes directly with the higher frequency counterpart of the target variable. The state space representation allows the use of the Kalman filter to obtain projections for both the observed and the state variables. Importantly, the Kalman filter can easily cope with quintessential features of a now-casting information set such as different number of missing data across series 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 observed at different frequencies. Dealing with missing data at the beginning of the sample is particularly relevant for emerging markets where data collection efforts are relatively recent.

An important feature of the framework proposed by Giannone et al. (2008) is that it allows to interpret and comment various data releases in terms of the signal they provide on current economic conditions. This is possible because the Kalman filter generates projections for all the variables in the the model and therefore allows to compute, for each data release, a model-based surprise, the *news*. Bańbura and Modugno (2010) have shown formally how to link such news to the resulting now-cast revision. In this way, the data releases are weighted in a model-based rigorous way and the role of different categories of data – surveys, financial, production or labor market – in signaling changes in economic activity can be evaluated. We regard this as a major step ahead with respect to the traditional approach of bridge equations.

Since the market as well as policy makers typically watch and comment many data, essentially following the data flow throughout the quarter, the now-cast model should ideally be able to handle a high-dimensional problem. This is indeed one of the features of the econometric model proposed by Giannone et al. (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 et al. (2008) the estimation procedure exploits the fact that relevant data series, although may be numerous, co-move quite strongly so that their dynamics 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 et al. (2008) is the two-step procedure proposed by Doz et al. (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 et al. (2012) have established the consistency and robustness properties when the size of the sample and the size of the cross-section are large.

The model of Giannone et al. (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 implemented in other central banks, including the European Central Bank (ECB, 2008) and in other institutions as, for example, the International Monetary Fund (Matheson, 2011). There have been many other studies: for the United States (Lahiri and Monokroussos, 2011); for the aggregate euro area (Angelini et al., 2010, 2011; Bańbura and Modugno, 2010; Bańbura and Rünstler, 2011; Camacho and Perez-Quiros, 2010); for the single euro area countries, including France (Bessec and Doz, 2011; Barhoumi et al., 2010), Germany (Marcellino and Schumacher, 2010), Ireland (D'Agostino et al., 2008; Liebermann, 2012b), the Netherlands (de Winter, 2011), see also Rünstler et al. (2009); for China (Yiu and Chow, 2010); for the Czech Republic (Arnostova et al., 2011); for New Zealand (Matheson, 2010); for Norway (Aastveit and Trovik, 2012); for Switzerland (Siliverstovs and Kholodilin, 2012; Siliverstovs, 2012).

Results in the literature have provided support for several 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 judgment. These results suggest that now-casting has an important place in the broader forecasting literature. Third, the now-casts become progressively more accurate as the quarter comes to a close and the relevant information accumulates, hence it is important to incorporate new data as soon as they are released. Fourth, the exploitation of timely data leads to improvement in the now-cast accuracy. In particular, the relevance of various data types is not only determined by the strength of their relationship with the target variable, as it is the case in traditional forecasting exercises, but also by their timeliness. Soft information has been found to be extremely important especially early in the quarter when hard information is not available. An extensive review of the literature, including empirical findings, is provided in the survey by Bańbura et al. (2011).

In this chapter we review different statistical approaches to now-casting in more detail and perform a new empirical exercise.

The focus of the review, as we stressed earlier, is on frameworks which provide a comprehensive approach to the problem of now-casting and are based on multivariate dynamic models that can be written in the state space form. Although most applications are based on the dynamic factor model, we also review papers based on mixed frequency VARs (e.g., Giannone et al., 2009b and Kuzin et al., 2011) as they fit within the general framework. By contrast, partial models such as the traditional bridge equations capture only a limited aspect of the now-casting process. However, since these models are still used by practitioners, we include them in our review and discuss recent refinements such as MIDAS equations (used for now-casting by, e.g., Clements and Galvão, 2008, 2009; Kuzin et al., 2011).

In the empirical part, we propose and evaluate a daily dynamic factor model for now-casting US GDP with real-time data and provide illustrations on how it can be used for the reading of the real-time data flow, including not only variables available at a monthly frequency like in previous literature, but also daily financial variables and those weekly variables which are typically watched by the market.

The chapter is organized as follows. The second section defines the problem of now-casting in general and discusses different approaches used in the literature to deal with it. In the third section, we define the daily model and provide results for the empirical application. Section 4 concludes. Two appendixes contain, respectively, technical details on the implementation of the model and further empirical results.

2. NOW-CASTING: PROBLEM AND OVERVIEW OF APPROACHES

Now-casting essentially involves obtaining a projection of a variable of interest on the available information set, say Ω_v. Index v can be associated with time of a particular data release. The data vintage index v should not be confused with the model time index t. Due to frequent and non-synchronous statistical data releases v is of high frequency and is irregularly spaced. Several releases within a single day could occur.

Typically the variable of interest is an indicator collected at rather low frequency and subject to a significant publication lag. The aim is to obtain its "early estimate" on the basis of high frequency, more timely information. This has a number of implications regarding the features of the information set Ω_v. First, Ω_v could contain data collected at a wide range of frequencies, from daily to annual. Second, as different types of 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. Key in now-casting is to use all the available information. This results in a so-called "ragged" or "jagged" edge of Ω_v. Finally, the information set could be very large.

Different solutions have been proposed to the problem of mixed frequency data, we review some of them in the context of now-casting in the following sections. We focus in particular on the approaches that treat the low frequency data as high frequency with periodically missing observations and specify the underlying model dynamics at high frequency. To this end, for the low frequency variables their high frequency unobserved "counterparts" are introduced. In this, the usual convention is to index the observations of low frequency variables by time indexes ts referring to the end of the respective observation intervals.

Before explaining the details, let us introduce some notation. For simplicity, in the main text we assume that the observation intervals for variables collected at low frequency are constant across time, i.e., each month or quarter would have a constant number of days. The case of irregular intervals is discussed in the Appendix. For some variable y, we will denote by y_t^k its "counterpart" defined at an observation interval of k periods. Note that it does not necessarily mean that the variable is collected at this interval. For example, we could have an indicator collected at monthly frequency expressed as a quarterly concept. In case y is collected at an interval k we will observe y_t^k for $t = k, 2k, 3k, \ldots$ In other words, for $k > 1$ the observations of y_t^k will be periodically missing. A vector of N variables, possibly defined at different observation intervals, will be denoted by $Y_t^{K_Y} = (y_{t,1}^{k_1}, y_{t,2}^{k_2}, \ldots, y_{t,N}^{k_N})'$. For each variable in the information set, the high frequency, possibly unobserved, construct will be denoted by $y_t = y_t^1$ and the corresponding vector of N variables by $Y_t = (y_{t,1}, y_{t,2}, \ldots, y_{t,N})'$.

Note that the time unit associated with t will depend on the particular framework. For some models t would correspond to months while for the others to days, for example. For a given economic concept, the corresponding k will depend on the time unit adopted for the model. For example, for industrial production, which is collected at monthly frequency, $k = 1$ in case of a monthly model and $k = 22$ (on average) if the model is specified at daily frequency. Typically some of the variables in Ω_v will be observed at high frequency, i.e., for some n we will have $k_n = 1$, but in general this is not a necessary condition.[1]

Given this notation, the information set can be defined as $\Omega_v = \{y_{t,n}^{k_n}, t = k_n, 2k_n, \ldots, T_n(v), n = 1, 2, \ldots, N\}$. $T_n(v)$ is a multiple of k_n and refers to the last observation of variable n in the data vintage v. Due to mixed frequency of the data-set and non-synchronous releases we will have in general $T_n(v) \neq T_m(v)$ for some $n \neq m$ leading to the ragged edge described above.

As some of the approaches will be focused on one particular variable of interest, without the loss of generality, we will assume that it is the first variable, $y_{t,1}^{k_1}$.

Scalar and vector random variables will be denoted by lower and upper case letters, respectively. Parameters will be denoted by Greek letters.

2.1. Temporal Aggregation

Since key element of now-casting methodology is dealing with mixed frequency data, the issue of temporal aggregation arises.

[1] For example, the model of Aruoba et al. (2009) in the current version is specified at daily frequency but the highest frequency of the data is weekly.

The focus of this chapter will be on the approaches in which the model is specified at high frequency. Therefore it is important to understand the relation between the high frequency variables, y_t, which for key economic concepts are unobserved, and the corresponding observed low frequency series, y_t^k, $k > 1$. The relation depends on whether the corresponding indicator is a flow or a stock variable and on how it is transformed before entering the model. As most of the now-casting applications are based on models specified for stationary variables,[2] y_t^k often corresponds to a (log-)differenced, at k interval, version of some raw series z_t^k. Let z_t denote the high frequency counterpart of z_t^k. For the stock variables, such as, e.g., price indexes, the following holds:

$$z_t^k = z_t, \quad t = k, 2k, \ldots,$$

while for the flow variables, most notably GDP, we have:

$$z_t^k = \sum_{i=0}^{k-1} z_{t-i}, \quad t = k, 2k, \ldots,$$

see, e.g., Harvey (1989). In case y_t^k is a differenced version of z_t^k, we have for the stock variables:

$$y_t^k = z_t^k - z_{t-k}^k = z_t - z_{t-k} = \sum_{i=0}^{k-1} \Delta z_{t-i} = \sum_{i=0}^{k-1} y_{t-i} = \sum_{i=0}^{k-1} \omega_i^{k,s} y_{t-i}, \quad t = k, 2k, \ldots,$$

where $y_t = \Delta z_t$, $\omega_i^{k,s} = 1$ for $i = 0, 1, \ldots, k-1$ and $\omega_i^{k,s} = 0$ otherwise. For the flow variables we have that

$$\begin{aligned} y_t^k &= z_t^k - z_{t-k}^k = \sum_{i=0}^{k-1} z_{t-i} - \sum_{i=k}^{2k-1} z_{t-i} \\ &= \sum_{i=0}^{k-1} (i+1) y_{t-i} + \sum_{i=k}^{2k-2} (2k-i-1) y_{t-i} = \sum_{i=0}^{2k-2} \omega_i^{k,f} y_{t-i}, \quad t = k, 2k, \ldots, \end{aligned}$$

where $\omega_i^{k,f} = i+1$ for $i = 0, \ldots, k-1$; $\omega_i^{k,f} = 2k-i-1$ for $i = k, \ldots, 2k-2$ and $\omega_i^{k,f} = 0$ otherwise. If z_t^k is a stationary flow variable the relation is the same as in the case of differenced stock variables.

Note that in case y_t^k is a log-differenced flow variable, we follow the approximation of Mariano and Murasawa (2003):

$$\begin{aligned} y_t^k &= \log\left(z_t^k\right) - \log\left(z_{t-k}^k\right) = \log\left(\sum_{i=0}^{k-1} z_{t-i}\right) - \log\left(\sum_{i=k}^{2k-1} z_{t-i}\right) \\ &\approx \sum_{i=0}^{k-1} \log\left(z_{t-i}\right) - \sum_{i=k}^{2k-1} \log\left(z_{t-i}\right) = \sum_{i=0}^{2k-2} \omega_i^{k,f} y_{t-i}, \quad t = k, 2k, \ldots, \end{aligned}$$

[2] See, e.g., Seong et al. (2007) for an approach with non-stationary data.

where $y_t = \Delta \log \left(z_t\right)$. The approximation allows to keep the observational constraints stemming from the temporal aggregation linear.[3] For example, for quarterly GDP in a monthly model we would have:

$$y_t^3 = y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4}, \quad t = 3, 6, \ldots,$$

see, e.g., Bańbura et al. (2011), Bańbura and Modugno (2010), Kuzin et al. (2011) or Mariano and Murasawa (2003).

2.2. Joint Models in a State Space Representation

The key feature of this type of approaches is that a joint model for $Y_t^{K_Y}$ is specified and that it has a state space representation:

$$Y_t^{K_Y} = \mu + \zeta(\theta)X_t + G_t, \quad G_t \sim i.i.d.\ N(0, \Sigma_G(\theta)), \tag{1}$$

$$X_t = \varphi(\theta)X_{t-1} + H_t, \quad H_t \sim i.i.d.\ N(0, \Sigma_H(\theta)), \tag{2}$$

where the measurement equation (1) links the vector of observed variables, $Y_t^{K_Y}$, to a vector of possibly unobserved state variables, X_t, and the transition equation (2) specifies the dynamics of the latter (see, e.g., Harvey, 1989, for a comprehensive treatment of state space models). We do not add index t in order not to complicate the notation but both the matrices of the coefficients, $\zeta(\theta)$ and $\varphi(\theta)$, as well as the covariance matrices of the disturbances, $\Sigma_G(\theta)$ and $\Sigma_H(\theta)$, could be time-varying.

Given a model with a representation (1) and (2) and the parameters θ, the Kalman filter and smoother provide conditional expectations of the state vector on the information set Ω_v and the associated precision[4]:

$$X_{t|\Omega_v} = \mathbb{E}_\theta\left[X_t|\Omega_v\right], \quad P_{t|\Omega_v} = \mathbb{E}_\theta\left[\left(X_t - \mathbb{E}_\theta\left[X_t|\Omega_v\right]\right)\left(X_t - \mathbb{E}_\theta\left[X_t|\Omega_v\right]\right)'\right].$$

Importantly, the Kalman filter and smoother can efficiently deal with any missing observations in $Y_t^{K_Y}$ and provide the conditional expectation for those. Consequently, now-casts or forecasts can be easily obtained for the target variable and for the predictors. As in this framework the problems of mixed frequency and ragged edge are essentially missing data problems, they are easily solved by Kalman filter and smoother apparatus. Last but not least, joint state space representation also allows to derive model-based news of statistical data releases and to link them to the now-cast revision, see Section 2.3.

[3] Proietti and Moauro (2006) propose to use a non-linear smoothing algorithm to impose the temporal constraint exactly. Proietti (2011) further shows how to account for cross-sectional observational constraints.

[4] In case the disturbances are not Gaussian the Kalman smoother provides the minimum mean square linear (MMSLE) estimates.

$\zeta(\theta)P_{t|\Omega_v}\zeta(\theta)'$ is sometimes referred to as "filter uncertainty" (Giannone et al., 2008) as it captures the part of the uncertainty underlying the now-cast of $Y_t^{K_Y}$ that is associated with signal extraction.[5]

Different versions of the general model given by (1) and (2) have been considered in the literature.

2.2.1. Factor Model

As stressed above the real-time data flow is inherently high dimensional. As a consequence it is important to use a parsimonious model that allows to avoid parameter proliferation but at the same time is able to capture the salient features of the data. A dynamic factor model is particularly suitable in this context. In a dynamic factor model, each series is modeled as the sum of two orthogonal components: the first, driven by a handful of unobserved factors captures the joint dynamics and the second is treated as an idiosyncratic residual. If there is a high degree of co-movement among the series, the bulk of the dynamics of any series can be captured by the few factors. There is considerable empirical evidence that indeed this is the case for large panels of macroeconomic variables (see Sargent and Sims, 1977; Giannone et al., 2004; Watson, 2004 and, for a recent survey see Stock and Watson, 2011) and this is why we have chosen this modeling strategy here.

The most common version in the context of now-casting specifies that the high frequency variables, Y_t, have a factor structure and that the factors, F_t, follow a vector autoregressive (VAR) process:

$$Y_t = \mu + \Lambda F_t + E_t, \quad E_t \sim i.i.d.\ N(0, \Sigma_E), \tag{3}$$

$$F_t = \Phi(L)F_t + U_t, \quad U_t \sim i.i.d.\ N(0, \Sigma_U). \tag{4}$$

The latter feature can be particularly important for now-casting, as in the presence of ragged edge, both cross-sectional and "dynamic" information is useful. Σ_E is assumed to be diagonal but, as discussed below, the estimates are robust to violations of this assumption.

This is the type of model that Giannone et al. (2008) have proposed to now-cast GDP from a large set of monthly indicators. In their application Y_t contains only monthly (observed) variables, hence, Eqs. (3) and (4) constitute a state space representation and Kalman filter and smoother can be run to obtain the estimates of the factors. The now-casts are then obtained via a regression of GDP on temporally aggregated factor estimates:

$$y_{t,1}^{k_1} = \alpha + \beta F_{t|\Omega_v}^{k_1} + e_t^{k_1}, \quad t = k_1, 2k_1, \ldots, \tag{5}$$

Giannone et al. (2008) estimate the state space representation (3) and (4) by a so-called two-step procedure. In the first step, the parameters of the state space representation are

[5] Note that, for given set of parameters θ and for t sufficiently large, such that the Kalman smoother has approached its steady state, filter uncertainty can be considered time invariant in the sense that it will not depend on t but rather on the shape of the ragged edge in Ω_v with respect to the target quarter, see Bańbura and Rünstler (2011) for a formal explanation.

estimated using principal components derived from a "balanced" panel of Y_t as factor estimates. The balanced panel is obtained by considering only the sample for which all observations are available.[6] In the second step, factors are re-estimated by applying the Kalman smoother to the entire information set.

As stressed in the introduction, this approach has been widely used and applied for different countries. Bańbura and Rünstler (2011) and Angelini et al. (2010) modify it slightly by including quarterly variables into the state space representation using the temporal aggregator variables as explained below.

Doz et al. (2012) show that large systems like (3) and (4) can be also estimated by maximum likelihood. They use the Expectation Maximization (EM) algorithm to obtain the maximum likelihood estimates. The EM algorithm is a popular tool to estimate the parameters for models with unobserved components and/or missing observations, such as (3) and (4). The principle is to write the likelihood in terms of both observed and unobserved data, in this case the state variables, and to iterate between two operations: (i) compute the expectation of the log-likelihood (sufficient statistics) conditional on the data using the parameter estimates from the previous iteration and (ii) re-estimate the parameters through the maximization of the expected log-likelihood. In case of (3) and (4) this boils down to iterating the two-step procedure until convergence, at each step correcting for the uncertainty associated with the estimation of the common factors, see Watson and Engle (1983) and Shumway and Stoffer (1982).

Maximum likelihood has a number of advantages compared to the principal components and the two-step procedure. First it is more efficient for small systems. Second, it allows to deal flexibly with missing observations. Third, it is possible to impose restrictions on the parameters. For example, Bańbura and Modugno (2010) impose the restrictions on the loadings to reflect the temporal aggregation. Bańbura et al. (2011) introduce factors that are specific to groups of variables.

Doz et al. (2011, 2012) show consistency of the two-step and maximum likelihood estimates, respectively. The asymptotic properties are analyzed under different sources of misspecification: omitted serial and cross-sectional correlation of the idiosyncratic components, and non-normality. It is shown that the effects of misspecification on the estimation of the common factors is negligible for large sample size (T) and the cross-sectional dimension (N). We would like to stress here that large cross-section is just an asymptotic device to study the properties of the estimates when more data are included and hence it does not mean that robustness is achieved only when the number of variables approaches infinity. How large is large enough is an empirical question, and Monte Carlo exercises of Doz et al. (2011, 2012) show that substantial robustness is achieved already with a handful of variables.

[6] Reduced rank of the disturbances in the factor VAR is often imposed. As discussed in Forni et al. (2009) this feature enforces dynamic heterogeneity in the factor structure.

These results provide theoretical ground for the use of maximum likelihood estimation for factor models in now-casting since they point to its robustness to misspecification. Maximum likelihood estimation with the EM algorithm for a factor model adapted to the now-casting problem has been proposed by Bańbura and Modugno (2010) and this is the approach we use in the empirical application of this chapter. Camacho and Perez-Quiros (2010) and Frale et al. (2011), for example, have also applied maximum likelihood estimation although with a different implementation.

2.2.2. Model with Daily Data

Most of the now-casting applications have been based on monthly and quarterly variables. Modugno (2011) develops a model with daily, weekly, and monthly stock data for now-casting inflation. In this chapter, we generalize this framework by adding flow variables. As above we assume that the high frequency concepts follow a factor model (3) and (4). Consequently for nth variable defined at k_n interval we have:

$$y_{t,n}^{k_n} = \sum_{i=0}^{2k_n-2} \omega_i^{k_n,\cdot} y_{t-i,n} = \sum_{i=0}^{2k_n-2} \omega_i^{k_n,\cdot} \left(\Lambda_{n,\cdot} F_{t-i} + e_{t-i,n} \right),$$

where $\omega_i^{k_n,\cdot} = \omega_i^{k_n,f}$ for the flow variables and $\omega_i^{k_n,\cdot} = \omega_i^{k_n,s}$ for the stock variables. $\Lambda_{n,\cdot}$ denotes the nth row of Λ.

To limit the size of the state vector, temporal aggregator variables for F_t are constructed. We need separate aggregators for each frequency and for stock and flow variables, $F_t^{k,f}$ and $F_t^{k,s}$, $k = k_q, k_m, k_w$, where $k = k_q, k_m$, and k_w refer to the (average) number of days in a quarter, month, and week, respectively.[7] These variables aggregate recursively F_t so that at the end of the respective period we have:

$$F_t^{k,\cdot} = \sum_{i=0}^{2k-2} \omega_i^{k,\cdot} F_{t-i}, \quad t = k, 2k, \ldots, \quad k = k_q, k_m, k_w.$$

The details on how the aggregators are constructed are provided in the Appendix.

Note that analogously to the common component, the idiosyncratic error in the measurement equation will be a moving average of the daily $e_{t,n}$. However, in the estimation we will assume that at k interval, at which it is defined, it is a white noise.

Another source of misspecification is due to conditional heteroskedasticity and fat tails which is typical of daily data. Fortunately, as discussed in Section 2.2, the factor model is robust to those misspecifications when the factors are extracted from many variables (see Doz et al., 2012). However, reducing misspecifications by explicitly modeling key data features might give sizeable advantages in finite samples. Important directions for future research, especially for daily data, consists in modeling stochastic volatility and rare big shocks, see Marcellino et al. (2012) and Curdia et al. (2012).

[7] In fact, we do not have quarterly stock variables in the data-set.

Let $Y_t^{k,\cdot}$ collect the variables observed at interval k (flows or stocks). The measurement equation can be written as follows:

$$\begin{pmatrix} Y_t^{k_q,f} \\ Y_t^{k_m,f} \\ Y_t^{k_m,s} \\ Y_t^{k_w,f} \\ Y_t^{k_w,s} \\ Y_t^{k_d} \end{pmatrix} = \begin{pmatrix} \tilde{\Lambda}^{q,f} & 0 & 0 & 0 & 0 & 0 \\ 0 & \tilde{\Lambda}^{m,f} & 0 & 0 & 0 & 0 \\ 0 & 0 & \Lambda^{m,s} & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{\Lambda}^{w,f} & 0 & 0 \\ 0 & 0 & 0 & 0 & \Lambda^{w,s} & 0 \\ 0 & 0 & 0 & 0 & 0 & \Lambda^{d} \end{pmatrix} \begin{pmatrix} \tilde{F}_t^{k_q,f} \\ \tilde{F}_t^{k_m,f} \\ F_t^{k_m,s} \\ \tilde{F}_t^{k_w,f} \\ F_t^{k_w,s} \\ F_t \end{pmatrix} + E_t^{K_Y}. \tag{6}$$

For the flow variables an auxiliary aggregator variable, $\bar{F}_t^{k,f}$, is necessary: $\tilde{F}_t^{k,f\prime} = (F_t^{k,f\prime}\bar{F}_t^{k,f\prime})$ and $\tilde{\Lambda}^{\cdot,f} = (\Lambda^{\cdot,f}\ 0)$, see the appendix for details.

The coefficients of the transition equation are time-varying:

$$\begin{pmatrix} I_{2r} & 0 & 0 & 0 & 0 & \mathcal{W}_t^{k_q,f} \\ 0 & I_{2r} & 0 & 0 & 0 & \mathcal{W}_t^{k_m,f} \\ 0 & 0 & I_r & 0 & 0 & \mathcal{W}_t^{k_m,s} \\ 0 & 0 & 0 & I_{2r} & 0 & \mathcal{W}_t^{k_w,f} \\ 0 & 0 & 0 & 0 & I_r & \mathcal{W}_t^{k_w,s} \\ 0 & 0 & 0 & 0 & 0 & I_r \end{pmatrix} \begin{pmatrix} \tilde{F}_t^{k_q,f} \\ \tilde{F}_t^{k_m,f} \\ F_t^{k_m,s} \\ \tilde{F}_t^{k_w,f} \\ F_t^{k_w,s} \\ F_t \end{pmatrix}$$

$$= \begin{pmatrix} \mathcal{I}_t^{k_q,f} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathcal{I}_t^{k_m,f} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathcal{I}_t^{k_m,s} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathcal{I}_t^{k_w,f} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathcal{I}_t^{k_w,s} & 0 \\ 0 & 0 & 0 & 0 & 0 & \Phi \end{pmatrix} \begin{pmatrix} \tilde{F}_{t-1}^{k_q,f} \\ \tilde{F}_{t-1}^{k_m,f} \\ F_{t-1}^{k_m,s} \\ \tilde{F}_{t-1}^{k_w,f} \\ F_{t-1}^{k_w,s} \\ F_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ U_t \end{pmatrix}, \tag{7}$$

where $\mathcal{W}_t^{\cdot,\cdot}$ contain appropriate aggregation weights and $\mathcal{I}_t^{\cdot,\cdot}$ are matrices of zeros and ones, see the appendix for details.[8]

The model is estimated by maximum likelihood using the EM algorithm, as explained in the appendix. Initial parameter values for the algorithm are obtained using principal components as factor estimates. To extract the principal components the missing observations are "filled in" using splines and then the data are filtered to reflect different aggregation intervals.

2.2.3. *Mixed Frequency VAR*

Another type of model that can be cast in a state space representation is a VAR. Different approaches have been considered to deal with the issue of mixed frequency. One solution, analogous to the approach explained above, is to specify that the high frequency concepts

[8] The transition equation is obtained by pre-multiplying (7) by the inverse of the left-hand side matrix.

follow a VAR:

$$\Psi(L)\Big(Y_t - \mu\Big) = E_t, \quad E_t \sim i.i.d.\ N(0, \Sigma_E)$$

and to derive the measurement equation using the (approximate) temporal aggregation relationships between the observed variables and Y_t as explained in Section 2.1, see, e.g., Mariano and Murasawa (2010). As in the case of factor models explained above, Kalman filter and smoother can be used to obtain the now-casts. Giannone et al. (2009b) and Kuzin et al. (2011) apply this type of model to now-cast euro area GDP with monthly indicators. Earlier applications include Zadrozny (1990) and Mittnik and Zadrozny (2004).

Another solution is sometimes referred to as "blocking," see, e.g., Chen et al. (2012). The model is specified at low frequency and the high frequency information is "distributed" into multiple series. For example, in a system with monthly and quarterly variables, blocking consists in having three different time series for each monthly variable, one for each of the three months of the quarter. For a detailed analysis of blocked linear systems see Chen et al. (2012). McCracken et al. (2013) use this approach for now-casting with a large Bayesian VAR.

VAR is a less parsimonious specification than a factor model. For large information sets two solutions to the curse of dimensionality problem could be adopted. Either the forecasts from many smaller systems could be combined or Bayesian shrinkage could be employed to avoid over-fitting in a large system, see Bańbura et al. (2010), Giannone et al. (2012) or Koop (2013) (see also the Chapter on Bayesian VAR in this volume by Karlsson, 2013). Recent papers that have used Bayesian shrinkage to handle large information sets in the context of now-casting are Bloor and Matheson (2011) and Carriero et al. (2012).

2.3. Now-Cast Updates and News

We argue that now-casting goes beyond producing a single prediction for a reference period. The aim is rather to build a framework for the reading of the flow of data releases in real time.

At each time that new data become available, a now-casting model produces an estimate of the variable of interest, say the current quarter growth rate of GDP, thereby providing a sequence of updates for this fixed event. Within a state space framework, the same model also produces forecasts for all variables we are interested in tracking so as to allow extracting the *news* or the "unexpected" component from the released data. Having model-based news for all variables allows obtaining the revision of the GDP now-cast as the weighted sum of those news where the weights are estimated by the model. The framework therefore provides a key for the understanding of changes in the estimates of current economic activity over time and helps evaluating the significance of each data publication.

Following Bańbura and Modugno (2010) we can explain these ideas formally.

Let us consider two consecutive data vintages, Ω_v and Ω_{v+1}. The information sets Ω_v and Ω_{v+1} can differ for two reasons: first, Ω_{v+1} contains some newly released figures, $\{y^{k_{n_j}}_{t_j,n_j}, j = 1, \ldots, J_{v+1}\}$, which were not available in Ω_v; second, some of the data might have been revised. To simplify the notation, in what follows we assume that no past observations for the variable of interest $y^{k_1}_{t,1}$ are contained in the release and that $k_{n_j} = 1$, $j = 1, \ldots, J_{v+1}$ so that $y^{k_{n_j}}_{t_j,n_j} = y_{t_j,n_j}$. The derivations can be modified to a general case in a straightforward manner. More importantly, we abstract from data revisions and therefore we have:

$$\Omega_v \subset \Omega_{v+1} \quad \text{and} \quad \Omega_{v+1} \backslash \Omega_v = \{y_{t_j,n_j}, j = 1, \ldots, J_{v+1}\},$$

hence the information set is "expanding." Note that since different types of data are characterized by different publication delays, in general we will have $t_j \neq t_l$ for some $j \neq l$.

Let us now look at the two consecutive now-cast updates, $\mathbb{E}\left[y^{k_1}_{t,1}|\Omega_v\right]$ and $\mathbb{E}\left[y^{k_1}_{t,1}|\Omega_{v+1}\right]$. The new figures, $\{y_{t_j,n_j}, j = 1, \ldots, J_{v+1}\}$, will in general contain some new information on $y^{k_1}_{t,1}$ and consequently lead to a revision of its now-cast. From the properties of conditional expectation as an orthogonal projection operator, it follows that:

$$\underbrace{\mathbb{E}\left[y^{k_1}_{t,1}|\Omega_{v+1}\right]}_{\text{new forecast}} = \underbrace{\mathbb{E}\left[y^{k_1}_{t,1}|\Omega_v\right]}_{\text{old forecast}} + \underbrace{\mathbb{E}\left[y^{k_1}_{t,1}|A_{v+1}\right]}_{\text{revision}},$$

where

$$A_{v+1} = (a_{v+1,1} \cdots a_{v+1,J_{v+1}})', \quad a_{v+1,j} = y_{t_j,n_j} - \mathbb{E}\left[y_{t_j,n_j}|\Omega_v\right], \;\; j = 1, \ldots, J_{v+1}.$$

A_{v+1} represents the part of the release $\{y_{t_j,n_j}, j = 1, \ldots, J_{v+1}\}$, which is "orthogonal" to the information already contained in Ω_v. In other words, it is the "unexpected" (with respect to the model), part of the release. Therefore, we label A_{v+1} as the *news*. Note that it is the news and not the release itself that leads to now-cast revision. In particular, if the new numbers in Ω_{v+1} are exactly as predicted, given the information in Ω_v, or in other words "there is no news," the now-cast will not be revised.

We can further develop the expression for the revision, that is the difference between the new and the old now-cast, as:

$$\mathbb{E}\left[y^{k_1}_{t,1}|A_{v+1}\right] = \mathbb{E}\left[y^{k_1}_{t,1}A'_{v+1}\right]\mathbb{E}\left[A_{v+1}A'_{v+1}\right]^{-1}A_{v+1}.$$

In what follows we abstract from the problem of parameter uncertainty.

For the model written as (1) and (2) with a diagonal Σ_G, this can be further developed as:

$$\mathbb{E}\left[y^{k_1}_{t,1}a_{v+1,j}\right] = \zeta_{1,\cdot}\mathbb{E}\Big[\left(X_t - \mathbb{E}\left[X_t|\Omega_v\right]\right)\left(X_{t_j} - \mathbb{E}\left[X_{t_j}|\Omega_v\right]\right)'\Big]\zeta'_{n_j,\cdot} \quad \text{and}$$
$$\mathbb{E}\left[a_{v+1,j}a_{v+1,l}\right] = \zeta_{n_j,\cdot}\mathbb{E}\Big[\left(X_{t_j} - \mathbb{E}\left[X_{t_j}|\Omega_v\right]\right)\left(X_{t_l} - \mathbb{E}\left[X_{t_l}|\Omega_v\right]\right)'\Big]\zeta'_{n_l,\cdot} + \Sigma_{G,jl}1_{j=l},$$

where $\Sigma_{G,jl}$ is the element of the Σ_G from the jth row and lth column. Kalman filter and smoother provide appropriate expectations.

As a result, we can find a vector $\mathcal{D}_{v+1} = (\delta_{v+1,1}, \ldots, \delta_{v+1,J_{v+1}})$ such that the following holds:

$$\underbrace{\mathbb{E}\left[y^{k_1}_{t,1}|\Omega_{v+1}\right] - \mathbb{E}\left[y^{k_1}_{t,1}|\Omega_v\right]}_{\text{revision}} = \mathcal{D}_{v+1}A_{v+1} = \sum_{j=1}^{J_{v+1}} \delta_{v+1,j}\Big(\underbrace{y_{t_j,n_j} - \mathbb{E}\left[y_{t_j,n_j}|\Omega_v\right]}_{\text{news}}\Big). \tag{8}$$

In other words, the revision can be decomposed as a weighted average of the news in the latest release. What matters for the revision is both the size of the news as well as its relevance for the variable of interest, as represented by the associated weight $\delta_{v+1,j}$. Formula (8) can be considered as a generalization of the usual Kalman filter update equation (see, e.g., Harvey, 1989, Eq. (3.2.3a)) to the case in which "new" data arrive in a non-synchronous manner.

Note that filter uncertainty for $y^{k_1}_{t,1}$ decreases with the new release and the reduction can be decomposed along similar lines.

Relationship (8) enables us to trace sources of forecast revisions.[9] 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.[10] In addition, we can produce statements like, e.g., "after the release of industrial production, the forecast of GDP went up because the indicators turned out to be (on average) higher than expected".[11]

2.4. "Partial" Models

In contrast to approaches described in Section 2.2, the methodologies that we label as "partial" do not specify a joint model for the variable of interest and for the predictors. One limitation of partial models is that without a joint representation the model-based news of releases and their impact on the now-cast cannot be derived. Other drawbacks include the need for auxiliary models or for separate set of parameters for each data vintage. In spite of those limitations we review also partial models here because they have a long tradition in policy institutions, in particular central banks.

Again let us assume for simplicity that $k_n = 1, n \neq 1$ so that $y^{k_n}_{t,n} = y_{t,n}$. In other words, all the predictors are observed at the same, high frequency. The methodologies presented below typically can be generalized to relax this restriction.

The following "partial" models have been studied in the literature.

[9] Note, that the contribution from the news is equivalent to the change in the overall contribution of the series to the forecast (the measure proposed in Bańbura and Rünstler, 2011) when the correlations between the predictors are not exploited in the model. Otherwise, those measures are different. In particular, there can be a change in the overall contribution of a variable even if no new information on this variable was released. Therefore news is a better suited tool for analyzing the sources of forecasts revisions, see Bańbura and Modugno (2010) for the details.

[10] If the release concerns only one group or one series, the contribution of its news is simply equal to the change in the forecast.

[11] This holds of course for the indicators with positive entries in $\delta_{v+1,j}$.

2.4.1. Bridge Equations

In this type of model, the now-cast and forecasts of $y_{t,1}^{k_1}$ are obtained via the following regression:

$$y_{t,1}^{k_1} = \alpha + \beta y_{t,n}^{k_1} + e_t^{k_1}, \quad t = k_1, 2k_1, \ldots, \tag{9}$$

where $y_{t,n}^{k_1}$ is a predictor aggregated to the *lower* frequency, i.e., the frequency of the target variable. Hence the mixed frequency problem is solved by temporal aggregation of the predictors to the lower frequency. To handle ragged edge auxiliary models, such as ARMA or VAR, are used to forecast $y_{t,n}$ to close the target period of interest.

This is the "traditional" now-casting tool, popularly employed at central banks to obtain early estimates of GDP or its components. The predictors are typically monthly, see, e.g., Kitchen and Monaco (2003), Parigi and Golinelli (2007), Parigi and Schlitzer (1995) and Baffigi et al. (2004).

Equation (9) is typically estimated by the OLS. It can be further extended to include more predictors or the lags of the dependent variable. In case the information set is large, forecast combination is often performed (Kitchen and Monaco, 2003; Diron, 2008; Angelini et al., 2011; Rünstler et al., 2009). Bridge equations can be also combined in a so-called bottom-up approach where one now-casts GDP by aggregating the now-casts of its components exploiting national accounts identities (see Hahn and Skudelny, 2008; Drechsel and Scheufele, 2012; Baffigi et al., 2004).

Note that the model of Giannone et al. (2008) can be also interpreted as "bridging with factors" as the factor estimates obtained by the Kalman filter and smoother would be plugged into an equation similar to (9) to obtain the now-casts, cf. Eq. (5).

2.4.2. MIDAS-Type Equations

In contrast to the previous approach in a MIDAS-type model the predictors are included in the regression at their original observation frequency:

$$y_{t,1}^{k_1} = \alpha + \beta \Gamma(L, \theta) y_{t-h_n,n} + e_t^{k_1}, \quad t = k_1, 2k_1, \ldots, \tag{10}$$

where $\Gamma(L,\theta)$ is a lag polynomial. Since for large k_1 many lags of the explanatory variable might be required, key in this approach is that $\Gamma(L,\theta)$ is parsimoniously parameterised. Various versions have been proposed (Ghysels et al., 2003), including exponential Almon polynomial for which $\Gamma(L,\theta) = \sum_{m=1}^{M} \gamma(m,\theta) L^m$ with $\theta = (\theta_1, \theta_2)$ and $\gamma(m,\theta) = \frac{\exp(\theta_1 m + \theta_2 m^2)}{\sum_{m=1}^{M} \exp(\theta_1 m + \theta_2 m^2)}$. In contrast to approaches explained above, MIDAS-type regression implies that the temporal aggregation weights are data driven.

Regarding the problem of ragged edge, the solution in this type of approach can be thought of as re-aligning each time series. The time series with missing observations at the end of the sample are shifted forward in order to obtain a balanced data-set with the most recent information.[12] The parameters in Eq. (10) depend on h_n, which is determined by

[12] Re-aligning has been a popular strategy do deal with ragged-edge data (see, e.g., Altissimo et al., 2001, 2010 and de Antonio Liedo and Muoz, 2010).

the difference between the forecast target period and the period of the last observation of the predictor. As a consequence, separate models need to be estimated for different data vintages as the corresponding h_n vary. The case of $h_n < k_1$, i.e., when some data referring to the target quarter are available, is sometimes labeled as MIDAS with leads (Andreou et al., 2013).

Applications of this type of model to now-casting include Clements and Galvão (2008, 2009) or Kuzin et al. (2011) who use monthly indicators to forecast GDP. Recently, Andreou et al. (2013) also include daily financial variables to the equation. Equation (10) is typically estimated by non-linear least squares. Clements and Galvão (2008) propose how to add a lag of the low frequency variable in order to avoid a seasonal response of the dependent variable to the predictors. They use the Broyden–Fletcher–Goldfarg–Shanno method to obtain the estimates of the parameters.

The MIDAS equations suffer from the curse of dimensionality problem and can include only a handful of variables. Forecast combination is a popular strategy to deal with large information sets (see, e.g., Andreou et al., 2013).

As an alternative, Marcellino and Schumacher (2010) propose to now-cast GDP from the following equation:

$$y_{t,1}^{k_1} = \alpha + \beta\Gamma(L,\theta)F_{t-h_F|\Omega_v} + e_t^{k_1}, \quad t = k_1, 2k_1, \ldots,$$

where $F_{t|\Omega_v}$ are factors estimated from a set of monthly predictors following the methodology of Giannone et al. (2008) and h_F corresponds to the difference between the forecast target period and the latest observation in the predictor set.

As we have already remarked, in order to understand why and how now-casts change with the arrival of new information it is important to have a joint model that allows to form expectations and hence derive the news component of data releases and their impact on the now-cast. Attempts to circumvent the problem within partial models has to be necessarily based on a heuristic procedure. For example Ghysels and Wright (2009) construct news using market expectations. The latter are linked to the change in the forecast by estimating additional auxiliary regressions.

3. EMPIRICAL APPLICATION

The aim of the empirical application is to establish whether daily and weekly variables contribute to the precision of the now-cast of quarterly GDP growth and to use our framework to study the extent to which stock prices are connected with macroeconomic variables.

The now-casting framework is the appropriate one for studying the role of financial variables for macro forecasting since it takes into account the timeliness of financial information. Although the role of financial variables for the forecast of real economic conditions has been studied extensively in the forecasting literature (see Stock and Watson, 2003;

Forni et al., 2003), results are typically based on models which do not take into account the publication lags associated with different data series and therefore miss timeliness as an essential feature of financial information in real-time forecasting. What is more, most of the now-casting studies we have reviewed here are based on monthly models. Daily variables, when included, are first converted to monthly frequency hence their advantage due to timeliness might be partly lost (see, e.g., Giannone et al., 2008; Bańbura and Rünstler, 2011). This study corrects this feature by treating all data at their original frequency.

3.1. Data

We are considering 24 series of which only GDP is quarterly. Table 4.1 provides the list of variables, including the observation frequency and the transformation we have adopted. 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. We have aimed at selecting the "headline" macroeconomic variables. Accordingly, the series we collected are marked on Bloomberg website as "Market Moving Indicators". The daily financial variables include S&P 500 stock price index, short- and long-term interest rates, effective exchange rate and the price of oil. To give an idea on the timeliness of each macroeconomic variable Table 4.1 also provides the publication delay, i.e., the difference (in days) between the end of the reference period and the date of the respective release for January 2011.[13] We can observe the typical pattern according to which soft data, notably surveys, are published more timely than the hard data.

Let us make some remarks about the criteria used for the selection of the macroeconomic variables. We only include the headlines of each macroeconomic report since these are the data followed by the market and extensively commented by the newspapers. For example, for the release of industrial production and capacity utilization we only include total indexes hence disregarding the sectoral disaggregation. The disaggregated data for each release were considered in Giannone et al. (2008) whose universe of data included around 200 time series. Bańbura and Modugno (2010) and Bańbura et al. (2011) analyze the marginal impact on the now-cast precision of disaggregated data and show that it is minimal, result which is supported by the observation that markets only focus on the headlines of each report. The same authors also show that the inclusion of disaggregated data does not deteriorate the performance of the model, supporting the results on the robustness of the factor model to data selection (see the empirical analysis in Bańbura et al., 2010 for longer horizons forecasting and the simulation study of Doz et al., 2011, 2012). In this chapter we have therefore decided to disregard them but the results just cited carry two important implications for empirical work in this field. First, the fact that

[13] The release dates typically vary from month to month. For example, industrial production is released between the 14th and 18th day of each month.

Table 4.1 Data

Number	Name	Frequency	Publication Delay (in Days After) Reference Period)	Transformation Log	Transformation Diff
1	Real Gross Domestic Product	Quarterly	28	×	×
2	Industrial Production Index	Monthly	14	×	×
3	Purchasing Manager Index, Manufacturing	Monthly	3		×
4	Real Disposable Personal Income	Monthly	29	×	×
5	Unemployment Rate	Monthly	7		×
6	Employment, Non-farm Payrolls	Monthly	7	×	×
7	Personal Consumption Expenditure	Monthly	29	×	×
8	Housing Starts	Monthly	19	×	×
9	New Residential Sales	Monthly	26	×	×
10	Manufacturers' New Orders, Durable Goods	Monthly	27	×	×
11	Producer Price Index, Finished Goods	Monthly	13	×	×
12	Consumer Price Index, All Urban Consumers	Monthly	14	×	×
13	Imports	Monthly	43	×	×
14	Exports	Monthly	43	×	×
15	Philadelphia Fed Survey, General Business Conditions	Monthly	−10		×
16	Retail and Food Services Sales	Monthly	14	×	×
17	Conference Board Consumer Confidence	Monthly	−5		×
18	Bloomberg Consumer Comfort Index	Weekly	4		×
19	Initial Jobless Claims	Weekly	4	×	×
20	S&P 500 Index	Daily	1	×	×
21	Crude Oil, West Texas Intermediate (WTI)	Daily	1	×	×
22	10-Year Treasury Constant Maturity Rate	Daily	1		×
23	3-Month Treasury Bill, Secondary Market Rate	Daily	1		×
24	Trade Weighted Exchange Index, Major Currencies	Daily	1		×

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

including disaggregated data does not worsen forecasting performance says that, if for the problem at hand we were interested in commenting them, we could include them in the model without paying a cost in terms of larger forecast error. Second, including variables with little marginal forecasting power does not hurt results and therefore it is not necessary to select variables using criteria outside the model; the model itself attributes the appropriate weight to the different predictors.

An alternative approach consists in selecting the variables using statistical criteria as suggested in Boivin and Ng (2006) and Bai and Ng (2008). We do not consider it for a number of reasons. First, the algorithms for selecting predictors have been developed for balanced panels and hence they are not suitable in the context of now-casting since they are designed to account only for the quality of the signal but not for timeliness. Second, empirically it has been found that, if data are collinear, there is no major difference in the forecasting performance of models based on selected or all available predictors (see De Mol et al., 2008).[14] Finally and most importantly, because of collinearity among predictors, variable selection is inherently unstable, i.e., the set of predictors selected is very sensitive to minor perturbation of the data-set, such as adding new variables or extending the sample length (see De Mol et al., 2008). Similar instabilities have also been found in the context of model selection and model averaging (Ouysse, 2011; Stock and Watson, 2012).

The out-of-sample now-cast evaluations are performed on the basis of real-time data vintages for the series described in Table 4.1. This implies that at each date in which we produce a forecast we use the data that were available just before or on the same day. The real-time database has been downloaded from ALFRED, the US real-time database maintained by the Federal Reserve Bank of St. Louis.

Notice that most of the now-casting studies we reviewed earlier involve out-of-sample forecast evaluations in *pseudo* rather than in fully real time like here. Pseudo real-time forecast evaluation design mimics the real-time situation in the sense that the data publication delays are replicated following a realistic *stylized* publication calendar and that the model is estimated recursively. However, given the difficulties of obtaining real-time data vintages for many series, final revised data are used throughout the exercise. Hence the effects of data revisions, which for some variables can be sizable, are disregarded. As noted by, e.g., Croushore (2011) this could affect the results of forecast evaluation and comparison. The few studies with real-time data include, e.g., Camacho and Perez-Quiros (2010); Lahiri and Monokroussos (2011); Liebermann (2012a); Siliverstovs (2012).

3.2. Now-Casting GDP

In this section we study the model performance for real GDP now-casting from different perspectives. In particular, we analyze the evolution of the now-cast and its uncertainty

[14] This result also emerges from a careful reading of the empirical results of Boivin and Ng (2006).

in relation to releases of different categories of data. We focus on GDP since it is the variable that best summarizes the state of the economy.

The benchmark factor model described in Section 2.2.2 including the variables listed in Table 4.1 is estimated following Bańbura and Modugno (2010) and is specified with one factor only. Our choice is mainly motivated by simplicity and by the fact that results based on two factors are qualitatively similar to those based on one factor.[15] In order to shed light on the importance of high frequency data for the accuracy of the now-cast, we also construct two alternative models. The first is a factor model analogous to the benchmark model, but at monthly frequency and based on the information set that excludes the weekly and daily variables. The second is a small monthly factor model based on five (hard) indicators, namely real GDP, industrial production, real disposable income, retail sales and employment. This is the data-set used in Stock and Watson (1989) for estimating a coincident indicator for the US economy and it is also considered by the the NBER's Business Cycle Dating Committee.

We also compare the performance of the benchmark model to that of bridge equations. For the latter we estimate a separate regression of GDP on each of the 23 predictors. In each case we aggregate the monthly predictors to quarterly frequency as explained in Section 2.1.[16] As mentioned in Section 2, we use an auxiliary model to produce the forecast for the predictors over the "missing" months in the quarter of interest. To this end we estimate an autoregressive (AR) models on the series and use the BIC criteria to select the number of lags.

We finally report results for the survey of professional forecasters (SPF) which are available in the middle second month of each quarter.

Depending on when in the quarter we perform the now-cast update, the availability of information differs. For example, in the first month we only have very limited information on the current quarter, in particular no hard data. To assess how the accuracy of the now-cast improves as information on the current quarter accumulates, we evaluate the now-cast at different points within a quarter. Precisely, we produce the now-cast four times per month, on the 7th, 14th, 21st, and 28th, starting in the first month of the current quarter up to the first month of the following quarter.[17] Each time, the model is re-estimated in order to take into account parameter uncertainty.

[15] A more systematic approach to the choice of the number of factors is to set the parameterizations in a data driven fashion. This can be done in different ways: (a) minimizing recursive mean square forecast error; (b) averaging across all possible parameterizations; and (c) applying recursively information criteria. The literature typically indicates that results are robust across model parameterizations (see, for example Angelini et al., 2011) although some papers have advocated pooling across specifications (see Kuzin et al., 2013). Bańbura and Modugno (2010) note that the recent recession period saw larger differences between specifications in terms of forecast accuracy and advocate pooling.

[16] All the bridge equation models in this exercise are based on monthly variables. Accordingly, daily and weekly data are first averaged over month and the auxiliary models are run at monthly frequency. Partial monthly information is thus disregarded.

[17] As US GDP is typically released around the end of the first month of the following quarter.

The evaluation sample is 1995–2010. For most of the series the data vintages are available since January 1995, first date of our evaluation. For the data for which real-time vintages starting from January 1995 are not available, we use the oldest vintage available and we apply a "stylized" publication delay, as provided in Table 4.1.[18] The estimation is recursive, i.e., the first observation in the sample is fixed to the first business day of January 1983. Note that Bloomberg Consumer Comfort Index dates back only to 1985, however the EM algorithm used here can deal with series of different lengths.[19]

Figure 4.1 reports the Root Mean Squared Forecast Error (RMSFE) from the real-time now-casting exercise for the three specifications of the factor models, for the bridge equations and the SPF. The RMSFE is computed with respect to GDP in the final data vintage.[20] The factor model containing the complete data-set is labeled as "Benchmark," that including only GDP and monthly variables as "Monthly" and the small one with only hard data as "BCDC." For the bridge equations we report the RMSFE of the average forecast from the 23 equations (in Table A in the appendix we report the RMSFE for each single bridge equation model). The dots indicate the RMSFE of the SPF. On the horizontal axis we indicate the day and the month of the respective now-cast update.

Overall, the results confirm qualitatively the findings of earlier pseudo real-time exercises.

For example, as found in earlier work, Figure 4.1 shows that, as the information accumulates, the gains in forecast accuracy based on the factor model are substantial. In the next subsection, we show this point formally via a statistical test. Clearly, the ability of the model to incorporate increasingly richer information as time progresses is key for improving now-cast accuracy.

The SPF, published the second month of the quarter, has comparable performance to our model at that date. This confirms the results that model-based now-casts fair well in comparison to institutional or private sector forecasts (Giannone et al., 2008; de Winter, 2011) with real-time data.

The performance of the bridge equations is inferior, indicating sizable gains from adopting a joint multivariate modeling strategy. This result has been found in, e.g., Angelini et al. (2011), Angelini et al. (2010) or Rünstler et al. (2009).

Two other results emerge from the analysis. First, surveys appear to have an important role in improving the accuracy of the model at all points of time as the now-casts from the model based on only hard data are less accurate (on this point see also Giannone

[18] For example, retail sales vintages are available from June 2001 only. For the period January 1995–May 2001 we use the data figures as available in June 2001 vintage, but every time we produce a forecast we assume the publication lag of 14 days (cf. Table 4.1), i.e., on the 14th of each month we add the observation relating to the previous month.

[19] This can be an important feature as for many interesting new data sources (e.g., Google trends) only limited back-data is available. In addition, for many economies, even among headline indicators many have been collected only since recently.

[20] "Final" vintage corresponds to the data available in June 2011.

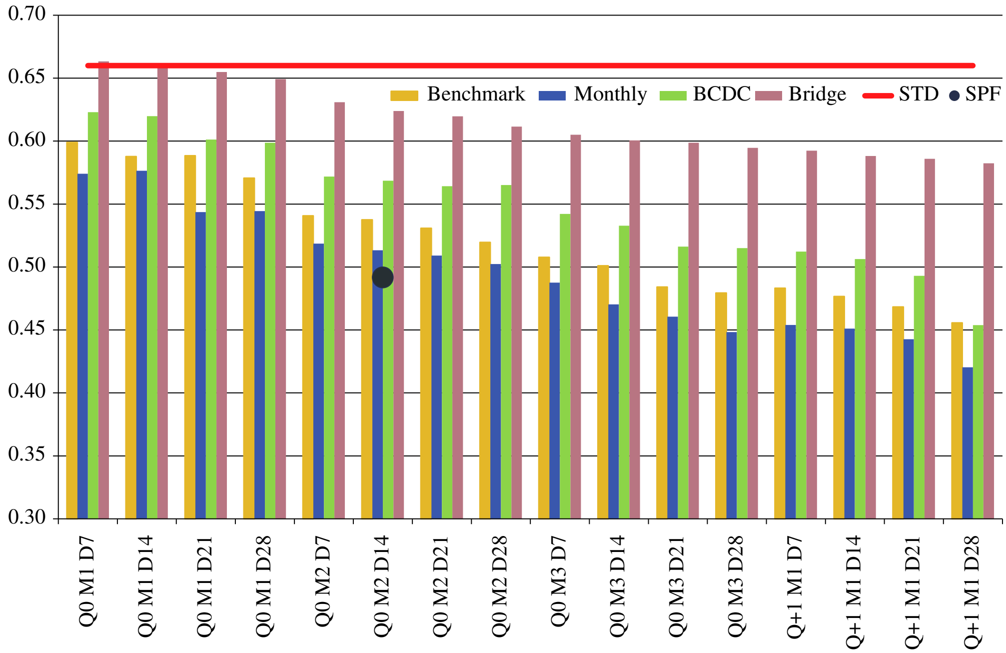

Figure 4.1 Root Mean Squared Forecast Error (RMSFE), GDP. *Notes:* Figure shows the Root Mean Squared Forecast Error (RMSFE) over period 1995–2010 for the model containing the complete data-set (Benchmark), the one that includes GDP and monthly variables only (Monthly), the one that uses the five hard variables considered by the NBER's Business Cycle Dating Committee (BCDC) and the RMSFEs of the average forecast produced by the bridge equations (Bridge) over the sample period 1995–2010. The forecast accuracy of the model is evaluated four times per month for four consecutive months (from the first month of the current quarter, Q0 M1 to the first month of the following quarter, Q+1 M1). The date of the update is indicated on the horizontal axis. The dot corresponds to the RMSFE for the survey professional forecasters (SPF).

et al., 2008; Bańbura and Rünstler, 2011; Lahiri and Monokroussos, 2011). Second, the inclusion of higher frequency data does not improve the results significantly.

Now-Casting After the Fall of Lehman Brothers

To further analyze the result that inclusion of daily and weekly variables does not improve now-cast accuracy, and as an illustration of how the sequence of news impacts the estimate of GDP and the associated uncertainty, we replicate a real-time now-cast from the benchmark specification of the GDP growth rate for the fourth quarter of 2008. Precisely, from October 1st 2008 until the end of January 2009, when the first official estimate of GDP was released, we produce daily updates of the now-cast, each time incorporating new data releases. This is a particularly interesting episode since it corresponds to the onset of the financial crisis following the bankruptcy of Lehman Brothers.

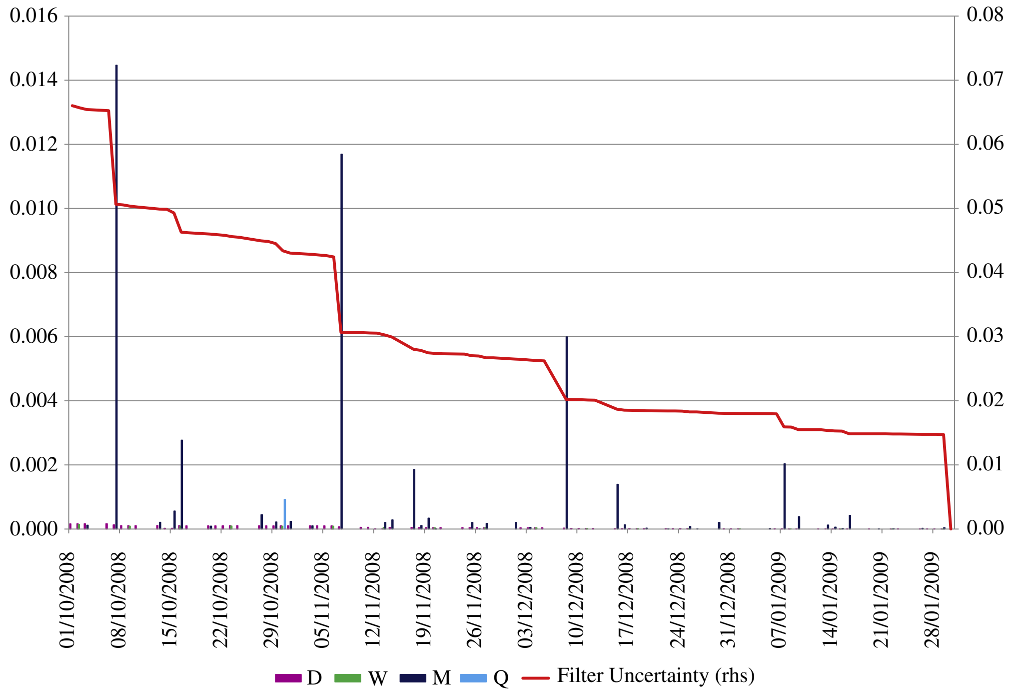

Figure 4.2 Filter uncertainty, GDP. *Notes:* Figure reports the evolution of filter uncertainty (RHS scale) and the contributions to its decline (LHS scale) for the GDP now-cast for fourth quarter of 2008. The data are grouped according to frequencies: daily ("D"), weekly ("W"), monthly ("M"), and quarterly ("Q"). The horizontal axis indicates the date of the now-cast update.

As explained in Section 2.3, the revisions of the GDP now-cast can be expressed as weighted sums of news from particular releases.[21] Similarly, the Kalman smoother output allows to decompose the gradual decline in filter uncertainty into contributions from particular (groups of) variables. For the purpose of the exercise we group the variables into the daily, weekly, monthly, and quarterly. As the decomposition is conditional on a given set of parameters, we produce all the updates using the parameters estimated over the sample 1983–2010.

Let us start with the filter uncertainty, see Figure 4.2. The chart indicates that macroeconomic monthly releases have the largest effects: the big spikes correspond to the release of the (monthly) employment situation report. Smaller spikes are experienced when industrial production is released. In contrast, daily and weekly data do not seem to have

[21] Recall that the decomposition does not take into account the effect of the revisions to back-data. Accordingly, the difference between two consecutive now-casts (based on vintages $\Omega_{\nu+1}$ and Ω_ν) in the exercise is equal to the sum of the effect of the news and of the effect of data revisions. The latter is defined as the difference between the now-cast based on the data availability pattern of vintage Ω_ν but with revised back-data as available in vintage $\Omega_{\nu+1}$ and the now-cast based on vintage Ω_ν.

much of an impact. To understand whether this result is explained by the fact that the effect of daily data is spread over time, we have computed the cumulative effect of daily and weekly data on the uncertainty from the first day of the current quarter to the day of GDP release. For that period total uncertainty is reduced from 0.066 to 0.015 (the day of the GDP release it collapses to zero) and 92% of it is due to the macro variables while the daily variables account only for the remaining 8%. This confirms the finding of a negligible role of high frequency variables, notably daily financial data. Finally, the impact of GDP release for the previous quarter is very small. This is explained by timeliness: once "early" information on the current quarter is released and incorporated in the model, information on the previous quarter GDP becomes redundant for the now-cast of the current quarter.

Figure 4.3 reports the evolution of the now-cast itself and the contribution of the news component of the various data groups to the now-cast revision.

Industrial production for September (published mid October) has the largest impact and leads to a substantial downward revision. This negative news in October is confirmed

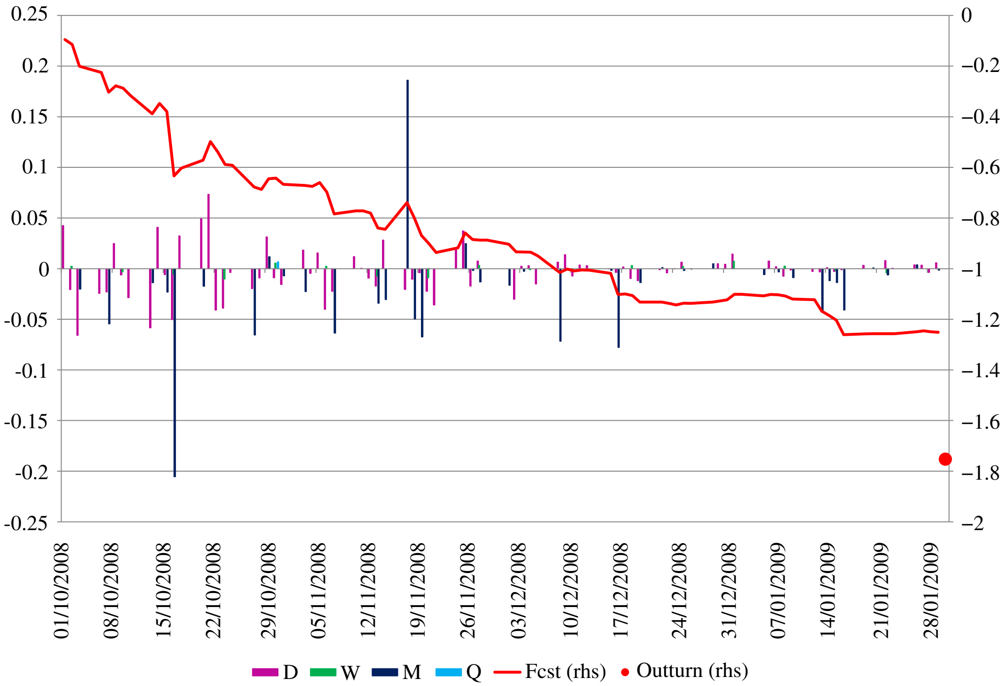

Figure 4.3 Contribution of news to now-cast revisions. *Notes:* Figure shows the evolution of the GDP now-cast (RHS scale) and how various data releases contribute to the now-cast revisions (LHS scale) for fourth quarter of 2008. The data are grouped according to frequencies: daily ("D"), weekly ("W"), monthly ("M"), and quarterly ("Q"). Dot indicates the outcome GDP growth (RHS scale, the latest vintage). The horizontal axis indicates the date of the now-cast update.

by subsequent data, both surveys and hard data. In fact, with all subsequent releases the tendency is for negative revisions.

News from daily financial variables are far from being small but they are volatile and lead to revisions in different directions. This explains the result of Figures 4.1 and 4.2. As will become clearer in the next section, it is the low frequency component of financial variables which is related to real economic activity while daily fluctuations might carry wrong signals.

Our results on the role of financial variables are in line with what has been found by Stock and Watson (2003) and Forni et al. (2003), who analyze the forecasting ability of financial variables for economic activity for longer horizon than ours and in models where they are aggregated at the monthly frequency and their timeliness is not exploited. A different conclusion is reached by Andreou et al. (2013) who find that financial variables have a sizeable impact on the now-cast accuracy of GDP.

With such different results obtained on the basis of different modeling approaches, more work is needed to reach a conclusion on this point. A difference between the application here and Andreou et al. (2013) is that the latter authors use a larger set of daily financial variables than the five we use.[22] However, results are hardly comparable since Andreou et al. (2013) treat monthly data such as surveys and the employment report, which we find very informative, as quarterly, thus disregarding their advantage in terms of timeliness.

Let us also observe that news from the weekly variables, unlike the daily financial news, are very small. Our conjecture for this finding is that initial jobless claims are rather noisy. This is line with the view of the NBER's Business Cycle Dating Committee, which does not use this series to determine the business cycle chronology (http://www.nber.org/cycles).

3.2.1. Does Information Help Improving Forecasting Accuracy? Monotonicity Tests

Figures 4.1–4.3 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), and we report the p-values for the "Benchmark" model, the models "Monthly" and "BCDC" as well as the average p-value of the bridge equations.[23]

[22] To reduce the dimensionality of the large panel they consider, they extract a small set of principal components and/or apply forecast combination methods.

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

Test 1: Monotonicity of the Forecast Errors

Let us define $\tilde{y}_t = y^{k_1}_{t,1}$ and $e_{t|\Omega_v} = \tilde{y}_t - \mathbb{E}\left[\tilde{y}_t|\Omega_v\right]$ 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 $v = 1, \ldots, V$.

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

The test is defined as follows:

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

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

Test 2: Monotonicity of the Mean Squared Forecast

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

The test is:

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

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

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

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

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: $\Delta^c_v = \mathbb{E}\left[\tilde{y}_{t|\Omega_v}\tilde{y}_t\right] - \mathbb{E}\left[\tilde{y}_{t|\Omega_{v+1}}\tilde{y}_t\right]$. The associated vector is defined as $\boldsymbol{\Delta}^c$ and the test is:

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

This test is closely related to the 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 4.2. Monotonicity cannot be rejected by any of the three tests confirming the visual evidence of Figures 4.1 and 4.2. The results for the individual bridge equations are provided in the appendix.

Table 4.2 Monotonicity Tests

	$\Delta^e \geq 0$	$\Delta^f \leq 0$	$\Delta^c \leq 0$
Benchmark	0.50	0.49	0.50
Monthly	0.50	0.50	0.50
BCDC	0.50	0.50	0.50
Bridge	0.50	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. For the bridge equations the table reports the average p-value.

3.3. 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 macroeconomic variables have become a popular tool to monitor business cycles conditions (see, e.g., Aruoba et al., 2009).[24] 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 now-cast).

Figure 4.4a plots this daily index (re-scaled to facilitate the comparison) against GDP growth and shows that it tracks GDP quite well. The index is based on the latest data vintage and, as in the case of the filter uncertainty and news, on the parameters estimated over the sample 1983–2010.

By appropriate filtering, this index can be aggregated to reflect quarterly growth rate[25] and we can then consider the projection of GDP on this quarterly aggregate (Figure 4.4b). 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.

3.4. Stock Prices

In this section we report some statistics on the degree of commonality of stock prices with macroeconomic variables.

[24] The Federal Reserve Bank of Philadelphia 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/.

[25] As the filter weights we use $\omega_i^{k_q,f}$, $i = 0, 1, \ldots, 2k_q - 2$, as explained in Section 2.

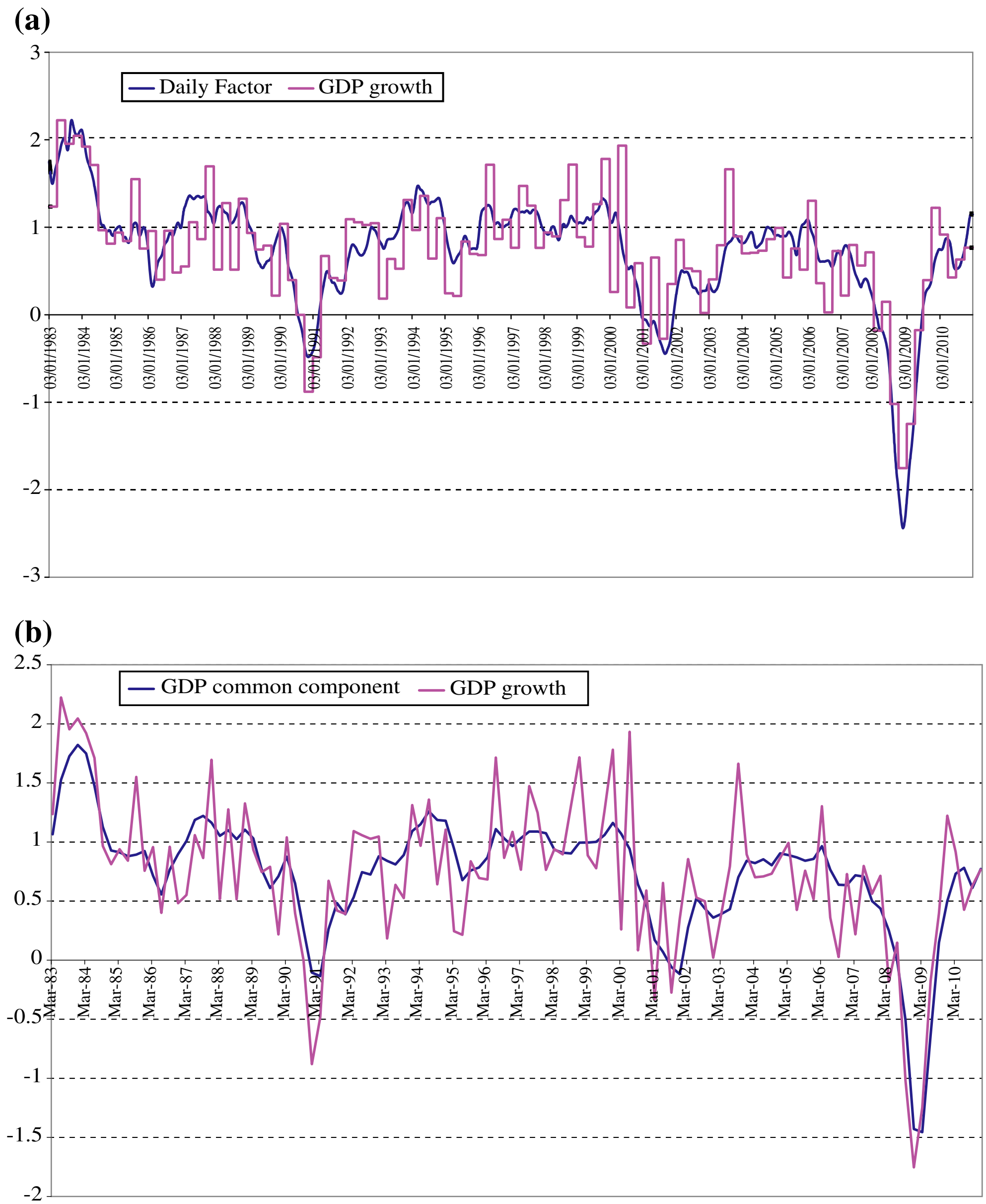

Figure 4.4 Daily factor, GDP and its common component. (a) Quarterly GDP growth and the daily factor. (b) Common component of GDP. *Notes:* The lower panel shows the GDP growth against its projection on the quarterly aggregation of the daily factor, i.e., against its common component.

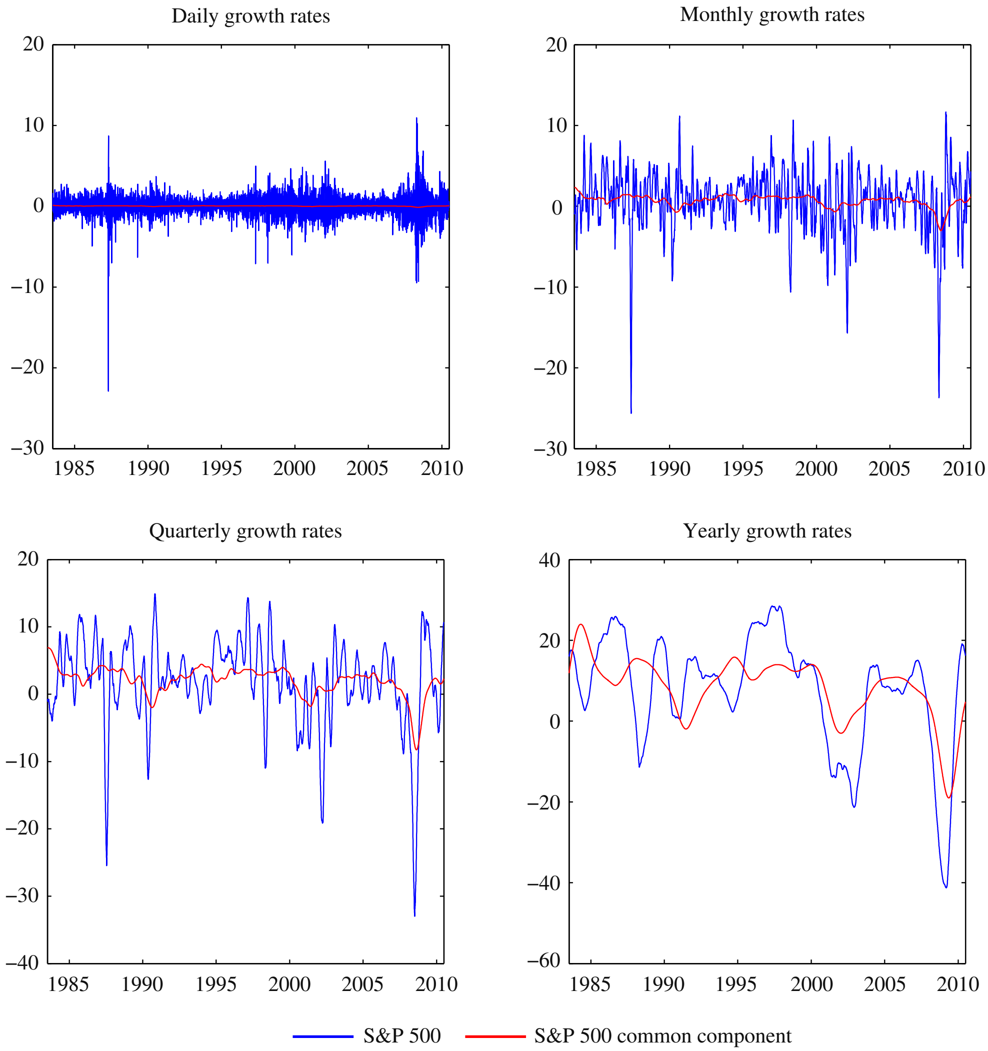

Figure 4.5 S&P 500 and its common component at different levels of time aggregation. *Notes:* Figure compares the S&P 500 and its common component at daily frequency filtered to match daily, monthly, quarterly, and yearly growth rates at the end of the respective periods.

First, we compute the common component of the S&P 500 index filtered to match daily, monthly, quarterly, and yearly growth rates at the end of the respective periods (Figure 4.5).[26] This is the signal of stock prices, extracting that part of their dynamics which is correlated with the macroeconomy. Figure 4.5 shows that the commonality is

[26] What we mean here by the monthly (quarterly and yearly) growth rate is the growth rate between the average last 22 (66 or 244) daily "level" observations and the average preceding 22 (66 or 244) observations. To obtain these growth rates we filter the daily growth rates using the weights $\omega_i^{k,f}$, $k = k_m$ ($k = k_q$ or $k = k_a$). Note that the resulting series is still at daily frequency.

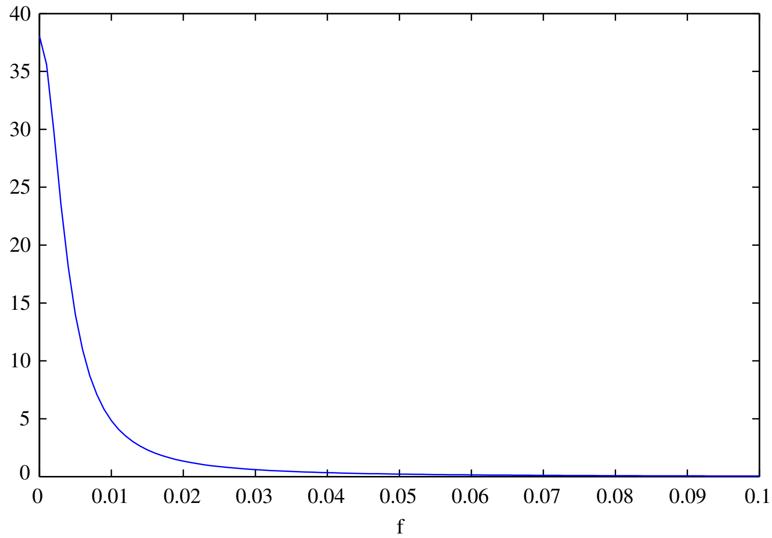

Figure 4.6 Spectral density ratio: S&P 500 vs its common component. *Notes:* Figure shows the ratio of spectral density of the common component to that of the S&P 500 of the series itself. *f* on the horizontal axis refers to the frequency.

not trivial when we consider longer term fluctuations, annual in particular, although the degree of commonality of the S&P 500 with the rest of the panel is less pronounced than what is estimated for GDP (Figure 4.4b). Clearly, although stock prices provide some false signals, they do go down in recessions.

Second, we compute the ratio of the spectral density of the common component to that of the series itself (at daily frequency).[27] Figure 4.6 shows that the bulk of commonality between the S&P 500 index and macroeconomic variables 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 (corresponding to the frequency $f = 2\pi/(8 * 264) = 0.003$) 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 periodicity lower than yearly ($f > 2\pi/264 = 0.024$), the ratio is below 2%.

Our model can be developed further to express the common component of stock prices in terms of news and model-based weights of those news. We leave this for further work.

[27] The spectral density is derived from the estimated parameters of the model.

4. CONCLUSIONS AND DISCUSSION ON FURTHER DEVELOPMENTS

In this chapter we have reviewed different approaches to now-casting, including traditional tools used in policy institutions (bridge equations).

A key point of our discussion is that the traditional approach is limited since it does not provide a framework for the reading of the flow of data as they become available throughout the quarter and for the evaluation of their impact on the updates of the now-cast.

We have distinguished between an approach which is based on the specification of a joint model for the target variable and for the predictors – which can therefore be used to derive model-based *news* associated to different data releases and to assess their impact on the now-cast – and what we call "partial" models based on single equations.

In the discussion and application we have stressed that the essential ideas of now-casting have been developed in two papers, Evans (2005) and Giannone et al. (2008), which propose a model in which the state space representation is exploited in order to produce sequences of now-casts in relation to the real-time data flow. We have then discussed advances and refinements of this approach as well as different solutions to technical aspects of the real-time analysis: mixed frequency, "jagged"/"ragged" edges, high-dimensionality. On the latter aspect we have referred to theoretical work on factor models for high-dimensional data and, in particular to Doz et al. (2011, 2012). The latter paper provides the theoretical justification for using maximum likelihood estimation for a "large" factor model which is the approach followed in the empirical application as proposed by Bańbura and Modugno (2010) and Modugno (2011) and which we consider the state of the art in this field.

The empirical analysis we have presented, based on a daily model and a real-time evaluation, confirms early results on the role of surveys (see Bańbura et al., 2011, for a recent review) and indicates that daily financial variables do not help improving the precision of the GDP now-cast. A formal test on the role of expanding information for improving the accuracy of the GDP now-cast shows that, as more data become available during the quarter, the precision indeed increases thereby confirming previous heuristic evidence on this point.

As a by-product of our methodology 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 the projection explains the bulk of GDP dynamics while it explains much less of daily fluctuations of stock prices. On the other hand, the index explains low frequency movements of stock prices indicating that financial variables are linked to macroeconomic fundamentals.

We have limited the empirical application to GDP and stock prices although the methodology can be applied to the now-casting of any of the variables included in the model. Liebermann (2012a), for example, provides a detailed study on the now-casting of a large variety of key monthly macroeconomic releases. The same framework has also been used for now-casting inflation (Modugno, 2011) and for now-casting the components of GDP (Angelini et al., 2010).

Our review has focused on what we regard as essential contributions to the now-casting literature and has omitted some important papers. For example, we have focused only on point now-cast. Now-cast densities in a state space framework could be obtained both via a frequentist (Aastveit et al., 2011) or Bayesian (Marcellino et al., 2012) approach. Further, Aastveit et al. (2011) consider now-cast combination where different classes of models, including VARs, factor models, bridge equations are combined. This approach is interesting but it is not obvious how to use it to extract the *news* and relate them to the sequence of now-cast updates, the feature that we have stressed here as being a key element of now-casting. Also, we have omitted a discussion of the so-called bottom-up approach that imposes national account identities for obtaining coherent now-cast for GDP ad its main components. Empirically, in terms of forecasting accuracy there are no major advantages of this approach with respect to our direct approach. However, the forecasts of components of GDP might be useful for economic interpretation and "story telling." We refer the reader to Angelini et al. (2010) for the adaptation of our model in this context.

Let us also stress that we have focused on now-casting rather than on the construction of synthetic indicators of real economic activity, the so-called diffusion indexes. Those indexes were introduced by Stock and Watson (1989) in small-scale models with only monthly indicators and estimated by maximum likelihood. In large-scale models they have been developed by Altissimo et al. (2001, 2010) and Chicago Fed (2001) using principal component analysis. Mariano and Murasawa (2003) extended the framework of Stock and Watson (1989) to include quarterly GDP as well as monthly variables and recently Aruoba et al. (2009) have also included weekly initial jobless claims using the same framework.

Although models to construct early estimates of GDP using selected monthly data have been around for a long time in policy institutions, now-casting as we have defined it here is a recent field for academic research and it is therefore likely to see new developments in many aspects. Beside research on modeling high frequency data, such as financial, which we have mentioned in the text, research on the potential value of other sources of information like google search volumes (see for example Vosen and Schmidt, 2011) is also interesting and we are sure will be the object of future investigation.

Another idea for further research is to link the high frequency now-casting framework with a quarterly structural model in a model coherent way. Giannone et al. (2009a) have suggested a solution and other developments are in progress. A by-product of this analysis is that one can obtain real-time estimates of variables such as the output gap or the natural rate of interest which can only be defined theoretically on the basis of a structural model.

As a final remark, let us stress that the now-casting models we have considered here are all linear and with constant parameters. Our choice is motivated by the fact that the empirical knowledge at present is mostly limited to this class of models. However, the events of the last few years have challenged the predictive power of all models, including now-casting. Extensions of our framework, incorporating, e.g., time-varying features, might address some of these challenges and we consider this a promising area for future research (see e.g. Carriero et al., 2012, and Guerin and Marcellino, 2013).

APPENDIX: DETAILS ON THE STATE SPACE REPRESENTATION AND ESTIMATION

Aggregator Variables

We explain how the aggregator variables can be recursively obtained so that

$$F_t^{k,\cdot} = \sum_{i=0}^{2k-2} \omega_i^{k,\cdot} F_{t-i}, \quad t = k, 2k, \ldots, \quad k = k_q, k_m, k_w.$$

As for the flow variables, the summation in $F_t^{k,f}$ goes over the current and previous observation interval, we also need auxiliary variables, $\bar{F}_t^{k,f}$. $F_t^{k,f}$ can be obtained recursively as follows:

$$\tilde{F}_t^{k,f} = \begin{pmatrix} F_t^{k,f} \\ \bar{F}_t^{k,f} \end{pmatrix} = \begin{cases} \begin{pmatrix} \bar{F}_{t-1}^{k,f} + \omega_{k-1}^{k,f} F_t \\ 0 \end{pmatrix}, & t = 1, k+1, 2k+1, \ldots, \\ \begin{pmatrix} F_{t-1}^{k,f} + \omega_{R(k-t,k)}^{k,f} F_t \\ \bar{F}_{t-1}^{k,f} + \omega_{R(k-t,k)+k}^{k,f} F_t \end{pmatrix}, & \text{otherwise,} \end{cases}$$

where $R(\cdot, k)$ denotes the positive remainder of the division by k (e.g., $R(-1, k) = k-1$). For the stock variables only single aggregator variable is necessary and we have:

$$F_t^{k,s} = \begin{cases} \omega_{k-1}^{k,s} F_t, & t = 1, k+1, 2k+1, \ldots, \\ F_{t-1}^{k,s} + \omega_{R(k-t,k)}^{k,s} F_t, & \text{otherwise.} \end{cases}$$

This can be implemented via the transition equation (7) with the following $\mathcal{W}_t^{k,\cdot}$ and $\mathcal{I}_t^{k,\cdot}$:

$$\mathcal{W}_t^{k,f} = \begin{cases} \begin{pmatrix} -\omega_{k-1}^{k,f} \\ 0 \end{pmatrix}, & t = 1, k+1, \ldots, \\ \begin{pmatrix} -\omega_{R(k-t,k)}^{k,f} \\ -\omega_{R(k-t,k)+k}^{k,f} \end{pmatrix}, & \text{otherwise,} \end{cases}$$

$$\mathcal{I}_t^{k,f} = \begin{cases} \begin{pmatrix} 0 & I_r \\ 0 & 0 \end{pmatrix}, & t = 1, k+1, \ldots, \\ I_{2r}, & \text{otherwise,} \end{cases}$$

$$\mathcal{W}_t^{k,s} = -\omega_{R(k-t,k)}^{k,s}, \quad \mathcal{I}_t^{k,s} = \begin{cases} 0, & t = 1, k+1, \ldots, \\ I_r, & \text{otherwise.} \end{cases}$$

Different Number of Days Per Period

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

$$z_t^k = \frac{k}{k_t} \sum_{i=0}^{k_t-1} z_{t-i}, \qquad t = k_1, k_1 + k_{k_1+1}, \ldots,$$

where k_t is the number of business days in the period (month or quarter) that contains day t and k is the average number of business days per period over the sample. This can be justified by the fact that data are typically working day adjusted. Consequently, $y_t^k = z_t^k - z_{t-k_t}^k$ becomes[28]

$$y_t^k = k \left(\sum_{i=0}^{k_t-1} \frac{i+1}{k_t} y_{t-i} + \sum_{i=k_t}^{k_t+k_{t-k_t}-2} \frac{k_t + k_{t-k_t} - i - 1}{k_{t-k_t}} y_{t-i} \right), \qquad t = k_1, k_1 + k_{k_1+1}, \ldots.$$

Hence we will have time-varying aggregation weights for the flow variables $\omega_{t,i}^{k,f} = k\frac{i+1}{k_t}$ for $i = 0, 1, \ldots, k_t - 1$, $\omega_{t,i}^{k,f} = k\frac{k_t+k_{t-k_t}-i-1}{k_{t-k_t}}$ for $i = k_t, k_t + 1, \ldots, k_t + k_{t-k_t} - 2$ and $\omega_{t,i}^{k,f} = 0$ otherwise. Formulas described above should be modified to reflect this.

Estimation by the EM Algorithm

We first explain the EM algorithm steps for a simpler state space representation, as given by (3) and (4), and then discuss how to modify the procedure for the more complicate representation (6) and (7).

For the state space representation (3) and (4) with one lag in the factor VAR we would have $\theta = (\mu, \Lambda, \Phi, \Sigma_E, \Sigma_U)$, where the only restriction is that Σ_E is diagonal. Let $T_v = \max_n T_v(n)$ denote the time index of the most recent observation in Ω_v. The log-likelihood can be written in terms of $Y = (Y_1, Y_2, \ldots, Y_{T_v})$ and $F = (F_1, F_2, \ldots, F_{T_v})$ as $l(Y, F; \theta)$. With some initial estimates of the parameters $\theta(0)$ the EM algorithm would proceed as follows:

$$\begin{aligned} &\text{E-step:} \quad L(\theta, \theta(j)) = \mathbb{E}_{\theta(j)}\left[l(Y, F; \theta)|\Omega_v\right], \\ &\text{M-step:} \quad \theta(j+1) = \arg\max_{\theta} L(\theta, \theta(j)). \end{aligned}$$

The new parameter estimates in the M-step can be obtained in two steps, first $\Lambda(j+1)$ and $\Phi(j+1)$ are given by:

$$\Lambda(j+1) = \left(\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[Y_t F_t' | \Omega_v\right] \right) \left(\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[F_t F_t' | \Omega_v\right] \right)^{-1}, \tag{11}$$

[28] If t is the last day of a period with k_t days then k_{t-k_t} refers to the number of days in the preceding period.

Table A Bridge Equations, RMSFEs and Monotonicity Tests

	IPTOT	PMI	DPRI	URATE	PYRLTOT	PCETOT	HSTAR	HSOLD	ORD	EXP	IMP	PHB
Q0 M1 D7	0.63	0.67	0.69	0.67	0.60	0.68	0.69	0.69	0.71	0.69	0.70	0.69
Q0 M1 D14	0.63	0.67	0.69	0.66	0.60	0.68	0.69	0.69	0.71	0.68	0.69	0.69
Q0 M1 D21	0.58	0.67	0.69	0.66	0.60	0.68	0.69	0.69	0.71	0.68	0.69	0.67
Q0 M1 D28	0.57	0.66	0.69	0.65	0.60	0.68	0.68	0.69	0.71	0.68	0.69	0.67
Q0 M2 D7	0.57	0.65	0.68	0.59	0.57	0.65	0.68	0.68	0.71	0.68	0.69	0.66
Q0 M2 D14	0.58	0.65	0.68	0.59	0.57	0.65	0.68	0.68	0.71	0.65	0.65	0.66
Q0 M2 D21	0.59	0.65	0.68	0.59	0.57	0.65	0.62	0.68	0.71	0.65	0.65	0.66
Q0 M2 D28	0.59	0.65	0.68	0.59	0.57	0.65	0.61	0.67	0.65	0.65	0.65	0.66
Q0 M3 D7	0.59	0.65	0.68	0.56	0.55	0.62	0.61	0.67	0.65	0.65	0.65	0.66
Q0 M3 D14	0.58	0.65	0.68	0.56	0.55	0.62	0.61	0.67	0.65	0.62	0.63	0.66
Q0 M3 D21	0.53	0.65	0.68	0.56	0.55	0.62	0.61	0.67	0.65	0.62	0.63	0.66
Q0 M3 D28	0.53	0.65	0.68	0.56	0.55	0.62	0.61	0.67	0.59	0.62	0.63	0.66
Q+1 M1 D7	0.53	0.65	0.69	0.55	0.54	0.59	0.61	0.67	0.58	0.62	0.63	0.66
Q+1 M1 D14	0.53	0.65	0.69	0.54	0.54	0.59	0.61	0.67	0.58	0.61	0.59	0.66
Q+1 M1 D21	0.50	0.65	0.69	0.54	0.54	0.59	0.59	0.67	0.58	0.61	0.59	0.66
Q+1 M1 D28	0.49	0.64	0.69	0.53	0.53	0.60	0.59	0.66	0.57	0.61	0.59	0.66

(Continued)

Table A Continued

	SALES	CONFCFB	PPIFG	CPITOT	BCC	ICSA	OIL	GS10	TB3	TWEXM	SP500
Q0 M1 D7	0.69	0.68	0.69	0.70	0.69	0.63	0.69	0.68	0.67	0.69	0.69
Q0 M1 D14	0.67	0.68	0.69	0.70	0.70	0.66	0.69	0.68	0.67	0.69	0.68
Q0 M1 D21	0.65	0.68	0.70	0.70	0.69	0.65	0.69	0.68	0.67	0.69	0.67
Q0 M1 D28	0.65	0.67	0.70	0.70	0.68	0.63	0.68	0.67	0.67	0.69	0.67
Q0 M2 D7	0.65	0.66	0.68	0.66	0.67	0.58	0.67	0.68	0.65	0.68	0.66
Q0 M2 D14	0.62	0.66	0.68	0.66	0.66	0.56	0.67	0.67	0.65	0.68	0.66
Q0 M2 D21	0.62	0.66	0.69	0.67	0.66	0.55	0.67	0.67	0.65	0.68	0.66
Q0 M2 D28	0.62	0.65	0.69	0.67	0.66	0.54	0.67	0.67	0.65	0.68	0.65
Q0 M3 D7	0.62	0.65	0.69	0.67	0.65	0.54	0.67	0.66	0.65	0.68	0.65
Q0 M3 D14	0.61	0.65	0.70	0.67	0.64	0.53	0.67	0.66	0.65	0.68	0.65
Q0 M3 D21	0.61	0.65	0.70	0.69	0.64	0.53	0.67	0.66	0.65	0.68	0.65
Q0 M3 D28	0.61	0.64	0.70	0.69	0.64	0.53	0.67	0.66	0.65	0.68	0.65
Q+1 M1 D7	0.61	0.64	0.70	0.69	0.64	0.53	0.67	0.66	0.65	0.68	0.65
Q+1 M1 D14	0.59	0.64	0.70	0.69	0.64	0.53	0.67	0.66	0.65	0.68	0.65
Q+1 M1 D21	0.59	0.64	0.70	0.69	0.64	0.53	0.67	0.66	0.65	0.68	0.65
Q+1 M1 D28	0.58	0.64	0.70	0.69	0.64	0.52	0.67	0.65	0.64	0.68	0.65

Notes: The upper part of the table reports the RMSFEs for each bridge equations over the sample period 1995–2010. The forecast accuracy of the model is evaluated four times per month (on the 7th, 14th, 21st, and 28th) for four consecutive months (from the first month of the current quarter, Q0 M1 to the first month of the following quarter, Q+1 M1).

$$\Phi(j+1) = \left(\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[F_t F_{t-1}' | \Omega_v\right]\right) \left(\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[F_{t-1} F_{t-1}' | \Omega_v\right]\right)^{-1}. \tag{12}$$

Second, given the new estimates of Λ and Φ, the covariance matrices can be obtained as follows:

$$\begin{aligned}\Sigma_E(j+1) &= \operatorname{diag}\left(\frac{1}{T_v}\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[\left(Y_t - \Lambda(j+1)F_t\right)\left(Y_t - \Lambda(j+1)F_t\right)' | \Omega_v\right]\right) \\ &= \operatorname{diag}\left(\frac{1}{T_v}\left(\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[Y_t Y_t' | \Omega_v\right] - \Lambda(j+1)\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[F_t Y_t' | \Omega_v\right]\right)\right)\end{aligned} \tag{13}$$

and

$$\Sigma_U(j+1) = \frac{1}{T}\left(\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[F_t F_t' | \Omega_v\right] - \Phi(j+1)\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[F_{t-1} F_t' | \Omega_v\right]\right), \tag{14}$$

see Watson and Engle (1983) and Shumway and Stoffer (1982). If Y_t did not contain missing observations ($\Omega_v = Y$), we would have that

$$\mathbb{E}_{\theta(j)}\left[Y_t Y_t' | \Omega_v\right] = Y_t Y_t' \quad \text{and} \quad \mathbb{E}_{\theta(j)}\left[Y_t F_t' | \Omega_v\right] = Y_t \mathbb{E}_{\theta(j)}\left[F_t' | \Omega_v\right],$$

which can be plugged to the formulas above. The expectations $\mathbb{E}_{\theta(j)}\left[F_t F_t' | \Omega_v\right]$, $\mathbb{E}_{\theta(j)}\left[F_t F_{t-1}' | \Omega_v\right]$ and $\mathbb{E}_{\theta(j)}\left[F_t | \Omega_v\right]$ can be obtained via the Kalman filter and smoother. When Y_t contains missing observations (11) and (13) become

$$\operatorname{vec}\left(\Lambda(j+1)\right) = \left(\sum_{t=1}^{T_v} \mathbb{E}_{\theta(j)}\left[F_t F_t' | \Omega_v\right] \otimes \mathcal{S}_t\right)^{-1} \operatorname{vec}\left(\sum_{t=1}^{T_v} \mathcal{S}_t Y_t \mathbb{E}_{\theta(j)}\left[F_t' | \Omega_v\right]\right) \tag{15}$$

and

$$\begin{aligned}\Sigma_E(j+1) = \operatorname{diag}\Bigg(\frac{1}{T_v}\sum_{t=1}^{T_v} \Big(&\mathcal{S}_t Y_t Y_t' \mathcal{S}_t' - \mathcal{S}_t Y_t \mathbb{E}_{\theta(j)}\left[F_t' | \Omega_v\right]\Lambda(j+1)'\mathcal{S}_t \\ &- \mathcal{S}_t \Lambda(j+1)\mathbb{E}_{\theta(j)}\left[F_t | \Omega_v\right] Y_t' \mathcal{S}_t + \mathcal{S}_t \Lambda(j+1)\mathbb{E}_{\theta(j)}\left[F_t F_t' | \Omega_v\right]\Lambda(j+1)'\mathcal{S}_t \\ &+ (I_N - \mathcal{S}_t)\Sigma_E(j)(I_N - \mathcal{S}_t)\Big)\Bigg),\end{aligned} \tag{16}$$

where $\mathcal{S}_t$ is a selection matrix, i.e., it is a diagonal matrix with ones corresponding to the non-missing observations in Y_t and zeros otherwise, see Bańbura and Modugno (2010).

For the daily model given by (6) and (7) the parameters are estimated using a modified version of the procedure just described. The necessary conditional expectations are provided by the Kalman filter and smoother applied to (6) and (7). $\Lambda^{\cdot,\cdot}$ are estimated blockwise, by frequency and by stock or flow type, using a formula similar to (15) in which Y_t and F_t are replaced by the appropriate block of $Y_t^{K_Y}$ and the corresponding aggregator variable, respectively. The estimate of the covariance matrix of $E_t^{K_Y}$ follows from (16) with F_t and Λ replaced by the entire state vector and the entire matrix of coefficients in the measurement equation (6) respectively. Finally the estimates for Φ and Σ_U follow from (12) and (14) respectively, by taking the elements from the conditional covariances of the state vector corresponding to F_t.

Computation of the News

To compute the news we need the conditional covariances of X_t and X_{t-i}:

$$\mathbb{E}\Big[\left(X_t - \mathbb{E}\left[X_t|\Omega_v\right]\right)\left(X_{t-i} - \mathbb{E}\left[X_{t-i}|\Omega_v\right]\right)'\Big].$$

One way to obtain them is to extend the state vector by its i lags. However, if i is large this could lead to a prohibitively large state vector. Instead we use the direct formulas for the covariances by De Jong and Mackinnon (1988).

ACKNOWLEDGMENTS

The authors would like to thank the Editors, Graham Elliott and Allan Timmermann, and the Referees for their suggestions. This research is supported by the IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy).

REFERENCES

Aastveit, K., Trovik, T., 2012. Nowcasting norwegian GDP: the role of asset prices in a small open economy. Empirical Economics 42 (1), 95–119.

Aastveit, K.A., Gerdrup, K.R., Jore, A.S., Thorsrud, L.A., 2011. Nowcasting GDP in Real-Time: A Density Combination Approach. Working Paper 2011/11, Norges Bank.

Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Hallin, M., Lippi, M., Reichlin, L., Veronese, G., 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., Cristadoro, R., Forni, M., Lippi, M., Veronese, G., 2010. New eurocoin: tracking economic growth in real time. The Review of Economics and Statistics 92 (4), 1024–1034.

Andreou, E., Ghysels, E., Kourtellos, A., 2013. Should macroeconomic forecasters look at daily financial data? Journal of Business & Economic Statistics 31 (2), 240–251.

Angelini, E., Bańbura, M., Rünstler, G., 2010. Estimating and forecasting the euro area monthly national accounts from a dynamic factor model. OECD Journal: Journal of Business Cycle Measurement and Analysis 2010 (1), 7.

Angelini, E., Camba-Méndez, G., Giannone, D., Reichlin, L., Rünstler, G., 2011. Short-term forecasts of euro area GDP growth. Econometrics Journal 14 (1), C25–C44.

Arnostova, K., Havrlant, D., Ruzicka, L., Toth, P., 2011. Short-term forecasting of Czech quarterly GDP using monthly indicators. Czech Journal of Economics and Finance 61 (6), 566–583.

Aruoba, S., Diebold, F.X., Scotti, C., 2009. Real-time measurement of business conditions. Journal of Business and Economic Statistics 27 (4), 417–427.

Baffigi, A., Golinelli, R., Parigi, G., 2004. Bridge models to forecast the euro area GDP. International Journal of Forecasting 20 (3), 447–460.
Bai, J., Ng, S., 2008. Forecasting economic time series using targeted predictors. Journal of Econometrics 146 (2), 304–317.
Bańbura, M., Modugno, M., 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., Rünstler, G., 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.
Bańbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian VARs. Journal of Applied Econometrics 25 (1), 71–92.
Bańbura, M., Giannone, D., Reichlin, L., 2011. Nowcasting. In: Clements, M.P., Hendry, D.F. (Eds.), Oxford Handbook on Economic Forecasting. Oxford University Press, pp. 63–90.
Barhoumi, K., Darn, O., Ferrara, L., 2010. Are disaggregate data useful for factor analysis in forecasting French GDP? Journal of Forecasting 29 (1–2), 132–144.
Bessec, M., Doz, C., Prevision de court terme de la croissance du PIB francais laide de modeles facteurs dynamiques. Documents de Travail Numéro 2011/01 de la DG Tresor 1, France, Tresor Direction Gnral, Economie et Prevision.
Bloor, C., Matheson, T., 2011. Real-time conditional forecasts with Bayesian VARs: an application to New Zealand. The North American Journal of Economics and Finance 22 (1), 26–42.
Boivin, J., Ng, S., 2006. Are more data always better for factor analysis? Journal of Econometrics 132 (1), 169–194.
Camacho, M., Perez-Quiros, G., 2010. Introducing the euro-sting: short term indicator of euro area growth. Journal of Applied Econometrics 25 (4), 663–694.
Carriero, A., Clark, T.E., Marcellino, M., 2012. Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility. Working Paper 1227, Federal Reserve Bank of Cleveland.
Chen, W., Anderson, B.D., Deistler, M., Filler, A., 2012. Properties of Blocked Linear Systems. Discussion Paper, Vienna University of Technology, Manuscript.
Chicago Fed, 2001. CFNAI Background Release. Discussion Paper. <http://www.chicagofed.org/economicresearchanddata/national/pdffiles/CFNAI bga.pdf>.
Clements, M.P., Galvão, A.B., 2008. Macroeconomic forecasting with mixed-frequency data. Journal of Business and Economic Statistics 26, 546–554.
Clements, M.P., Galvão, A.B., 2009. Forecasting US output growth using leading indicators: an appraisal using MIDAS models. Journal of Applied Econometrics 24 (7), 1187–1206.
Croushore, D., 2011. Frontiers of real-time data analysis. Journal of Economic Literature 49 (1), 72–100.
Curdia, V., Del Negro, M., Greenwald, D., 2012. Rare Shocks. Great Recessions. Discussion Paper, Federal Reserve Bank of New York, Manuscript.
D'Agostino, A., McQuinn, K., O'Brien, D., 2008. Now-casting Irish GDP. Research Technical Papers 9/RT/08, Central Bank & Financial Services Authority of Ireland (CBFSAI).
de Antonio Liedo, D., Muoz, E.F., 2010. Nowcasting Spanish GDP Growth in Real Time: "One and A Half Months Earlier". Banco de Espaa Working Papers 1037, Banco de Espaa.
De Jong, P., Mackinnon, M.J., 1988. Covariances for smoothed estimates in state space models. Biometrika 75 (3), 601–602.
De Mol, C., Giannone, D., Reichlin, L., 2008. Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components? Journal of Econometrics 146 (2), 318–328.
de Winter, J., 2011. Forecasting GDP Growth in Times of Crisis: Private Sector Forecasts Versus Statistical Models. DNB Working Papers 320, Netherlands Central Bank, Research Department.
Diron, M., 2008. Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data. Journal of Forecasting 27 (5), 371–390.
Doz, C., Giannone, D., Reichlin, L., 2011. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. Journal of Econometrics 164 (1), 188–205.
Doz, C., Giannone, D., Reichlin, L., 2012. A maximum likelihood approach for large approximate dynamic factor models. Review of Economics and Statistics 94 (4), 1014–1024.

Drechsel, K., Scheufele, R., 2012. A Comparison of Bottom-Up Approaches and Direct Forecasts of German GDP in a Data-Rich Environment. Discussion Paper, Halle Institute for Economic Research (IWH), Manuscript.

ECB, 2008. Short-term forecasts of economic activity in the euro area. In: Monthly Bulletin. European Central Bank, April, pp. 69–74.

Evans, M.D.D., 2005. Where are we now? Real-time estimates of the macroeconomy. International Journal of Central Banking 1 (2).

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

Forni, M., Giannone, D., Lippi, M., Reichlin, L., 2009. Opening the black box: structural factor models with large cross sections. Econometric Theory 25 (05), 1319–1347.

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

Ghysels, E., Wright, J.H., 2009. Forecasting professional forecasters. Journal of Business and Economics Statistics 27 (4), 504–516.

Ghysels, E., Santa-Clara, P., Sinko, A., Valkanov, R., 2003. MIDAS Regressions: Further Results and New Directions. Working Paper, UNC and UCLA.

Giannone, D., Reichlin, L., Sala, L., 2004. Monetary policy in real time. In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual. MIT Press, pp. 161–200.

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

Giannone, D., Monti, F., Reichlin, L., 2009a. Incorporating conjunctural analysis in structural models. In: Wieland, V. (Ed.), The Science and Practice of Monetary Policy Today. Springer, Berlin, pp. 41–57.

Giannone, D., Reichlin, L., Simonelli, S., 2009b. Nowcasting euro area economic activity in real time: the role of confidence indicators. National Institute Economic Review 210, 90–97.

Giannone, D., Lenza, M., Primiceri, G.E., 2012. Prior Selection for Vector Autoregressions. CEPR Discussion Papers 8755, C.E.P.R. Discussion Papers.

Guerin, P., Marcellino, M., 2013. Markov-switching MIDAS models. Journal of Business & Economic Statistics 31 (1), 45–56.

Hahn, E., Skudelny, F., 2008. Early Estimates of Euro Area Real GDP Growth – A Bottom Up Approach from the Production Side. Working Paper Series 975, European Central Bank.

Harvey, A., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Karlsson, S., 2013. Forecasting with Bayesian VAR models. In: Graham Elliott, A.T. (Ed.), Handbook of Economic Forecasting. North Holland.

Kitchen, J., Monaco, R.M., 2003. Real-time forecasting in practice: the US treasury Staff's real-time GDP forecast system. Business Economics, 10–19.

Koop, G., 2013. Forecasting with Medium and Large Bayesian VARs. Working Papers 1117, University of Strathclyde Business School, Department of Economics, Journal of Applied Econometrics 28 (2), 177—203.

Kuzin, V., Marcellino, M., Schumacher, C., 2011. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. International Journal of Forecasting 27 (2), 529–542.

Kuzin, V., Marcellino, M., Schumacher, C., 2013. Pooling Versus Model Selection For Nowcasting GDP With Many Predictors: Empirical Evidence For Six Industrialized Countries. Journal of Applied Econometrics 28 (3), 392–441.

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

Liebermann, J., 2012a. Real-Time Forecasting in a Data-Rich Environment. Discussion Paper 39452, MPRA Paper.

Liebermann, J., 2012b. Short-term forecasting of quarterly gross domestic product growth. Central Bank of Ireland Quarterly Bulletin 3, 74–84.

Marcellino, M., Schumacher, C., 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.

Marcellino, M., Porqueddu, M., Venditti, F., 2012. Short-Term GDP Forecasting with a Mixed Frequency Dynamic Factor Model with Stochastic Volatility. Discussion Paper, European University Institute, Manuscript.

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

Mariano, R.S., Murasawa, Y., 2010. A coincident index, common factors, and monthly real GDP. Oxford Bulletin of Economics and Statistics 72 (1), 27–46.

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.

Matheson, T., 2011. New Indicators for Tracking Growth in Real Time. IMF Working Papers 11/43, International Monetary Fund.

McCracken, M.W., Sekhposyan, T., Owyang, M., 2013. Real-Time Forecasting with a Large, Mixed Frequency, Bayesian VAR. Manuscript.

Mittnik, S., Zadrozny, P.A., 2004. Forecasting Quarterly German GDP at Monthly Intervals Using Monthly IFO Business Conditions Data. CESifo Working Paper Series 1203, CESifo Group Munich.

Modugno, M., 2011. Nowcasting Inflation Using High Frequency Data. International Journal of Forecasting.

Ouysse, R., 2011. Comparison of Bayesian moving Average and Principal Component Forecast for Large Dimensional Factor Models. Discussion Papers 2012-03, The University of New South Wales, School of Economics.

Parigi, G., Golinelli, R., 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., Schlitzer, G., 1995. Quarterly forecasts of the Italian business cycle by means of monthly indicators. Journal of Forecasting 14 (2), 117–141

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

Proietti, T., 2011. Estimation of common factors under cross-sectional and temporal aggregation constraints. International Statistical Review 79 (3), 455–476.

Proietti, T., Moauro, F., 2006. Dynamic factor analysis with non-linear temporal aggregation constraints. Journal of the Royal Statistical Society Series C 55 (2), 281–300.

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

Sargent, T.J., Sims, C.A., 1977. Business cycle modeling without pretending to have too much a-priori economic theory. In: Sims, C. (Eds.), New Methods in Business Cycle Research. Federal Reserve Bank of Minneapolis.

Seong, B., Ahn, S.K., Zadrozny, P.A., 2007. Cointegration Analysis with Mixed-Frequency Data. CESifo Working Paper 1939, CESifo Group Munich.

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

Siliverstovs, B., Kholodilin, K.A., 2012. Assessing the real-time informational content of macroeconomic data releases for now-/forecasting GDP: evidence for Switzerland. Journal of Economics and Statistics/Jahrbücher für Nationalökonomie und Statistik 232 (4), 429–444.

Siliverstovs, B., Kholodilin, K.A., 2012. Assessing the Real-Time Informational Content of Macroeconomic Data Releases for Now-/Forecasting GDP: Evidence for Switzerland. KOF Working Papers 10-251, KOF Swiss Economic Institute, ETH Zurich.

Stock, J.H., Watson, M.W., 1989. New Indexes of Coincident and Leading Economic Indicators. NBER Macroeconomics Annual 1989, vol. 4, NBER Chapters. National Bureau of Economic Research, Inc., pp. 351–409.

Stock, J.H., Watson, M.W., 2003. Forecasting output and inflation: the role of asset prices. Journal of Economic Literature 41 (3), 788–829.

Stock, J.H., Watson, M.W., 2011. Dynamic factor models. In: Clements, M.P., Hendry, D.F. (Eds.), Oxford Handbook on Economic Forecasting. Oxford University Press.

Stock, J.H., Watson, M.W., 2012. Generalized shrinkage methods for forecasting using many predictors. Journal of Business and Economics Statistics 30 (4), 481–493.

Vosen, S., Schmidt, T., 2011. Forecasting private consumption: survey-based indicators vs. Google trends. Journal of Forecasting 30 (6), 565–578.
Watson, M.W., 2004. Comment on Giannone, Reichlin, and Sala. In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual. MIT Press, pp. 216–221.
Watson, M.W., Engle, R.F., 1983. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. Journal of Econometrics 23, 385–400.
Wolak, F.A., 1987. An exact test for multiple inequality and equality constraints in the linear regression model. Journal of the American Statistical Association 82, 782–793.
Wolak, F.A., 1989. Testing inequality constraints in linear econometric models. Journal of Econometrics 31, 205–235.
Yiu, M.S., Chow, K.K., 2010. Nowcasting Chinese GDP: information content of economic and financial data. China Economic Journal 3 (3), 223–240.
Zadrozny, P., 1990. Estimating a Multivariate ARMA Model with Mixed-Frequency Data: An Application to Forecating US GNP at Monthly Intervals. Working Paper Series 90-6, Federal Reserve Bank of Atlanta.