# Real-Time Forecast Density Combinations*

## Forecasting US GDP Growth Using Mixed-Frequency Data

Thomas B. Götz Alain Hecq†
Jean-Pierre Urbain

Maastricht University, SBE, Department of Quantitative Economics

April 16, 2012

**Abstract**

We combine the issues of dealing with variables sampled at mixed frequencies and the use of real-time data. In particular, the repeated observations forecasting (ROF) analysis of Stark and Croushore (2002) is extended to an autoregressive distributed lag setting in which the regressors may be sampled at higher frequencies than the regressand. For the US GDP quarterly growth rate, we compare the forecasting performances of an AR model with several mixed-frequency models among which the MIDAS approach. The additional dimension provided by different vintages allows us to compute several forecasts for a given calendar date and use them to construct forecast densities. Scoring rules are employed to test for their equality and to construct combinations of them. Given the change of the implied weights over time, we propose time-varying ROF-based weights using vintage data which present an alternative to traditional weighting schemes.

**JEL Codes:** C32, C53

**Keywords:** Real-Time Data, MIDAS, Repeated Observations Forecasting, Scoring Rule, Data Revision

*We thank Franz Palm and the participants of the CIRANO workshop on Data Revision in Macroeconomic Forecasting and Policy, October 7-8, 2011, in Montreal for useful suggestions and comments on earlier versions of the paper.

†Corresponding author: Alain Hecq, Department of Quantitative Economics of the School of Business and Economics at Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: a.hecq@maastrichtuniversity.nl, Tel.: +31 43 388 3798, Fax: +31 43 388 2000.

# 1 Introduction

Economic time series released by statistical offices and other providers of data are available in mixed frequencies. As an example, one can simultaneously find intra-day stock prices and quarterly national account aggregates. In econometrics, there are two main solutions to this problem, first, temporally aggregating the higher-frequency variables into low-frequency ones (averaging for flow and skip-sampling for stock data) and, second, considering a mixed-frequency model. The former approach, however, might lead to a loss of information due to the deletion of high-frequency observations. Hence, it is reasonable to believe that the forecasting performance of a low-frequency series might be improved by making use of the additional information contained in the higher frequency variables by considering the latter solution. Among the different approaches that have been developed, we consider the MIDAS framework (see Ghysels, Santa-Clara and Valkanov 2004) in this paper, a method that allows to integrate series sampled at different frequencies in a parsimonious manner.

Stark and Croushore (2002) analyzed another issue researchers face in practice, i.e. the use of real-time versus latest-available data. To be more precise, the authors argue that in the literature a new model is often developed based on the results it achieves employing a data set which is different from the one researchers could use in real time (Croushore and Stark 2001). Furthermore, Stark and Croushore (2002) investigate how model selection, i.e. the lag order in autoregressive models, changes as different data releases, or vintages, are concerned. Interestingly, measures of forecast accuracy may be deceptively lower when latest-available data are dealt with than when real-time data are used (Stark and Croushore 2002). For an overview of the existing literature on real-time data analysis we refer the reader to Croushore (2011).

In this paper we consider both issues introduced above. In particular, we apply and extend the repeated observation forecasting (ROF) approach proposed by Stark and Croushore (2002) and examined by Croushore (2006) by considering an autoregressive distributed lag (ADL) setting in which the regressors may be sampled at higher frequencies than the regressand. As such, a wider range of approaches may be considered than the ARIMA models of Stark and Croushore (2002) giving rise to the question whether a certain model uniformly outperforms other approaches across different vintages. Given a calendar date, the use of each vintage leads to a different forecast of the variable of interest enabling us to produce a forecast density. Comparisons of the forecast densities produced by the different forecasting methods is then undertaking employing a Diebold-Mariano (1995)-type test for the equality of forecast densities. In particular, we extend this approach by applying it to density forecasts produced by different vintages as opposed to being generated by "time series models, for which parameters are estimated with a moving window of [...] observations" (Diks, Panchenko, Sokolinskiy and van Dijk 2011a).

As opposed to choosing a single model, a combination of the models we investigate is considered. As discussed widely in the literature, combining the information present in all models may lead to diversification gains (Bates and Granger 1969, Stock and Watson 2004 or Timmermann 2006). Just as this argument applies to point forecasting, it may do so for the combination of forecast densities, see e.g. Aastveit, Gerdrup, Jore and Thorsud (2011).

We will investigate whether such a combination improves the forecasting performance and if so, how the composition (i.e. weights) of the forecast densities alters through time. Given the change of the weights across different calendar dates, we propose time-varying ROF-based weights which present an alternative to traditional weighting schemes such as the one of Bates and Granger (1969). We illustrate our approach in an empirical analysis where we aim at fore-, and later, nowcasting the US GDP quarterly growth rate. Aastveit et al. (2011) combine density nowcasts for GDP growth from different model classes using real-time vintage data and is therefore probably closest related to our study.

The rest of the paper is organized as follows. In Section 2 we give the notation for a mixed-frequency real-time data analysis. Section 3 summarizes the repeated observation forecasting and nowcasting approaches for our particular framework. Section 4 describes the methodology to compare and combine density forecasts. In Section 5 the models we use for now- or forecasting are presented. Section 6 describes the series and summarizes the outcomes of the empirical analysis. Section 7 concludes.

## 2 Notations for mixed-frequency real-time data

We observe $N$ regularly spaced vintages available for a maximum of $T$ and $T \times m_j$ $(j = 1, \ldots, K)$ observations for the low frequency variable $y$ and the high-frequency variables $x_1, \ldots, x_K$, respectively. In order to avoid confusions in the terminology, in this article, a vintage is a moment at which a series is published, for instance June 2011. The index $t$ represents the low-frequency series and runs from 1 to $T$. For a high-frequency regressor $x_j$, the number of high-frequency observations per $t$-period equals $m_j$. In a quarter/months example, $m_j = 3$; in a month/(working) days example $m_j = 20$ or 22.

Assuming for instance a publication lag of one period, $y_{t-1}^t$ denotes the figure published at time $t$ of the value of $y$ for time $t-1$. The first difference operator $\Delta = (1-L)$ runs over both indexes with $\Delta y_{t-1}^t = (1-L)y_{t-1}^t = y_{t-1}^t - y_{t-2}^{t-1}$. This notational convention is used for instance in Hecq and Jacobs (2009).

The notation becomes more evolved for the high-frequency variables, $x_1, \ldots, x_K$. For now, the $j$-index is suppressed for explanatory convenience. We use $x_{t-1,m-i}^{t,m-g}$ with an additional index $m-i$ to refer to the high-frequency observation within each $t$-period, where for $m-i$, $i$ satisfies $0 \leq i \leq m-1$. If $i = 0$, the $m$-index is often suppressed such that $x_{t-1,m}^{t,m-g} \equiv x_{t-1}^{t,m-g}$ meaning it is the end-of-period observation such as the last day in a month. $x_{t-1,1}^{t,m-g}$ is the first day of the month in a month/day analysis or the first month of the quarter in a quarter/month setting. Finally, $x_{t-1,m-m}^{t,m-g} \equiv x_{t-2}^{t,m-g}$. For each $x$ the superscript refers to the vintage just as for the low-frequency variable. In this paper, vintages for the regressors may appear at a higher frequency than for the regressand such that, in a quarter/month example, $x_{2012Q1,2}^{2012Q1,3}$ denotes the figure published in March 2012 of the value of $x$ for February 2012. So, the same rules as for $i$ apply to $g$, i.e. $0 \leq g \leq m-1$ and so on.

Table 1 illustrates this notation using a quarter/month example for a close-up corresponding to the last three observed quarters. To simplify, we assume the low-frequency (quarterly)

variable to be revised on a low-frequency basis noting, however, that an extension towards high-frequency vintages (e.g. monthly) is easily applicable leading to a "rectangular" shape of the table as far as the regressand is concerned. Additionally, a publication lag of one period is assumed for both, the regressor and the regressand, such that the table possesses a triangular shape on a monthly basis for $x$ and on a quarterly basis for $y$. Note also that a nowcast is assumed to be computed in the last month of a quarter. However, a nowcast might as well be obtained during the first or second month as explained in a later section.

# 3 Repeated observations forecasting approaches

A common practice in empirical work is to use the last available time series to evaluate forecasts. This means that at period $T$, June 2011 say, one collects the historical time series for $y_{t-1}^{T}$ where $t = 2, \ldots, T$ assuming a publication lag of one period. Subsequently, a one-step ahead point forecast for $\hat{y}_T$ might be obtained. Stark and Croushore (2002) have proposed an interesting tool, which they call repeated observations forecasting (ROF hereafter), to evaluate whether the forecast accuracy is sensitive to the vintage chosen. Roughly speaking, they take one particular calendar date $t^*$ and look at the historical series $y_{t^*-1}^{v}$ for a set of vintages $v = 1, \ldots, V$. Finally, they have a sequence of $V$ (in this case, one step-ahead) forecasts for the same point $\hat{y}_{t^*}^{v}$, $v = 1, \ldots, V$, which they report in a graph.

However, Stark and Croushore (2002) exclusively use ARIMA models for that exercise. Matters become more complicated when we want to use explanatory variables which are possibly available for different vintages and with a different frequency of observations. We study three approaches whereby most attention is given to the first one. These are (1) the ROF à la Stark and Croushore (2002) employing forecast densities, (2) a multivariate dimension ROF and (3) nowcasting of the first diagonal release $y_t^t$, $t = 1, \ldots, T$.

The ROF (1) is basically similar to the one in Stark and Croushore (2002) with the particularity that we consider mixed-frequency models employing exogenous variables $x$. Note that an extension towards $h$-step-ahead (direct or indirect) forecasts is straightforwardly done such that the methodology presented later applies in such a setting as well. Nevertheless, in order to focus on the methodology of obtaining weights based on forecast densities constructed by different models, we disregard any additional vintage-dimension that might arise due to the use of high-frequency variables. This means that, for instance, in quarter $t-3$ of Table 1, we take the low- and high-frequency variables at the same time release, i.e. at the end-of-quarter vintages,

$$
\begin{aligned}
t-2 \quad &: \quad y_{t-3}^{t-2} = f(x_{t-3,m-1}^{t-2,m}, x_{t-3,m-2}^{t-2,m}, y_{t-4}^{t-2}, x_{t-4,m}^{t-2,m}, x_{t-4,m-1}^{t-2,m}, x_{t-4,m-2}^{t-2,m}, y_{t-5}^{t-2}, ..) \\
t-1 \quad &: \quad y_{t-3}^{t-1} = f(x_{t-3,m-1}^{t-1,m}, x_{t-3,m-2}^{t-1,m}, y_{t-4}^{t-1}, x_{t-4,m}^{t-1,m}, x_{t-4,m-1}^{t-1,m}, x_{t-4,m-2}^{t-1,m}, y_{t-5}^{t-1}, ..) \\
t \quad &: \quad y_{t-3}^{t} = f(x_{t-3,m-1}^{t,m}, x_{t-3,m-2}^{t,m}, y_{t-4}^{t}, x_{t-4,m}^{t,m}, x_{t-4,m-1}^{t,m}, x_{t-4,m-2}^{t,m}, y_{t-5}^{t}, ..).
\end{aligned}
$$

in order to obtain $\hat{y}_{t-3}^{t-2}, \hat{y}_{t-3}^{t-1}$ and $\hat{y}_{t-3}^{t}$ from which to compute a forecast density. How to precisely

Table 1: Real-time data mixed-frequency matrix

| $\frac{Vintages}{Calendar\ time}$ | $\dots$ | $t-2,m-2$ | $t-2,m-1$ | $t-2,m$ | $t-1,m-2$ | $t-1,m-1$ | $t-1,m$ | $t,m-2$ | $t,m-1$ | $t,m$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $t-3,m-2$ | $\dots$ | $x_{t-3,m-2}^{t-2,m-2}$ | $x_{t-3,m-2}^{t-2,m-1}$ | $x_{t-3,m-2}^{t-2,m}$ | $x_{t-3,m-2}^{t-1,m-2}$ | $x_{t-3,m-2}^{t-1,m-1}$ | $x_{t-3,m-2}^{t-1,m}$ | $x_{t-3,m-2}^{t,m-2}$ | $x_{t-3,m-2}^{t,m-1}$ | $x_{t-3,m-2}^{t,m}$ |
| $t-3,m-1$ | $\dots$ | $x_{t-3,m-1}^{t-2,m-2}$ | $x_{t-3,m-1}^{t-2,m-1}$ | $x_{t-3,m-1}^{t-2,m}$ | $x_{t-3,m-1}^{t-1,m-2}$ | $x_{t-3,m-1}^{t-1,m-1}$ | $x_{t-3,m-1}^{t-1,m}$ | $x_{t-3,m-1}^{t,m-2}$ | $x_{t-3,m-1}^{t,m-1}$ | $x_{t-3,m-1}^{t,m}$ |
| $t-3,m$ | $\dots$ | $x_{t-3,m}^{t-2,m-2}$ | $x_{t-3,m}^{t-2,m-1}$ | $x_{t-3,m}^{t-2,m};y_{t-3}^{t-2}$ | $x_{t-3,m}^{t-1,m-2}$ | $x_{t-3,m}^{t-1,m-1}$ | $x_{t-3,m}^{t-1,m};y_{t-3}^{t-1}$ | $x_{t-3,m}^{t,m-2}$ | $x_{t-3,m}^{t,m-1}$ | $x_{t-3,m}^{t,m};y_{t-3}^{t}$ |
| $t-2,m-2$ | | | $x_{t-2,m-2}^{t-2,m-1}$ | $x_{t-2,m-2}^{t-2,m}$ | $x_{t-2,m-2}^{t-1,m-2}$ | $x_{t-2,m-2}^{t-1,m-1}$ | $x_{t-2,m-2}^{t-1,m}$ | $x_{t-2,m-2}^{t,m-2}$ | $x_{t-2,m-2}^{t,m-1}$ | $x_{t-2,m-2}^{t,m}$ |
| $t-2,m-1$ | | | | $x_{t-2,m-1}^{t-2,m}$ | $x_{t-2,m-1}^{t-1,m-2}$ | $x_{t-2,m-1}^{t-1,m-1}$ | $x_{t-2,m-1}^{t-1,m}$ | $x_{t-2,m-1}^{t,m-2}$ | $x_{t-2,m-1}^{t,m-1}$ | $x_{t-2,m-1}^{t,m}$ |
| $t-2,m$ | | | | *nowcast* $y_{t-2}^{t-2}$ | $x_{t-2,m}^{t-1,m-2}$ | $x_{t-2,m}^{t-1,m-1}$ | $x_{t-2,m}^{t-1,m};y_{t-2}^{t-1}$ | $x_{t-2,m}^{t,m-2}$ | $x_{t-2,m}^{t,m-1}$ | $x_{t-2,m}^{t,m};y_{t-2}^{t}$ |
| $t-1,m-2$ | | | | | | $x_{t-1,m-2}^{t-1,m-1}$ | $x_{t-1,m-2}^{t-1,m}$ | $x_{t-1,m-2}^{t,m-2}$ | $x_{t-1,m-2}^{t,m-1}$ | $x_{t-1,m-2}^{t,m}$ |
| $t-1,m-1$ | | | | | | | $x_{t-1,m-1}^{t-1,m}$ | $x_{t-1,m-1}^{t,m-2}$ | $x_{t-1,m-1}^{t,m-1}$ | $x_{t-1,m-1}^{t,m}$ |
| $t-1,m$ | | | | | | | *nowcast* $y_{t-1}^{t-1}$ | $x_{t-1,m}^{t,m-2}$ | $x_{t-1,m}^{t,m-1}$ | $x_{t-1,m}^{t,m};y_{t-1}^{t}$ |
| $t,m-2$ | | | | | | | | | $x_{t,m-2}^{t,m-1}$ | $x_{t,m-2}^{t,m}$ |
| $t,m-1$ | | | | | | | | | | $x_{t,m-1}^{t,m}$ |
| $t,m$ | | | | | | | | | | *nowcast* $y_t^t$ |

*Note:* A quarter/month example of a mixed-frequency real-time data set is depicted. $x$ is the high-frequency (monthly) variable and is updated on a monthly basis, $y$ is the low-frequency (quarterly) variable and is updated on a quarterly basis. If new data release of $y$ occurs on a monthly basis as well, the table contains a "rectangular" shape. The bottom entry in the quarterly vintages are nowcasts of $y$ because we estimate the value of $y$ at a calendar date for which a data release has just occurred.

construct these forecast densities will be the subject of the next section. Alternatively, one can compute forecast errors in order to compare the results with forecasts using data that are not revised. Stark and Croushore (2002) discuss the issue of which observations to use as "actuals" in order to compare different methods graphically or to compute forecast errors. In particular, they consider three possibilities namely (a) the respective observation four quarters later, (b) the corresponding figure in the latest-available data set (last vintage) or (c) the first observation after a comprehensive benchmark, as mentioned for instance in Jacobs, Sturm and van Norden (2010), has occurred. The authors emphasize that as soon as real-time data are used, the choice of "actuals" may have important implications. To be more precise, if final-vintage values are chosen, using latest-available instead of real-time data may lead to deceptively lower RMSEs. Hence, when using latest available data for a model comparison or selection study, for example, the researcher should use data which was available to the researchers at the respective moment in time, i.e. real-time data, in order to conduct a "fair" study. Although not displayed in this paper for brevity, this result can be extended to an ADL setting.

As far as the classical ROF is concerned, we propose and employ a different observation to be considered as the "actual" value, i.e. the realization in the respective vintage itself. We refer to these as the real-time realizations or real-time "actuals" and the justification to employ this choice of "actuals" is given in the next section on the construction of the forecast densities.

For (2), it is seen in Table 1 that several vintages of $x$ can be used for the same vintage of $y$. This means that we have an additional dimension compared to the previous ROF study. Note that for every high-frequency regressor, the researcher is able to choose from a different set of vintages such that every combination of low-and higher-frequency vintages may be considered. Take a quarterly dependent and monthly independent variable as an example and assume both variables are revised on a quarterly and monthly basis, respectively. Then, if we want to forecast, say $y_{2009Q1}$ using information up to March 2009 we need not only rely on the vintage of December 2008 (remember the publication lag), but we may use the monthly vintages October 2008 until February 2009 for $x$ as well.

As far as nowcasting (3) is concerned, we consider the diagonal sequence $\hat{y}_t^t, \hat{y}_{t-3}^{t-3}, \hat{y}_{t-2}^{t-2} \ldots$ using a certain high-frequency vintage of the explanatory variable $x$ (for instance, always using the last monthly vintage of a quarter in a quarter/month example). Note, however, that depending on which vintage we use, we also need to forecast $x$ by a certain number of periods. As an example, assume again a quarter/month example where we intend to nowcast the third quarter of 2011 in September 2011 such that we employ the last monthly vintage per quarter for the regressor. Keeping in mind the publication lag, we have quarterly observations until the second quarter and monthly observations until August 2011. Note that if monthly vintages were available for the quarterly variable, these vintages could be chosen as well for nowcasting in order to employ most recently revised data. Hence, we need to forecast the observation for $x$ in September 2011. If the second (first) monthly vintage per quarter was chosen for the regressor, we needed to forecast two (three) months ahead and so on. Forecasts are computed using an $ARMA(p, q)$ model where the autoregressive and moving average orders do not change across vintages (an extension along these lines is straightforwardly done).

As already explained before, we extend the analysis of Stark and Croushore (2002) by considering an ADL setting where the regressors may be sampled at high-frequencies than the regressand. As such, more models are available than the ones dealt with in Stark and Croushore and it becomes interesting to investigate whether one of the models and, in particular, the (ECM-)MIDAS model designed to cope with variables sampled at different frequencies, dominates the others or whether a combination approach should be followed. The methodology to compare and combine several models is described in the following section.

# 4 Comparing and Combining Forecast Densities

In order to compare the forecast accuracy of different models across a range of vintages, we make use of a Diebold-Mariano (1995)-type test for the equality of two forecast densities which result from two competing models. The test is described, among others, in Diks et al. (2011a). Although they apply it to copula-based density forecasts, it is possible to employ the approach in our case as well because "the only requirement that [the authors] impose on the forecast methods is that the density forecasts depend on a finite number [...] of most recent observations".

The approach works as follows. Consider a particular calendar date, say $t^*$, and two competing forecast methods, say $A$ and $B$, each yielding, as discussed above, one one-step-ahead (or $h$-step-ahead) forecast of $y^v_{t^*}$ for each vintage $v$ considered, i.e. $v = t^* + 1, \ldots, T$. From the resulting vectors of one-step-ahead forecasts, i.e. $\widehat{\mathbf{y}}_{A,t^*}$ and $\widehat{\mathbf{y}}_{B,t^*}$, forecast densities are constructed, i.e. $\widehat{f}_{A,t^*}$ and $\widehat{f}_{B,t^*}$. In this paper, we construct forecast densities by applying kernel density estimation techniques. To be more precise, let us assume that the $T - t^*$ one-step-ahead forecasts have some underlying unknown distribution $f$. Then, the kernel density estimator is

$$\widehat{f}_{\cdot,t^*}(y) = \frac{1}{(T-t^*)h} \sum_{i=1}^{T-t^*} K(\frac{y - \widehat{y}_{i,\cdot,t^*}}{h}),$$

where $K(\cdot)$ is the kernel function and $h$ is the bandwidth parameter. In this paper, the Gaussian kernel, i.e.

$$K(x) = \frac{1}{\sqrt{2\pi}} exp(\frac{1}{2}x^2)$$

is considered where the bandwidth is chosen of the order $\widehat{\sigma}(T - t^*)^{-1/5}$, where $\widehat{\sigma}$ is an estimator of the standard deviation. The results are robust to the use of other kernels, e.g. the Parzen or Triangle kernel (see, for example, Greene 2008).

We follow scoring rules in order to evaluate density forecast, an avenue often followed in this field of literature (see Diebold and Lopez 1996). A scoring rule is a loss function whose value depends on the density forecast, $\widehat{f}_{\cdot,t^*}$, and the vector of "actual" values, $\mathbf{y}_{t^*}$. Hence, it may be denoted as $S(\widehat{f}_{t^*}, \mathbf{y}_{t^*})$ and is constructed in such a way that a 'better' density forecast gets a higher score. In this paper we will make use of the logarithmic scoring rule

$$S(\widehat{f}_{\cdot,t^*}, \mathbf{y}_{t^*}) = log[\widehat{f}_{\cdot,t^*}(\mathbf{y}_{t^*})],$$

which is shown to be a proper scoring by Diks, Panchenko and van Dijk (2011b) meaning that, under this scoring rule, "an incorrect density forecast does not receive a higher average score than the true conditional density".

At this stage, it should become clear why we choose real-time realizations over final-vintage data as "actuals". If we used latest-available data as "actuals", the scoring vector would consist of merely one element making it rather uninteresting to base forecast density comparisons on. Employing real-time realizations allows us to investigate whether one model outperforms another one in terms of real-time forecasting, which is precisely what we aim for.

Similarly to the Diebold-Mariano (1995) test for equal forecast accuracy, we can compare two forecast methods by defining the score difference as

$$d_{t^*} = S(\widehat{f}_{A,t^*}, \mathbf{y}_{t^*}) - S(\widehat{f}_{B,t^*}, \mathbf{y}_{t^*})$$

and formulating the null hypothesis of equal scores as $H_0 : E(d_{t^*}) = 0$. Letting $\overline{d_{t^*}}$ denote the sample mean of the score differences, we may use the following Diebold-Mariano (1995)-type $t$-test to investigate the null hypothesis

$$t_{A,B} = \frac{\overline{d_{t^*}}}{\sqrt{\widehat{\sigma}^2_{d_{t^*}}/(T - t^*)}},$$

where $\widehat{\sigma}^2_{d_{t^*}}$ is a heteroskedasticity and autocorrelation-consistent (HAC) variance estimator of $\sigma^2_{d_{t^*}} = Var(\sqrt{T - t^*} * \overline{d_{t^*}})$. Diks et al. (2011a) show, by referring to the proof in Giacomini and White (2006), that the $t$-test is standard normally distributed under a set of conditions on the stochastic processes considered and on the (variance of the sample mean of the) score differences. We will use this approach in order to compare all models under consideration with each other to determine whether one of them dominates the others.

However, even if it turns out that one model does so, it is known in the literature for point forecasting (see e.g. Timmermann 2006) that combinations of forecasts offer diversification gains making it attractive to combine the information present in the different forecasting methods under consideration. Likewise, being equipped with forecast densities produced by different models, a density combination approach is a natural avenue to follow as in Aastveit et al. (2011). Justifications for density combinations can be found in Mitchell and Hall (2005) or Hall and Mitchell (2007).

Similar to other studies on forecast density combinations we consider a so-called linear opinion pool, i.e. defining the combined forecast density as a linear combination of the individual density forecasts:

$$\widehat{f}_{Combination,t^*}(\mathbf{y}) = \sum_{i=1}^{M} \phi_i \widehat{f}_{i,t^*}(\mathbf{y}),$$

where $M$ is the number of models to combine. We follow Jore, Mitchell and Vahey (2010) in defining the recursive log score weights of model $i = 1, \ldots, M$ as

$$\phi_i = \frac{exp(\overline{S_i})}{\sum_{i=1}^{M} exp(\overline{S_i})},$$

where now,

$$\overline{S_i} = \frac{1}{T - t^*} \sum_{v=t^*+1}^{T} log[\widehat{f}_{i,t^*}(\widehat{y}_{t^*}^{v})].$$

# 5 Models

In this section we describe several models whose real-time forecasting performances we compare to the one of an autoregressive model employing the approach illustrated in the previous section. Keeping in mind that we combine the issues of dealing with variables sampled at different frequencies and the use of real-time data, we consider the previously mentioned MIDAS approach as well as two classical models of temporal aggregation. We, however, abstract from other models such as factor models which present a popular choice when fore- or nowcasting GDP growth rates (Giannone et al. 2008). Although the methodology described in this paper can just as well be applied to any other model with a possibly large set of regressors such as Andreou, Ghysels and Kourtellos (2010b), one has to obtain and clean, i.e. create missing data releases, correcting for possible mismatches or dealing with multiple vintages for one time period, a real-time data set for each regressor concerned, a task which may be very time-consuming or even impossible (due to non-existent data) such that we present the methodology for a small set of models and regressors.

Furthermore, we take into account the possibility of cointegration between the variables by considering the models with and without a cointegrating relationship. Although we aim at fore- or nowcasting GDP growth rates, it is well possible that GDP is cointegrated with one or more regressors. As shown by e.g. Götz, Hecq and Urbain (2011), disregarding (including) a long-run relationship in the presence (absence) of cointegration leads to a considerably worsened forecasting performance. Hence, all models (except the autoregressive one) are in an error-correction format in which a long-run term ($z_{t-1}$ below) is either in- or excluded.

MIDAS regression models have been introduced by Ghysels et al. (2004) and aim at preserving information present in the high-frequency variables while estimating parameters in a parsimonious way. Estimating standard linear regression models unrestrictedly might be unappealing due to parameter proliferation (Andreou et al. 2010b). If $y_t$ is a quarterly variable and $x_{t,m-i}$ is daily, we might have to estimate over 50 parameters. In an (ECM-)MIDAS model we hyper-parameterize the polynomial lag structure yielding

$$\alpha^*(L)\Delta y_t = \quad c + \delta z_{t-1} + \beta \sum_{i=0}^{m-1} w_{i+1}(\theta)\Delta_m x_{t,m-i} + u_t, \tag{1}$$

where $\alpha^*(L)$ is a lag polynomial allowing for autoregressive terms, $z_{t-1}$ denotes the disequilibrium error in the previous period and $w_{i+1}(\theta)$ are weights that sum up to one in order to identify the scale coefficient $\beta$. The weights are based on an underlying function for which different specifications are proposed in the literature (see, for instance, Ghysels, Sinko and Valkanov 2007). Due to its dominance documented in Ghysels and Valkanov (2006) we employ the exponential Almon lag polynomial which is shown to be low-dimensional and extremely flexible allowing the weight-determination to be completely data-driven. Often, decaying weights are detected due to more recent observations being more important.

Note that in the equation above one weight function is specified for the whole set of short-run dynamics terms. There are two comments to be made here: First, the purely data-driven determination of the weight function allows us to include more short-run variables than theoretically necessary since the redundant ones will be assigned a weight of zero. Second, it is also possible to assign one weight function to every low-frequency period of the short-run variables (see Ghysels et al. 2007). Nevertheless, given that in our case the high frequency is monthly and the low frequency is quarterly ($m = 3$ hence), such an approach seems unappealing.

For the exact determination of the above ECM-MIDAS model, which is derived from a mixed-frequency autoregressive distributed lag model, we refer the reader to Götz et al. (2011). As far as the long-run term is concerned, we consider three possible cases:

(i) $z_{t-1} = y_{t-1} - \gamma x_{t-1}$: 'same-period'-case,

(ii) $z_{t-1} = y_{t-1} - \gamma x_{t,m-i}$ with $i < m$: 'x-after-y'-case,

(iii) $z_{t-1} = y_{t-1} - \gamma x_{t-1,m-i}$ with $i < m$: 'x-before-y'-case.

The first case corresponds to the standard framework in which the two series are sampled at the same moment. As emphasized by Götz et al. (2011), in the mixed frequency modeling there are alternative intuitive possibilities and for real-time data, keeping in mind the release of the data by statistical offices, the 'x-after-y case' is appealing. We illustrate in more detail how the timing of the high-frequency observation appearing in the long-run relationship impacts the derivation of the short-run dynamics. In any case, the short-run dynamics terms are high-frequency differences which can be estimated by an ECM-MIDAS regression. Due to the stationarity of these terms, independent from whether we assume $x$ to be I(0) or I(1), the estimations do not suffer from computational complications.

An alternative to ECM-MIDAS is the classical approach of temporally aggregating the high-frequency series into a low-frequency one as discussed by Marcellino (1999) amongst others. Averaging or Point-in-time sampling are the common principles here. The former approach takes the average of the high-frequency observations in a certain $t$-period as the corresponding low-frequency observation. The latter approach takes one specific observation in a period (often the last one) to be the corresponding $t$-observation.

Both approaches are in fact special cases of the ECM-MIDAS model considered before. Note that equation (1) actually summarizes all three models, ECM-MIDAS, Point-in-Time sampling and Averaging, depending on the structure of the weights:

$$w_{i+1}(\theta) = w_j(\theta_1, \theta_2) = \frac{exp(\theta_1 j + \theta_2 j^2)}{\sum_{j=1}^{k} exp(\theta_1 j + \theta_2 j^2)} \text{ (Almon Lag Polynomial; MIDAS)},$$

$$w_{i+1}(\theta) = w_{i+1} = \begin{cases} 1 \text{ if } i = 0 \\ 0 \text{ else} \end{cases} \text{ (Point-in-Time sampling)},$$

$$w_{i+1}(\theta) = w_{i+1} = \frac{1}{m} \; \forall i = 1, \ldots, m \text{ (Averaging)}.$$

As explained before, in order to flexibly deal with the presence or absence of cointegration in the just described models, the analysis is conducted once with and once without the inclusion of a long-run relationship. As in Götz et al. (2011) we base the choice of the timing in the disequilibrium error on an approach similar to the two-step Engle-Granger approach (1987). In particular, using the last vintage available, we screen through the possible cases described above ('x-before-y', 'same-period' and 'x-after-y') and pick the same-period case (for simplicity) given that the variables cointegrate for all candidates of $x$. If they do so only for some specific candidates, we choose one of these arbitrarily due to the invariance of temporal aggregation to the property of cointegration (Marcellino 1999). The resulting long-run relationship will be plugged into an error-correction model. Note that this timing will subsequently be employed for all vintages under investigation. Furthermore, for each vintage the cointegrating coefficient will be estimated by OLS.

Finally, in order to analyze whether an ADL setting improves real-time forecasts at all, we consider an $ARIMA(4, 1, 0)$ as is done almost throughout the whole investigation in Stark and Croushore (2002).

Hence, we have the following seven cases:

1. Restricted short-run (ECM-MIDAS); long-run relationship included,
2. Restricted short-run (ECM-MIDAS); long-run relationship excluded,
3. Average sampling; long-run relationship included,
4. Average sampling; long-run relationship excluded,
5. Point-in-Time sampling; long-run relationship included,
6. Point-in-Time sampling; long-run relationship excluded,
7. $ARIMA(4, 1, 0)$.

# 6 Empirical analysis

## 6.1 Data

We extract real time data sets for the US from http://alfred.stlouisfed.org/. $y$ is the quarterly real gross national product (ref. GNPC96), seasonally adjusted. The series is observed from 1960Q1 until 2010Q3 (for this exercise we have extracted the series in February 2011). There are monthly vintages from July 1986 to December 2010 where we, however, disregard the additional monthly vintages (i.e. keep the end-of-quarter vintages as quarterly vintages) in order to focus on the additional vintage-dimension originating from employing high-frequency regressors. For $x$ we consider, for the same time span, the monthly seasonally adjusted industrial production index (ref. INDPRO) and the daily S&P 500 stock index (ref. SP500). Monthly vintages from July 1986 to December 2010 are used for both regressors ($m_{INDPRO} = 3$). This means that there are 20 daily observations per vintage which, in turn, implies $m_{SP500} = 60$ (In order to have the same number of working days per vintage, and thereby quarter, we took the maximum number of working days which were available in each month over the time period considered. This lead to 20 daily observations per month). Note that the S&P500 stock index is not revised at all such that along different vintages new data becomes available, but existing ones do not change. Note that one could easily generate daily vintages as well. We, however, facilitate the analysis by only considering monthly vintages for both regressors.

Moreover, the vintages do not match exactly for the three series $y$, $x_{INDPRO}$ and $x_{SP500}$. The first mismatch comes from the fact that they are not published on the same day. We do not do anything about that. When there are two vintages for the same month (this happens for INDPRO), we take the vintage at the half of the month. The second problem is that there are, at different releases, missing vintages for the two variables. To facilitate the computation we create missing data releases such that $v = v - 1$, i.e. missing values will be assigned their respective value in the previous vintage. The alternative would have been to delete uncommon vintages.

A further remark concerns the fact that figures of both series are computed with reference to a base year. Due to changes in the base years, the values jump at the respective dates. Note that because we forecast growth rates of output, this is not an issue for our ROF analysis.

Note that the empirical analysis deals with merely two regressors whereas some alternative studies on forecasting GDP growth employ hundreds of variables or bivariate regressions. Furthermore, we consider only seven models as described in a previous section. We want to stress again that, in this paper, emphasis is on the methodology and its implications, i.e. the weights implied by the combinations of the various forecast densities constructed using ROF, in contrast to presenting an alternative regressor set producing possibly superior forecasts of GDP growth. The analysis presented in the next section is easily extended to include a larger set of regressors and/or forecasting models such as e.g. factor models as explained before.

## 6.2 Model Comparison and Combination: ROF

Before comparing (and combining) the forecast densities produced by the seven models introduced above, we analyze whether cointegration between $y$ and any of the two regressors is present. As illustrated above, using the last vintage available, we screen through the candidate timings and report the Engle-Granger (1987) ADF-$t$-test as well as the $p$-value associated with Johansen's trace test in Table 2 below. Note that GDP and S&P500 do not seem to be cointegrated such that only the results for a potential cointegrating relationship between GDP and industrial production ($x$ in the table) are presented.

| Error-Correction term | Timing | $t_{ADF}$ (p-value) | Johansen's Trace Test for no CE(s) p-value |
|---|---|---|---|
| $y_{t-1} - \beta x_{t,2}$ | 'x-2-after-y' | $-3.257$ | 0.0445 |
| $y_{t-1} - \beta x_{t,1}$ | 'x-1-after-y' | $-3.037$ | 0.0925 |
| $y_{t-1} - \beta x_{t-1,3}$ | 'Same Period' | $-3.238$ | 0.0898 |
| $y_{t-1} - \beta x_{t-1,2}$ | 'x-1-before-y' | $-3.131$ | 0.1041 |
| $y_{t-1} - \beta x_{t-1,1}$ | 'x-2-before-y' | $-2.946$ | 0.1444 |

Table 2: $t_{ADF}$ and Johansen's Trace test

Given the outcome of the $t_{ADF}$ statistic and Johansen's Trace test, there seems to be merely a weak cointegrating relationship of the 'x-after-y'-kind at the second month between the two variables. For the other candidate timings Johansen's trace test does not reject the null hypothesis of no cointegrating equations at the 5% significance level. Hence, the ECM-MIDAS model including a long-run term employs the $x$-variable 'two months after' the regressand where the cointegrating vector will be re-estimated by OLS for every vintage as mentioned before. In case of one of the classical aggregation models, the $x$-variable entering the long-run term will be computed accordingly as the average or the end-of-quarter observation, respectively. Note that due to the weakness of the cointegrating relationship, which might be even more pronounced for earlier vintages, we may expect the models excluding a cointegrating relationship to perform quite well.

In order to compare whether one of the seven models introduced above uniformaly dominates the others, we conduct the ROF (repeated observation forecasting) and compare the approaches at hang using the methodology presented in Section 4. As an illustrative example, three dates, roughly equally distant from each other, are considered, 1986Q3, 1991Q4 and 1996Q1; the results of the analysis for all calendar dates are presented subsequently. It emerged that even after 20 quarters there are still a lot of revisions in the growth rates taking place. This is emphasized in Figure 1 where even after a large amount of new data releases has occurred, there are quite large movements for some dates, e.g. 1986Q3.

The growth rates of quarterly real gross national product for the respective periods are forecasted in each vintage and the resulting one-step-ahead forecasts are used to construct forecast density functions. Note that for each date a different number of one-step ahead forecasts is calculated due to a different number of vintages to choose from (97, 76 and 59, respectively).

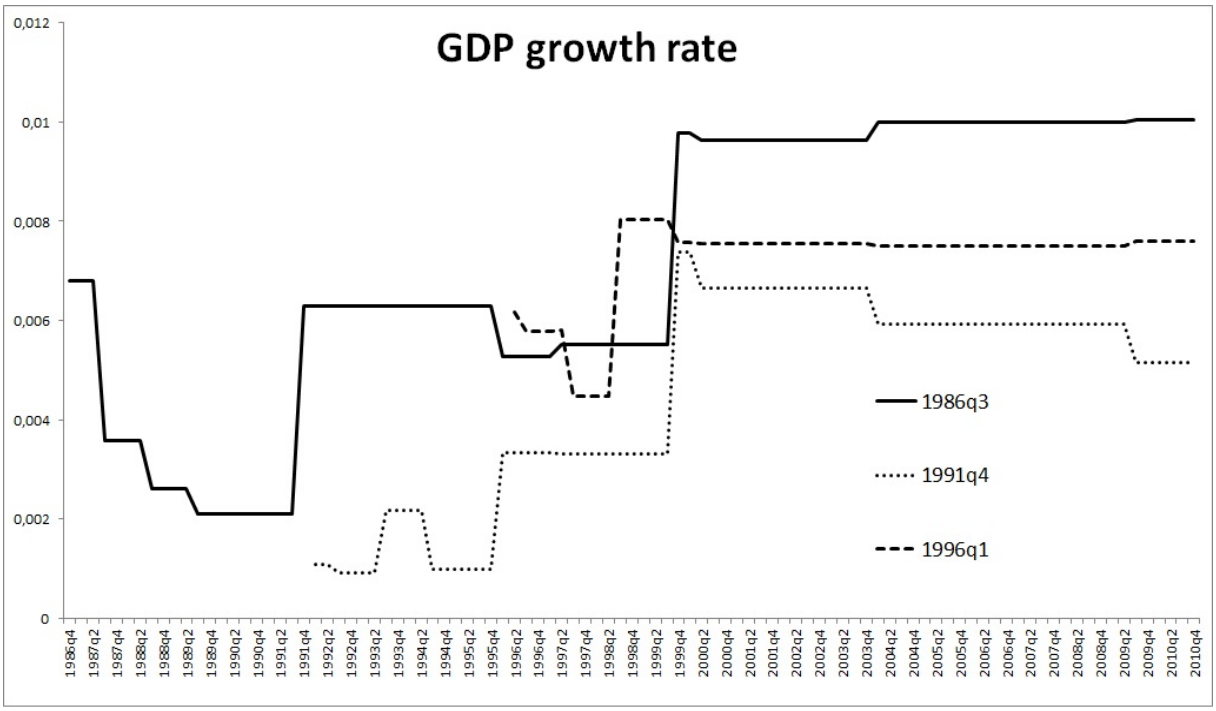


Figure 1: Revised GDP growth rate along monthly vintages

Point-in-time sampling is denoted by PIT, Average sampling by AV and the real-time realizations by D_Q. Whenever a long-run relationship is included, COINT is appearing in the label.

The forecast densities based on the one-step-ahead forecasts produced by the respective seven models and a combination forecast density based on the scores of these seven model-based forecast densities are plotted in Figures 2 to 4. Additionally, histograms of the real-time realizations, which are used as "actuals" in this analysis, are displayed in the top right corners. Finally, the vertical bars represent the realizations of the final vintage, i.e. another value commonly used as "actual" in studies dealing with real-time data.

The first striking observation about these figures is that we are dealing with density functions which are sometimes spread over a wide data range. This implies that the forecasts of the GDP growth rates differ substantially as new data releases take place, just as the corresponding realizations do (see Figure 1). Furthermore, it turns out that in all three figures each model results in quite different forecast densities. In the case of 1986Q3 for example, MIDAS_COINT produces a rather flat forecast density whereas the one of PIT is more concentrated on a smaller region. Also, the AR-model- and the AV- or the PIT_COINT-based forecast densities put positive probabilities on almost entirely disjoint data regions. Similar observations can be

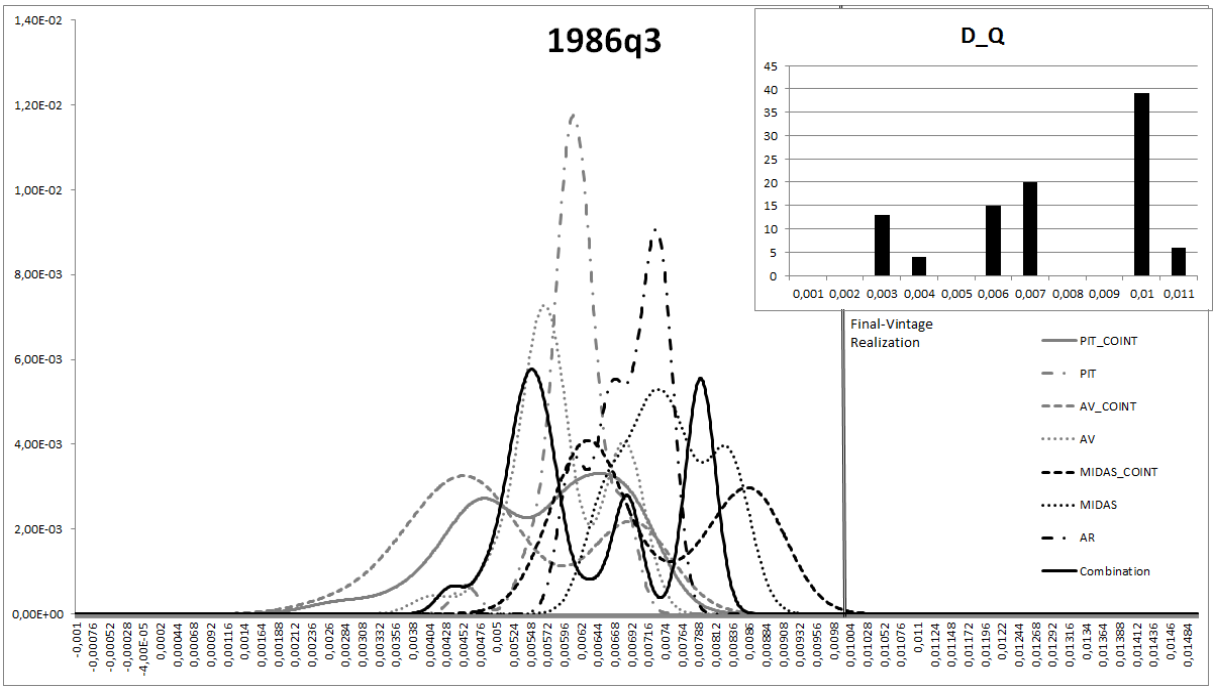


Figure 2: Forecast Densities and Real-time actuals 1986Q3

made for the other forecast densities in the case of 1991Q4 or 1996Q1.

In order to judge whether and in what sense one model outperforms another one, it becomes crucial to choose which observations are taken as "actuals". When comparing two models we take the real-time realizations as "actuals" as discussed before. A histogram of these values is given in top right corner of each figure. If a model was able to forecast the GDP growth rate in real-time perfectly, i.e. produce the real-time realization for each vintage, the forecast density would exactly reproduce the shape implied by the histogram. The less it mimics its shape, the lower the score of the model concerned and the more likely it is going to be dominated by a competing approach. Based on the plot of the forecast densities produced by each model we can already form an expectation of which model is superior to another one.

Take, for example, the case of 1991Q4 and note that the real-time realizations range from 0.001 to 0.008, their histogram has a tri-modal shape with local maxima at 0.001 and 0.004 and a global one at 0.006. The AR-model-implied forecast density ranges only from 0.005 to 0.008 (i.e. scoring zero on the smaller real-time realizations) with two modes at 0.006 and 0.0075. All other approaches result in forecast densities ranging from roughly 0.003 to 0.0085 with a bimodal (except for PIT_COINT) shape where the first one is a local maximum at around 0.004 and the second is a global maximum at around 0.006 to 0.0075 depending on which approach is concerned. Hence, none of the models seem to be able to forecast the mode at 0.001 (they seem to have overestimated the growth rate for the respective vintages) although all models appear

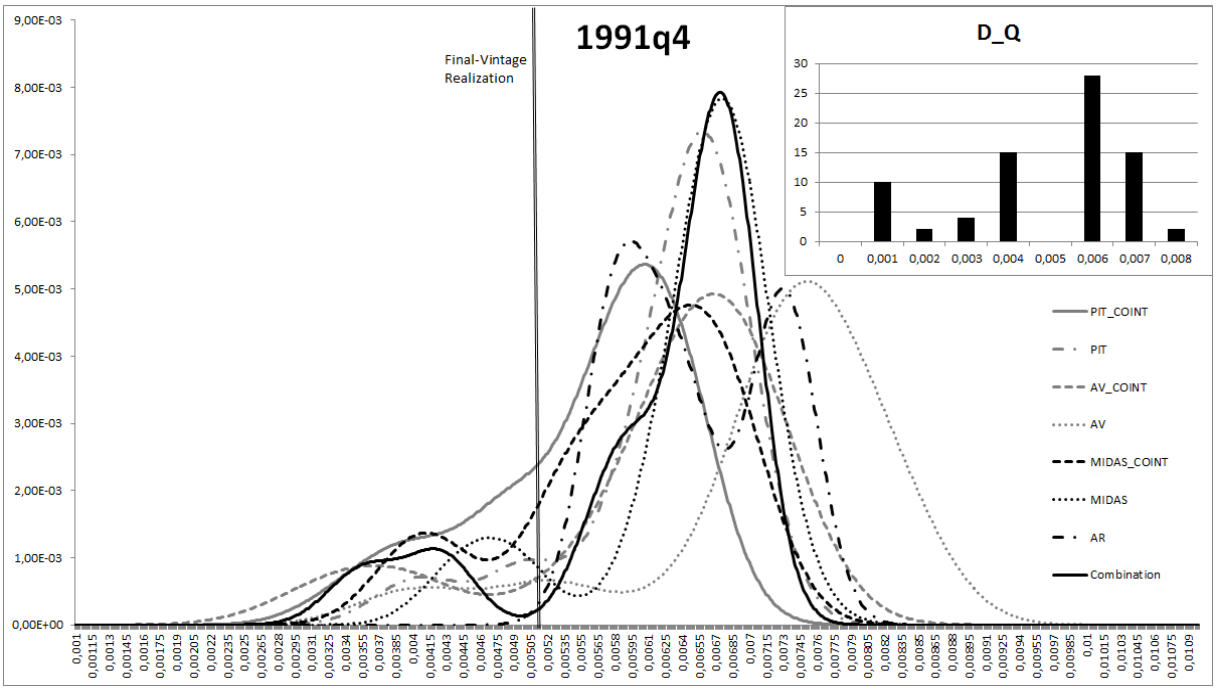


Figure 3: Forecast Densities and Real-time actuals 1991Q4

to reproduce the shape of the histogram quite well. Whether a model significantly outperforms the others remains to be formally tested employing the Diebold-Mariano-type test.

The situation may, however, not be as "good" as it is for 1991Q4 in the sense that the forecast densities reproduce the histogram of the real-time realizations quite well. In the case of 1986Q3 none of the seven models' forecast densities seem to display the global maximum in the "actuals" at 0.01. In fact, only the forecast density produced by MIDAS_COINT places positive probability on its value which is why it will probably outperform the other approaches in terms of the Diebold-Mariano-type test (because the other models score zero on almost half of the real-time realizations). Hence, if we test whether, for instance, AV displays an equally well forecasting performance as PIT, it should be pointed out that we may consider two "bad" models. The test merely tells us whether or not one approach dominates another one, not whether this approach has a good forecasting performance in terms of the "actuals" we consider. Wether or not some, or even all, models are entirely useless can be assessed graphically.

Note that our conclusions might change entirely if we focus on another value as "actual". Often the final-vintage (or latest-available) realization is taken to be the "actual" value implying that the forecasting model placing most weight on its value will be the preferred one. In Figures 2 to 4 the final-vintage realization is represented by a vertical bar such that for 1986Q3 MIDAS_COINT dominates the others, for 1991Q4 we would choose PIT_COINT whereas for

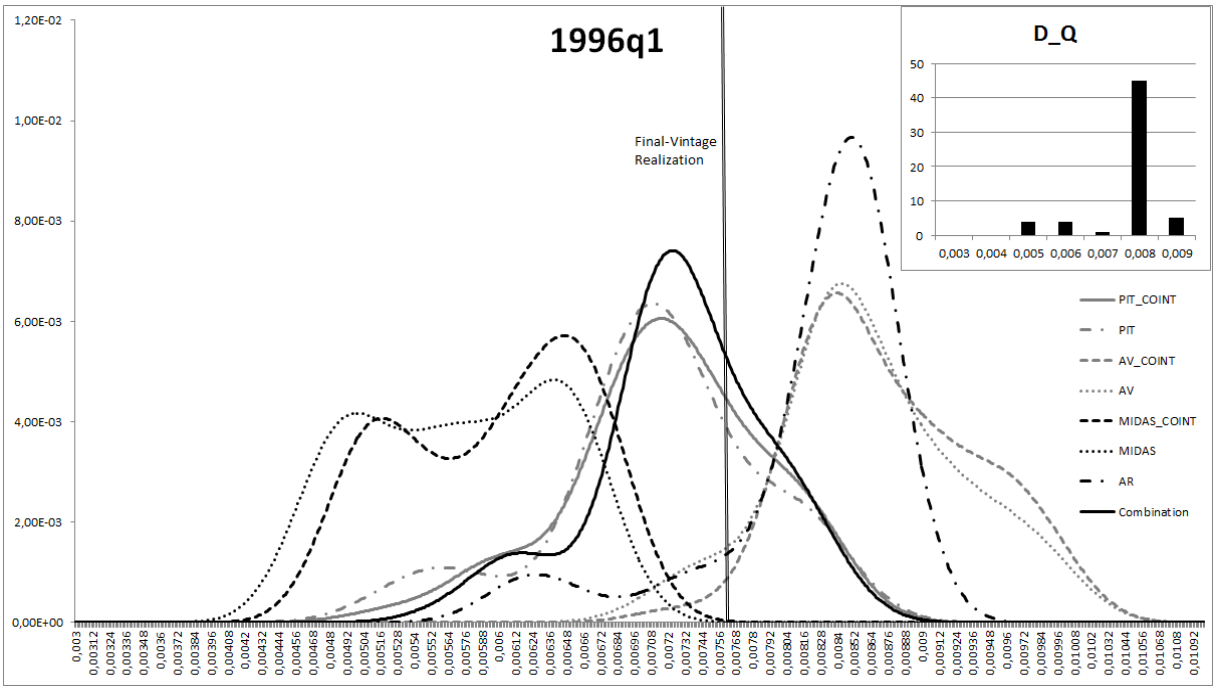


Figure 4: Forecast Densities and Real-time actuals 1996Q1

1996Q1 our choice would be either of the two. It may however occur that none of the models place any positive probability on the latest-available realization (it almost happens for 1986Q3). In such a situation the model being "closest" to its value could be chosen (for 1986Q3 the AR-model if MIDAS_COINT is left out of consideration).

As explained above, we produce a combined density forecast based on the scores of the seven model-based forecast densities. As argued extensively in the point forecast literature such an approach may lead to improved results due to diversification arguments. Even though MIDAS_COINT might dominate the other models for 1986Q3, some of the other approaches may carry useful information for real-time forecasting. The resulting combined density forecasts are represented by solid black lines in Figures 2 to 4. It emerges that the shape of its forecast density replicates the shape of the histogram very well. For example, in the case 1986Q3 the combined density forecast mimics the tri-modality of the histograms and for 1991Q4 it detects the minimum at 0.005 almost perfectly.

Table 3 summarizes the results for the three respective dates. In particular, the $t_{A,B}$-tests of Section 4 are displayed where $A$ is the model mentioned in the respective row and $B$ the one in the respective column. As such, a positive (negative) value indicates that the row-model achieves a larger (smaller) average score than the column-model, i.e. model $A$ outperforms (is dominated by) model $B$. Whether or not this difference is significantly different from zero needs to be assessed by means of comparisons to respective critical values. The implied weights given

to each model when combining forecast densities are given as well.

| *1986q3* | PIT_COINT | PIT | AV_COINT | AV | **MIDAS_COINT** | MIDAS | AR |
|---|---|---|---|---|---|---|---|
| PIT_COINT | | 6.9002 | -4.8966 | 6.6075 | **-1.4522** | 2.1985 | 5.6717 |
| PIT | | | -6.8153 | -7.0083 | **-6.2709** | -3.9448 | -0.6280 |
| AV_COINT | | | | 6.3743 | **-0.5119** | 2.6541 | 5.8807 |
| AV | | | | | **-4.6560** | -0.3681 | 4.1459 |
| **MIDAS_COINT** | | | | | | **4.0086** | **6.7711** |
| MIDAS | | | | | | | 7.1393 |
| Weights | 0.11 | 0.00 | 0.34 | 0.00 | **0.55** | 0.00 | 0.00 |
| *1991q4* | PIT_COINT | PIT | **AV_COINT** | AV | MIDAS_COINT | MIDAS | AR |
| PIT_COINT | | 2.9566 | **-2.4422** | 4.4038 | 2.9372 | 3.2949 | 3.5717 |
| PIT | | | **-2.8053** | -1.9089 | -2.9647 | 3.7185 | 3.6557 |
| **AV_COINT** | | | | **4.8891** | **2.7374** | **3.1184** | **3.5029** |
| AV | | | | | 1.2742 | 2.6630 | 3.3982 |
| MIDAS_COINT | | | | | | 3.4732 | 3.6220 |
| MIDAS | | | | | | | 3.6398 |
| Weights | 0.23 | 0.05 | **0.46** | 0.15 | 0.10 | 0.01 | 0.00 |
| *1996q1* | **PIT_COINT** | PIT | AV_COINT | AV | MIDAS_COINT | MIDAS | AR |
| **PIT_COINT** | | **1.0920** | **2.5243** | **2.2983** | **6.0188** | **6.2524** | **3.6826** |
| PIT | | | 2.4553 | 2.2216 | 6.2752 | 6.4841 | 3.3106 |
| AV_COINT | | | | -3.6960 | -0.7280 | -0.3271 | -2.1340 |
| AV | | | | | -0.2828 | 0.1479 | -1.7252 |
| MIDAS_COINT | | | | | | 6.8273 | -1.6390 |
| MIDAS | | | | | | | -2.3112 |
| Weights | **0.33** | 0.32 | 0.03 | 0.05 | 0.07 | 0.04 | 0.15 |

Table 3: Density Comparisons for the three dates and the implied weights to get the combined density forecasts

First of all, note that a model having only negative figures in its column and only positive ones in its row shows a better forecasting accuracy than all competing approaches and is repeated in bold letters above for each date under consideration. Wether or not such a difference is significantly different form zero needs to be assessed by comparing the $t_{A,B}$-statistics to asymptotically standard normal (otherwise bootstrapped) critical values. With respect to the discussion about Figures 2 to 4, it turns out that for 1986Q3 MIDAS_COINT is indeed (significantly) dominating the other models. For 1991Q4 AV_COINT leads to the best forecasting performance and the AR-model is significantly outperformed by all approaches as was expected due to its nonrecognition of the two local maxima smaller than 0.006. For 1996Q1 the model yielding the "best" forecasting performance is PIT_COINT. These observations are confirmed by the implied weights given to each model when combining their forecast densities. Note, however, that the dominating model does not get a weight of one, but that some, or even all (1996Q1), other models are assigned positive weights as well.

Note that the AR-model is assigned a zero weight for two of the considered dates. It turns out, when considering all calendar dates, that the autoregressive model is very often outperformed by all competing approaches. In the case of 2001Q3 e.g. the growth rate rose for

the past four quarters, but declined precisely in the third quarter of 2001. In such a situation the autoregressive model is not able to predict this decline whereas the other models, taking into account mid-quarter information, are better able to do so. Hence, adding explanatory variables improves the forecasting performance across different vintages.

As became evident from the figures and the table above, none of the models seems to dominate the other ones uniformly. In fact, for all three dates, a different model is assigned the largest weight when combining forecast densities which, in turn, seems to improve on the individual model-based forecast densities. So far, however, we have merely looked at three dates. How does the weight distribution over the seven models under consideration evolve over all calendar dates? If they turn out to be rather constant, we would have found a weight distribution which could be employed for future calendar dates. The following figure graphs the weights of each of the seven models for all calendar dates under consideration. If the weights stay constant across calendar dates, we would expect rectangular areas (of possibly different height) for each approach.

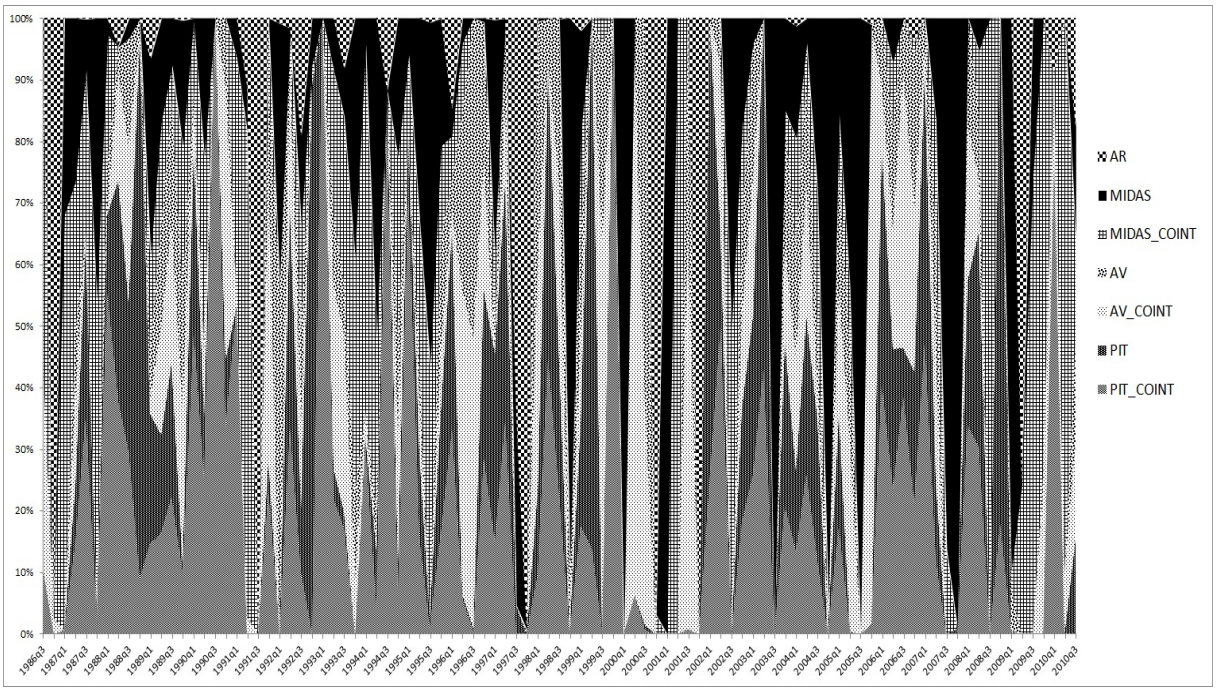


Figure 5: Weights given to the seven models over all calendar dates

Clearly, Figure 5 suggests that the weights do not stay constant along different calendar dates. The weight distribution seems to alter immensely as we move from the left to the right of the figure. Hence, instead of constant weights we should consider time-varying weights.

What does the whole previous analysis imply for a practitioner who is aiming at conducting a forecasting study employing the latest-available data set? Assume that he or she has several

forecasting models to choose from and is also familiar with traditional forecast combination approaches such as equally weighting the forecasts or applying a Bates and Granger (1969) weighting scheme. Our time-varying weights obtained by considering the real-time forecasting performance using a combination of forecast densities offer an alternative choice to the researcher. Note that classical weighting schemes such as the two mentioned above assign a constant weight to each forecast whereas our weights possibly change for every calendar date under consideration.

Do our time-varying real-time weights yield a better forecast accuracy than the individual models or other combination schemes such as e.g. Bates and Granger (1969) weights? To investigate this, we consider forecasting US GDP growth from 1995Q1 until 1999Q4 employing the latest-available vintage. In particular, we conduct 20 one-step-ahead out-of-sample forecasts using an estimation period from 1960Q2 to 1994Q4 and compute the RMSEs using each model individually, combining them using equal weights for the forecasts, using Bates and Granger (1969) weights and using our proposed time-varying ROF-based weights. The outcome of the analysis is summarized in the first column of Table 4.

| Model | RMSE (1995Q1-1999Q4) | RMSE (2005Q4-2010Q3) |
|---|---|---|
| PIT_COINT | 0.0205 | 0.0219 |
| PIT | 0.0206 | 0.0230 |
| AV_COINT | 0.0233 | 0.0220 |
| AV | 0.0232 | 0.0229 |
| MIDAS_COINT | 0.0182 | 0.0275 |
| MIDAS | 0.0185 | 0.0275 |
| AR | 0.0193 | 0.0377 |
| Bates and Granger | 0.0190 | 0.0221 |
| Equal Weights | 0.0194 | 0.0229 |
| ROF-based | 0.0160 | 0.0211 |

Table 4: RMSEs of the individual models and the competing combination methods

According to Table 4, employing ROF-based weights as described in this paper yields considerably better forecasts. Hence, for this data set and the models employed here, it seems that the use of a real-time data set to construct forecast densities from which to deduce time-varying weights improves the forecasting performance.

The previous small exercise focused on a period in the past, but in how far do the results change when we consider a more recent period to forecast? Note that our weights are determined by combining the forecast densities of several models which were constructed employing the forecasts based on different vintages for a particular calendar date. As there are fewer vintages to choose from for more recent dates, it implies that the weights themselves are based on fewer observations, i.e. forecasts which were used to construct forecast densities. To investigate whether the ROF-based weights suffer from the problem of being based on too few vintages, we consider forecasting the US GDP growth rate from 2005Q4 until 2010Q3, i.e. again producing

20 one-step-ahead out-of-sample forecasts. Again, we compare the forecasting performance of our seven approaches and the three combination methods from before with each other where we construct the ROF-weights exactly as before ignoring the fact that the forecast densities are based on very few vintages. The results are displayed in the second column of Table 4.

It seems that the use of the ROF-based weights leads to a forecasting performance that is not significantly different from the other combination approaches, analyzed using Diebold-Mariano (1995) tests, although it yields the lowest RMSE. Hence, when more recent calendar dates are dealt with, the superiority of forecast combinations employing our time-varying weights might be lost, at least in this particular study. To circumvent this problem, one could follow two approaches. First, the weights could be forecasted using a VAR model. For this particular study a VAR(0) is chosen according to information criteria such that the simple averages of the weights, displayed in Table 6, could be chosen as weights when forecasting the more recent period. Second, one could forecast vintages as is done in Clements and Galvão (2011) or Une (2011) in order to increase the number of data releases and thereby get a more reliable weight construction.

| Model | PIT_COINT | PIT | AV_COINT | AV | MIDAS_COINT | MIDAS | AR |
|---|---|---|---|---|---|---|---|
| Weights | 0.1998 | 0.1533 | 0.1478 | 0.1038 | 0.1394 | 0.1834 | 0.0725 |
| | (0.0244) | (0.0214) | (0.0206) | (0.0137) | (0.0227) | (0.0301) | (0.025) |

Table 5: Averages of the weights

The previous small forecasting exercises showed that the time-varying ROF-based weights may lead to an improved forecasting performance for the empirical application at hand. In general, however, for which type of data do we expect our approach to perform well? Given the way the weights are determined, we expect that such situations occur in case of model uncertainty or if the data exhibit some mixture distribution. In the latter case the combined forecast density might correspond to an estimation of the mixture probability density function. A closer investigation of this issue might be undertaken via a Monte Carlo study where it is, however, far from obvious which data generating process to employ. Are the data generated by a linear combination of different models? Which set of models should we choose? Does one model dominate the others uniformly over all vintages and if not, which one dominates for which vintages? Or should the weights be chosen randomly in each vintage? Even if we allow for the latter, we believe that the true underlying mechanics in such a real-time setting are far too complex to be accurately modeled such that we do not expect new insights from a Monte Carlo study.

### 6.3 Multivariate Dimension ROF

As already explained before, extending the analysis of Stark and Croushore (2002) to an ADL setting with possibly higher-frequency regressors, enables the researcher to make use of additional dimensions of vintages originating from the high-frequency variables. So far, we have focused on the vintages corresponding to the last month of a quarter. It might, however, be

the case, that using the vintages corresponding to the first or second month of a quarter yields considerably better forecasting performances than using the end-of-quarter vintages. If so, this might have an impact on which data releases to base our forecast density functions on and how we should compute nowcasts of GDP (how many observations of the regressors we need to forecast) as indicated in the introduction. Even if it turns out that employing different monthly vintages does not lead to a significantly different forecasting performance, we would have gained the insight that we can make use of the most recent observations without expecting a worse forecasting performance than when we, for instance, waited another month for revised data.

In order to find out whether employing a certain vintage yields considerably better forecasting performances, we consider five candidate vintages per quarterly vintage of $y$ similarly to the candidate timings in the MIDAS models before: The $x$-vintage two months before the one of $y$, one month before that of $y$, the same vintage as $y$, the vintage of $x$ one month after the one of $y$ and, finally, two months after the one of $y$. Similar to the previous section, one-step-ahead forecasts are computed for each candidate vintage employing real-time data. The RMSEs are calculated in a similar way as the scores earlier, i.e. employing the real-time realizations as "actuals". This results in five RMSEs which can be compared with the modified version of the Diebold-Mariano test (Harvey, Leybourne and Newbold 1997). To focus on the question whether the use of one particularly timed data release yields a significantly better forecast accuracy than another one, we only consider the MIDAS model in this section.

The RMSEs for the cases 'x-2-vintages-before-y', 'x-1-vintage-before-y', 'same-vintage', 'x-1-vintage-after-y' and 'x-2-vintages-after-y' are 0.04113, 0.0415, 0.04111, 0.04164 and 0.04182, respectively, and Table 6 summarizes the modified Diebold-Mariano test statistics.

| Modified DM statistic | x-1-vint-before-y | same-vint | x-1-vint-after-y | x-2-vint's-after-y |
|---|---|---|---|---|
| x-2-vint's-before-y | -0.66411 | -1.1938 | -0.75762 | -0.97631 |
| x-1-vint-before-y | | -1.36613 | -0.4092 | -0.77747 |
| same-vint | | | 0.41172 | -0.13662 |
| x-1-vint-after-y | | | | -1.05856 |

Table 6: Diebold-Mariano test statistics for the different candidate vintages

With respect to the results in Table 6, we can conclude that none of the different high-frequency vintages yields better forecasting performances than the others. Hence, we may freely choose which $x$-vintage to employ for a real-time data analysis and nowcasting. It implies that for either fore- or nowcasting we can make use of the most recent observations available to us.

Intuitively, one would expect the forecasting performance of later vintages to improve because new or revised information has been released. Note, however, that we are merely employing one high-frequency variable for which data are revised and that this variable appears only three times as often as the dependent variable. According to Foroni, Marcellino and Schumacher (2011), mixed-frequency models outperform models which disregard the higher frequency in the regressors only for large enough frequency differences i.e. if the regressor(s) is (are) available at much higher frequency than the regressand. A similar argument might hold for data releases

taking place at a higher frequency for some variables. In fact, Aastveit et al. (2011) study how the forecast of GDP becomes more concentrated as new data becomes available during a quarter. However, they also consider a wide range of regressors with different frequencies such that much more data releases occur between two GDP releases. We expect that adjustments in both directions (more variables and/or sampling at a higher frequency) have a corresponding effect on the outcome of this analysis.

### 6.4 Nowcasting

Given a certain model or combination approach, the mixed-frequency nature of the data enables the researcher to choose from an enlarged set of vintages containing more (high-frequency) information. Hence, it is of interest in how far the use of these extra information influences the performance of using real-time data as opposed to latest-available data. More importantly, employing real-time data in combination with mixed-frequency variables enables the user to nowcast the dependent variable employing the most recent information available at that time. The issue of nowcasting GDP in particular is dealt with extensively in the literature, e.g. Evans (2005) and Giannone, Reichlin and Small (2008) who develop methods to estimate the current state of the economy continually employing the flow of information of various data releases. For an extensive survey we refer to Bańbura, Giannone and Reichlin (2011a).

Bańbura, Giannone and Reichlin (2011b) extend the framework of Giannone et al. (2008) to accommodate a general mixed-frequency data set. Other studies employing mixed-frequency data for nowcasting include Giannone, Reichlin and Simonelli (2009) who develop a mixed-frequency VAR for nowcasting GDP in the euro area, Marcellino and Schumacher (2010) who propose a so-called Factor-MIDAS model for (fore- and) nowcasting or Kuzin, Marcellino and Schumacher (2009) who compare the two latter approaches.

Given the outcome of the previous analysis, we construct nowcasts using the end-of-quarter vintages for the regressors due to simplicity. Of course, we could also employ any other (high-frequency) vintage for the regressors since the forecasting performance does not alter significantly. As an illustration, we plot the nowcasts of US GDP growth against the latest-available and real-time (since GDP growth for the current vintage is not released yet, we consider the corresponding value appearing the following vintage as real-time observation) observations in Figure 6.

Note that, as already indicated before, nowcasts may also be done during the quarter by employing additional monthly vintages for the quarterly variable GDP, for instance. Say we want to nowcast the third quarter of 2011 in August. Assuming only quarterly vintages of GDP, we only have information about the first quarter of GDP (since the last vintage for GDP appeared in June giving a value for the first quarter of 2011 due to the publication lag). On the contrary, employing higher-frequency vintages enables us to use a revised version of the second quarter of 2011 in August. Hence, an extension along this line is worth undertaking since for both, the regressand as well as the regressors, it is valuable to make use of high-frequency information whenever possible in order to include the most recent observations into fore- or nowcasts.

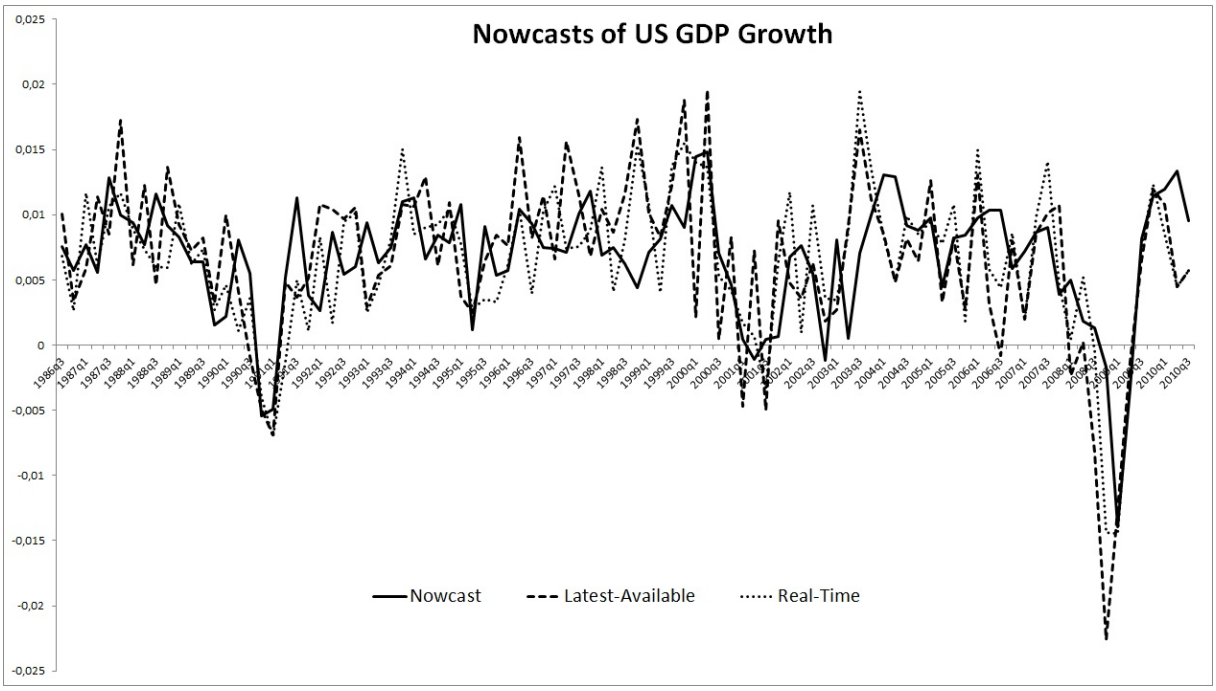


Figure 6: Nowcasts (solid line), latest-available (dashed line) and real-time data (dotted line) of US GDP growth

# 7 Conclusion

This paper combines the issues of dealing with variables that are sampled at mixed-frequencies and working with real-time data sets. As such, the analysis of Stark and Croushore (2002) is extended to an ADL setting where the regressors are possibly sampled at higher frequencies than the regressand. Thereby, a wider range of models could be considered by the practitioner and it is proposed to assess the superiority of one model over the other in terms of real-time forecasting by means of the repeated observations forecasting (ROF) approach introduced in Stark and Croushore (2002).

In particular, density forecasts are constructed for each model summarizing the extent to which forecasts of a particular calendar date differ across vintages. A combination of the forecast densities is constructed and the implied weights for each calendar date are investigated. These ideas are illustrated by means of an empirical analysis where the US GDP quarterly growth rate is fore- and nowcasted.

It is found that forecasts as well as the real-time realizations vary immensely across vintages. Employing a Diebold-Mariano (1995)-type test to compare forecast densities it is seen that no model outperforms all the others across all calendar dates. Nevertheless, a combination approach leads to a combined forecast density being able to mimic the real-time realizations

quite well for the dates concerned. The implied weights given to each model's forecast density, however, are not constant across time such that we propose a time-varying weight structure. This weight assignment is an alternative option for practitioners aiming to conduct forecast exercises using the latest-available data release and performs rather well in this particular study.

Working with high-frequency regressors gives rise to the use of high-frequency vintages in order to make use of the extra information inherent in the high-frequency variables. To this end, the multivariate dimension ROF is introduced. It is found that no vintage yields considerably better forecasting performance than the others. Allowing for more high-frequency regressors is, however, likely to lead to the outcome that employing new information as it becomes available will improve on forecasting performances.

This analysis should be extended along several lines. Most importantly, the set of regressors and models should be increased. Including variables of different and higher frequencies might lead to new insights into the multivariate dimension ROF and the question of which vintages to employ. Additionally, it might lead to more real-time forecasts on which to base our weight construction on. Furthermore, it will be interesting to see whether the ROF-based weights will behave in a systematic manner if more models and/or high-frequency regressors are used. This would give our time-varying ROF-based weights a more intuitive and founded background.

Our analysis shows that the time-varying ROF-based weights may present a competitive alternative to existing combination schemes such as e.g. Bates and Granger (1969) weights. Their construction, however, is dependent on the existence of a real-time data set. Hence, as Fernandez and Swanson (2009), this work stresses the need to make real-time data sets available to empirical researchers and appreciates recent developments in this matter (e.g. Statistics Canada or Reserve Bank of New Zealand).

## References


[1] Aastveit, K.A., Gerdrup, K.R., Jore, A.S., and Thorsrud, L.A. (2011), "Nowcasting GDP in Real-Time: A Density Combination Approach", Working Paper, fra Norges Bank, 2011/11.

[2] Andreou, E., Ghysels, E., and Kourtellos, R. (2010a), "Regression models with mixed sampling frequencies", *Journal of Econometrics*, 158(2), 246-261.

[3] Andreou, E., Ghysels E., and Kourtellos, R. (2010b), "Should macroeconomic forecasters use daily financial data and how?", mimeo Rimini Centre for Economic Analysis.

[4] Bańbura, M., Giannone, D., and Reichlin, L. (2011a), "Nowcasting", In *Oxford Handbook on Economic Forecasting*, ed. by M.P. Clements, and D.F. Hendry, pp. 63-90, Oxford University Press.

[5] Bańbura, M., Giannone, D., and Reichlin, L. (2011b), "Nowcasting with daily data", mimeo European Central Bank.

[6] Bates, J.M., and Granger, C.W.J. (1969), "The combination of forecasts", *Operations Research Quarterly*, 20, 451-468.

[7] Clements, M. P., and A. B. Galvão (2011), "Forecasting with Vector Autoregressive Models of Data Vintages: US output growth and inflation", *International Journal of Forecasting*, Forthcoming.

[8] Croushore, D., and Stark, T. (2001), "A real-time data set for macroeconomists", *Journal of Econometrics*, 105(1), 111-130.

[9] Croushore, D. (2006), "Forecasting with Real-Time Macroeconomic Data", In *Handbook of Economic Forecasting*, 1 (eds G. Elliott, C.W.J. Granger and A. Timmermann), 961-982. Amsterdam: North-Holland.

[10] Croushore, D. (2011), "Frontiers of Real-Time Data Analysis", *Journal of Economic Literature*, 49(March), 72-100.

[11] Diebold, F.X., and Lopez, J.A. (1996), "Forecast evaluation and combination", In *Handbook of Statistics*, 14 (eds G.S. Maddala and C.R. Rao), 241-268, Amsterdam: North-Holland.

[12] Diebold, F.X., and Mariano, R.S. (1995), "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, 13(July), 253-263.

[13] Diks, C., Panchenko, V., Sokolinskiy, O., and van Dijk, D. (2011a), "Comparing the Accuracy of Copula-Based Multivariate Density Forecasts in Selected Regions of Support", mimeo University of Amsterdam.

[14] Diks, C., Panchenko, V., and van Dijk, D. (2011b), "Likelihood-based scoring rules for comparing density forecasts in tails", *Journal of Econometrics*, 163, 215-230.

[15] Engle, R. F., and Granger, C. W. J. (1987), "Co-integration and error correction: Representation, estimation, and testing", *Econometrica*, 55(2), 251-76.

[16] Evans, M.D.D. (2005), "Where Are We Now? Real-Time Estimates of the Macroeconomy", *International Journal of Central Banking*, 1(2), 127-175.

[17] Fernandez, A., and Swanson, N.R. (2009), "Real-Time Datasets really do make a Difference: Definitional Change, Data Release, and Forecasting", Departmental Working Papers 201113, Rutgers University, Department of Economics.

[18] Foroni, C., Marcellino, M., and Schumacher, C. (2012), "U-MIDAS: Midas regressions with unrestricted lag polynomials", CEPR Discussion Paper No. DP8828.

[19] Ghysels, E. Santa-Clara, P., and Valkanov, R. (2004), "The midas touch: Mixed data sampling regression models", CIRANO Working Papers No. 2004s-20.

[20] Ghysels, E., Sinko, A., and Valkanov, R. (2007), "Midas regressions: Further results and new directions", *Econometric Reviews*, 26(1), 53-90.

[21] Ghysels, E., and Valkanov, R. (2006), "Linear time series processes with mixed data sampling and midas regression models", mimeo University of North Carolina.

[22] Giacomini, R., and White, H. (2006), "Tests of conditional predictive ability", *Econometrica*, 74, 1545-1578.

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

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

[25] Götz, T., Hecq A., and Urbain, J.-P. (2011), "Forecasting Mixed-Frequency Time Series With ECM-MIDAS Models", Maastricht University METEOR Research Memorandum 12/012.

[26] Greene, W.H. (2008), *Econometric Analysis* (6th ed.), Pearson Prentice Hall.

[27] Hall, S.G., and Mitchell, J. (2007), "Combining density forecasts", *International Journal of Forecasting*, 23(1), 1-13.

[28] Harvey, D., Leybourne, S., and Newbold, P. (1997), "Testing the equality of prediction mean squared errors", *International Journal of Forecasting*, 13 (June), 281-291.

[29] Hecq, A., and Jacobs, J.P.A.M. (2009), "On the VAR-VECM Representation of Real Time Data", mimeo Maastricht University.

[30] Jacobs, J.P.A.M., Sturm, J.-E., and van Norden, S. (2010), "Modeling Multivariate Data Revisions", mimeo University of Groningen.

[31] Jore, A.S., Mitchell, J., and Vahey, S.P. (2010), "Combining forecast densities from vars with uncertain instabilities", *Journal of Applied Econometrics*, 25(4), 621-634.

[32] Kuzin, V.N., Marcellino, M., and Schumacher, C. (2009), "MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area", Deutsche Bundesbank, Research Centre, Discussion Paper Series 1: Economic Studies No. 07/2009.

[33] Marcellino, M. (1999), "Some consequences of temporal aggregation in empirical analysis", *Journal of Business & Economic Statistics*, 17(1), 129-36.

[34] Marcellino, M., and 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.

[35] Mitchell, J., and Hall, S.G. (2005), "Evaluating, comparing and combining density forecasts using the klic with an application to the bank of England and niesr 'fan' charts of inflation", *Oxford Bulletin of Economics and Statistics*, 67(s1), 995-1033.

[36] Stark, T., and Croushore, D. (2002), "Forecasting with a real-time data set for macroeconomists", *Journal of Macroeconomics*, 24, 507-531.

[37] Stock, J.H., and Watson, M. (2004), "Combination forecasts of output growth in a seventcountry data set", *Journal of Forecasting*, 23, 405-430.

[38] Timmermann, A. (2006), "Forecast Combinations", In G. Elliott, C.W.J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, 1, 136-196, Amsterdam: Elsevier.

[39] Une, M.Y. (2011), "Forecasting Final Data Revisions in Real Time", mimeo Queen Mary University of London.