

Combining distributions of real-time forecasts: An application to U.S. growth

Citation for published version (APA):

Götz, T. B., Hecq, A. W., & Urbain, J. R. Y. J. (2014). Combining distributions of real-time forecasts: An application to U.S. growth. (GSBE Research Memorandum; No. 027). Maastricht: GSBE.

Document status and date:

Published: 01/01/2014

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Thomas B. Götz, Alain Hecq,
Jean-Pierre Urbain

**Combining Distributions of
Real-Time Forecasts: An
Application to U.S. Growth**

RM/14/027
(RM/12/021-revised-)

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O. Box 616
NL- 6200 MD Maastricht
The Netherlands

COMBINING DISTRIBUTIONS OF REAL-TIME FORECASTS: AN APPLICATION TO U.S. GROWTH*

Thomas B. Götz[†] Alain Hecq
Jean-Pierre Urbain

Maastricht University, SBE, Department of Quantitative Economics

July 7, 2014

Abstract

We extend the repeated observations forecasting (ROF) analysis of Stark and Croushore (2002) to allow for regressors of possibly higher sampling frequencies than the regressand. For the U.S. GNP quarterly growth rate, we compare the forecasting performances of an AR model with several mixed-frequency models among which is the MIDAS approach. Using the additional dimension provided by different vintages we compute several forecasts for a given calendar date and subsequently approximate the corresponding distribution of forecasts by a continuous density. Scoring rules are then employed to construct combinations of them and analyze the composition and evolution of the implied weights over time. Using this approach, we not only investigate the sensitivity of model selection to the choice of which data release to consider, but also illustrate how to incorporate revision process information into real-time studies. As a consequence of these analyses, we introduce a new weighting scheme that summarizes information contained in the revision process of the variables under consideration.

JEL Codes: C32, C53

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

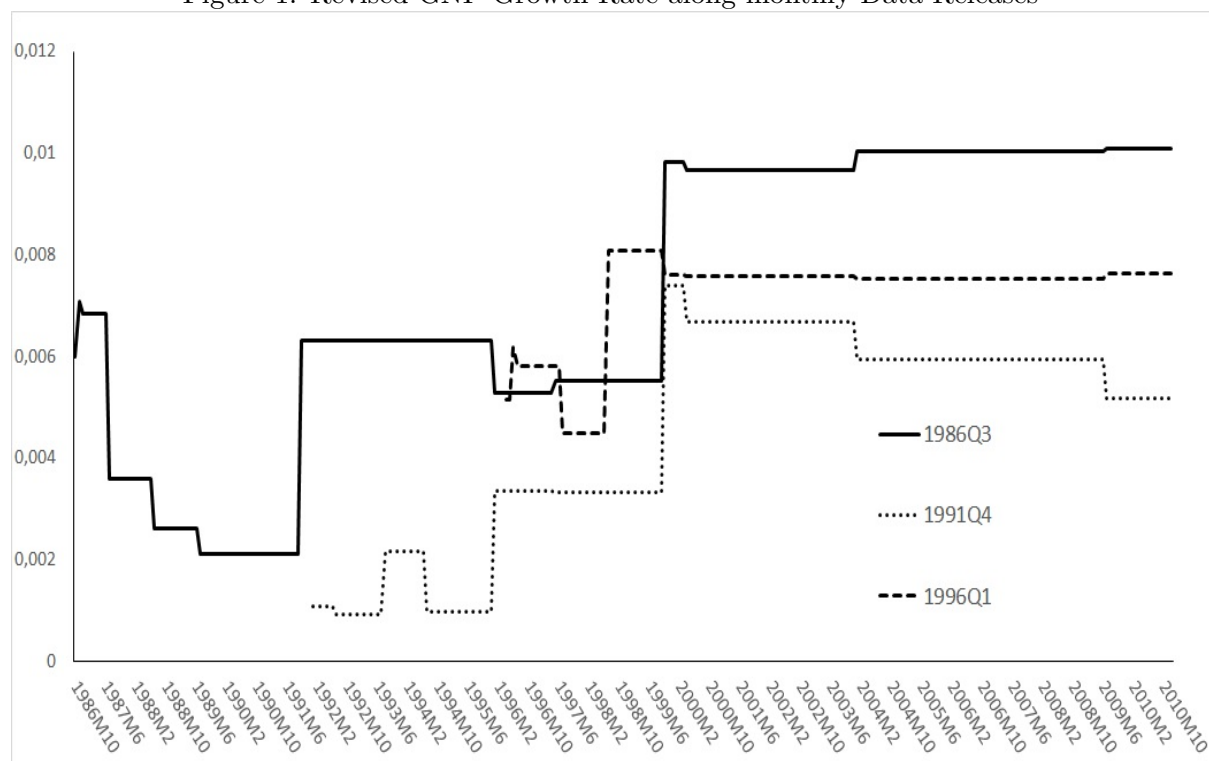
*We thank Franz Palm, Gianni Amisano and participants of the CIRANO workshop on Data Revision in Macroeconomic Forecasting and Policy in Montreal 2011, the 66th European Meeting of the Econometric Society in Malaga 2012, the 2012 Conference on Real-Time Data Analysis, Methods, and Applications in Philadelphia, the 7th Netherlands Econometric Study Group Meeting in Groningen 2012 and the 34th International Symposium on Forecasting in Rotterdam 2014 for useful suggestions and comments on earlier versions of the paper. The usual disclaimer applies.

[†]Correspondence to: Thomas B. Götz, Maastricht University, School of Business and Economics, Department of Quantitative Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: t.gotz@maastrichtuniversity.nl, Tel.: +31 43 388 3578, Fax: +31 43 388 20 00

1 Introduction

In economics, especially macroeconomics, many variables are subject to revisions. Take, for example, the quarterly growth rate of the U.S. real gross national product (GNP hereafter), for which a revised value is published each month.¹ Figure 1 displays the evolution of U.S. real GNP growth for three dates, 1986Q3, 1991Q4 and 1996Q1.² Note that even after 20 quarters there are still a lot of revisions taking place in the growth rates. Apart from many large movements being caused by so-called major revisions, or comprehensive benchmarks (Jacobs et al., 2013), also many smaller adjustments are found for all three dates, even after a large amount of data releases have already occurred.

Figure 1: Revised GNP Growth Rate along monthly Data Releases



Note: The GNP growth rates in 1986Q3, 1991Q4 and 1996Q1 are displayed for all applicable data releases until 2010Q4. Data releases progress on the x-axis, whereas values of the GNP growth rates are represented on the y-axis.

However, instead of making use of the entire history of a revised variable, i.e., instead of

¹Data releases for the nowadays more commonly modeled gross domestic product are available much later than for GNP, which is why we opted for the latter in this work.

²The data for this chapter were extracted from the databank ALFRED[®] of the Federal Reserve Bank of St. Louis: <http://alfred.stlouis.org>.

using a real-time data set, researchers often solely rely on the latest-available data set, i.e., the values represented by the right end points of the lines in Figure 1. In doing so, a data set is used that is different from the one researchers could have used in real time (Croushore and Stark, 2001, Croushore, 2011). This can have a drastic impact on the development of economic specifications, though. Stark and Croushore (2002) investigate how model selection, i.e., the lag order in autoregressive models, changes as different data releases, or vintages, are considered. Also, measures of forecast accuracy may be deceptively lower when latest-available data are dealt with than when real-time data are used. This makes many of the 'horse races' (Diebold, 2012), in which different models compete for the lowest root mean squared error (RMSE), questionable.

The contribution of this paper is threefold. First, in view of the previous argument, we analyze whether model selection outcomes based on latest-available data are robust to the vintage of data that is employed. To this end we extend the repeated observation forecasting approach proposed by Stark and Croushore (2002) to an autoregressive distributed lag (ADL hereafter) setting. Second, the construction and evaluation of densities of vintage-specific forecasts allow us to measure the sensitivity of these predictions to different data releases. By analyzing this issue, we consider and measure data uncertainty, a feature not often dealt with in the literature, contrary to parameter and model uncertainty. Third, using combinations of the aforementioned densities, we propose a weighting scheme that incorporates information contained in the revision process of the variables under consideration. These weights can then be used for fore- or nowcasting in real time.³

The rest of the paper is organized as follows. Section 2 contains the notation for a mixed-frequency real-time data set, a description of the repeated observation forecasting approach used in our particular framework and the methodology on how to construct densities of forecasts. Subsequently, the aforementioned techniques are illustrated with empirical data. In Section 3 we describe how to combine densities of forecasts based on scoring rules, whereby we discuss how to compute the associated weight for each model under consideration. Again, we continue by applying these methods to actual data. In Section 4 we show how to use the previous techniques in a real time study. Section 5 concludes.

To avoid confusion from the outset, a 'real-time data set' describes a data set that contains variables, which are subject to revisions. The term 'real time' alone, however, refers to using, at each moment in time, only data that has been available at that moment in time. In other words, one may consider only data that is released at or prior to the moment of time under consideration (Clements and Galvão, 2011).

³Note that we present an application to the U.S. growth rate involving a very restricted set of models (listed in Appendix A) and variables. Nevertheless, the analysis itself and the corresponding main conclusions can be applied and generalized to any other viable setting, i.e., one in which some (or all) variables are subject to revisions.

2 Extending the ROF Approach

2.1 Notation

We observe regularly spaced monthly vintages available for T and $T \times m_j$ ($j = 1, \dots, K$) observations of a low-frequency, here quarterly, variable y and K high-frequency variables x_1, \dots, x_K . A vintage is defined to be a moment at which a series is published, for instance 20th of June 2011. Let us denote the first and last available vintages by V_F and V_L , respectively. 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 . Likewise, the number of data releases within each low-frequency period can be defined as m_v . Due to the low frequency being quarterly and vintages occurring on a monthly basis, we will have $m_v = 3$ here.

Assuming a publication lag of one month for both, the regressors and the regressand,⁴ y_{t-1}^v , $v > t-1$, denotes the $(t-1)$ -value of y published in vintage v . The notation becomes more evolved for the high-frequency variables, where the j -index is suppressed now for explanatory convenience. We use $x_{t-i/m}^v$, $v > t - i/m$, where i selects the corresponding high-frequency observation. To be more precise, if $i = 0$, x_t^v represents the value of x at the end of the low-frequency period t (in vintage v). Consequently, the first high-frequency observation in period t is stored in $x_{t-(m-1)/m}^v$. Trivially, $x_{t-m/m}^v \equiv x_{t-1}^v$. This notation has become standard in the mixed-frequency literature (Clements and Galvão, 2008, 2009 or Götz et al., 2013, 2014) and is here extended to a real-time data set.⁵ Finally, with vintages potentially appearing on a high-frequency basis, v above can be a fraction similar to the subscript of the high-frequency regressors.

Considering a quarter/month-example, $x_{t-1-1/3}^{t-2/3} = x_{t-4/3}^{t-2/3}$ denotes the figure published in the first month of quarter t of the value of x two months earlier, i.e., in the middle month of quarter $t-1$. Table 1 illustrates the notation for a quarter/month-example, i.e., for $m = 3$.

2.2 Repeated Observation Forecasting

A common practice in empirical work is to use the latest-available data set to evaluate forecasts. This means that at period T one collects the historical time series for y_{t-1}^T where $t = 2, \dots, T$. Subsequently, a one-step-ahead point forecast for \hat{y}_T may be obtained. Stark and Croushore (2002) proposed an interesting graphical tool, which they call repeated observations forecasting, to evaluate whether the forecast accuracy is sensitive to the vintage chosen. For our variables y and X that are available at different frequencies, the approach works as follows. Let us take a particular low-frequency period t^* , where $\lceil V_F - 1 \rceil \leq t^* \leq \lceil V_L - 1 \rceil$, because no data releases

⁴Indeed, for many macroeconomic variables, e.g., GNP, the first figure in quarter t becomes available the first month of quarter $t+1$. Likewise, we have to wait for the coming month to get a first estimate of a regressor such as the industrial production index in the current month. For regressors that are not revised no publication delay exists in practice. Nevertheless, in order to keep the setup generic, we treat such regressors as being revised over time as well and pretend a publication lag of one month to exist here as well.

⁵High-frequency variables are usually labeled as $x_{t-i/m}^{(m)}$ in the mixed-frequency literature. As the superscript is, however, already reserved for the data releases, we deviate from the standard notation slightly in this respect.

Table 1: Revision Triangle for Mixed-Frequency Data

$Vintage$ $Time$	\dots	$t-2-1/3$	$t-2$	$t-1-2/3$	$t-1-1/3$	$t-1$	$t-2/3$	$t-1/3$	t
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$t-2-2/3$	$x_{t-2-2-1/3}^{t-2}$	$x_{t-2-2/3}^{t-2}$	$x_{t-2-2-2/3}^{t-2}$	$x_{t-1-2/3}^{t-1}$	$x_{t-1-1/3}^{t-1}$	$x_{t-2-2/3}^{t-1}$	$x_{t-2-2/3}^{t-2}$	$x_{t-2-2/3}^{t-1/3}$	$x_{t-2-2/3}^t$
$t-2-1/3$	$x_{t-2-2-1/3}^{t-2}$	$x_{t-2-1/3}^{t-2}$	$x_{t-2-2-1/3}^{t-2}$	$x_{t-2-1/3}^{t-1}$	$x_{t-1-1/3}^{t-1}$	$x_{t-2-1/3}^{t-1}$	$x_{t-2-1/3}^{t-2}$	$x_{t-2-1/3}^{t-1/3}$	$x_{t-2-1/3}^t$
$t-2$	y_{t-2}^{t-2}	y_{t-2}^{t-2}	y_{t-2}^{t-2}	y_{t-2}^{t-2}	y_{t-2}^{t-2}	y_{t-2}^{t-2}	y_{t-2}^{t-2}	y_{t-2}^{t-2}	y_{t-2}^t
$t-1-2/3$				$x_{t-1-2/3}^{t-1}$	$x_{t-1-1/3}^{t-1}$	$x_{t-2-2/3}^{t-1}$	$x_{t-1-2/3}^{t-2}$	$x_{t-1-2/3}^{t-1/3}$	$x_{t-1-2/3}^t$
$t-1-1/3$				$x_{t-1-1/3}^{t-1}$	$x_{t-1-1/3}^{t-1}$	$x_{t-1-1/3}^{t-1}$	$x_{t-1-1/3}^{t-2}$	$x_{t-1-1/3}^{t-1/3}$	$x_{t-1-1/3}^t$
$t-1$							y_{t-1}^{t-1}	y_{t-1}^{t-1}	y_{t-1}^t
$t-2/3$							$x_{t-2/3}^{t-2}$	$x_{t-2/3}^{t-1/3}$	$x_{t-2/3}^t$
$t-1/3$							$x_{t-1/3}^{t-1}$	$x_{t-1/3}^{t-1/3}$	$x_{t-1/3}^t$
t									

Note: A quarter/month-example ($m = 3$) of a mixed-frequency real-time data set is depicted. x is the high-frequency (monthly) variable, whereas y is the low-frequency (quarterly) variable. Both variables are updated on a monthly basis, i.e., $m_v = 3$. A publication delay of one month is assumed for both, x and y .

are available prior to V_F , and consider the historical series $y_{t-1}^v, t = 2, \dots, t^*$, and X_{t-i/m_j}^v with $t = 1, \dots, t^*, i = 0, \dots, m_j - 1$,⁶ for a set of monthly vintages $v = t^* + 1/m_v, \dots, V$. Note that not necessarily $V = V_L$ if one wants to base the ROF analysis on, e.g., a balanced or lower number of vintages. Given a certain model, a one-step-ahead point forecast for $\hat{y}_{t^*}^v$ can be computed for each vintage v such that we end up with a sequence of $V^* \equiv m_v(V - t^*)$, in this case one-step-ahead, forecasts for the same point.

Stark and Croushore (2002) report these forecasts using box-plot-like graphs. In contrast to their study we approximate the corresponding distribution of forecasts by a continuous density. Combining these model-specific densities implies weights for each model, that allow researchers

- (i) to assess whether model 'horse races' are vintage-sensitive, and
- (ii) to include information contained in the revision process into real-time studies.

Importantly, for (i) we set $V = V_L$ and thereby directly extend the analysis of Stark and Croushore (2002) by investigating the sensitivity of forecasts to the data release considered. For (ii), however, we fix $V = T^*$, where we denote by T^* the currently last available vintage. In other words, to mimic what a researcher could have done at a specific moment in time, we only consider data that is released no later than T^* , i.e., 'today' (Clements and Galvão, 2011). Note that in the analysis to come we only consider one-step-ahead forecasts, whereby an extension towards h -step-ahead (direct or indirect) forecasts is straightforwardly done.

Stark and Croushore (2002) discuss the issue of which observations to use as realizations, or 'actuals', in order to compare different methods graphically or to compute forecast errors. Depending on the choice of the realization, the outcome of an analysis may change substantially. Stark and Croushore (2002) consider three possibilities for the actual value, namely (a) the respective observation four quarters later, (b) the corresponding figure in the latest-available data set or (c) the first observation after a comprehensive benchmark has occurred. In this paper we propose a different set of observations as 'actuals', namely the realization in the respective vintage itself. To be more precise, the one-step-ahead forecast of y_{t^*} corresponding to vintage v , i.e., $\hat{y}_{t^*}^v$, is compared to $y_{t^*}^v$ and a forecast error is computed as $e_{t^*}^v = \hat{y}_{t^*}^v - y_{t^*}^v$. We refer to this set of 'actuals' as same-vintage realizations or same-vintage 'actuals'.

2.3 Constructing Densities of Forecasts

Given a model specification, we obtain a sequence of V^* one-step-ahead forecasts, i.e., $\hat{\mathbf{y}}_{t^*} = (\hat{y}_{t^*}^{t^*+1/m_v}, \dots, \hat{y}_{t^*}^V)'$, using the ROF approach as described above. Remember that $V = V_L$, i.e., the last data release of our data set, or $V = T^*$, i.e, the last available vintage in a real-time analysis, depending on whether we analyze case (i) or (ii) above (see Sections 3.4 and 4, respectively). The distribution of these forecasts is then approximated by a continuous density, i.e., \hat{f}_{t^*} , via kernel density estimation techniques. To be more precise, assuming the V^* one-step-ahead forecasts to have some underlying unknown distribution f , the kernel density estimator

⁶Here we assume that the value of X is published before the one of y , which is reasonable for most macroeconomic mixed-frequency data sets.

is

$$\hat{f}_{t^*}(y) = \frac{1}{(V^*)h} \sum_{i=t^*+1/m_v}^V K\left(\frac{y - \hat{y}_{t^*}^i}{h}\right), \quad (1)$$

where $K(\cdot)$ is the kernel function and h is the bandwidth parameter. In this paper, the Gaussian kernel, $K(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$ is considered. The bandwidth is chosen of the order $\hat{\sigma}(V^*)^{-1/5}$, where $\hat{\sigma}$ is an estimator of the standard deviation.⁷

2.4 Application I

Let us illustrate the techniques introduced thus far using actual data. To this end, we consider the quarterly real gross national product (ref. GNPC96 and referred to as *GNP* hereafter), seasonally adjusted, as dependent, low-frequency, variable y . The series is observed from 1960Q1 ($t = 1$) until 2010Q3 ($T = 203$).⁸ For the regressors we consider, for the same time span, the monthly seasonally adjusted industrial production index (ref. INDPRO and labeled *IPI* hereafter) and the daily S&P 500 stock index (ref. SP500, *SP* hereafter). Consequently, $m_{IPI} = 3$ and $m_{SP} = 60$.⁹ Note that all variables are measured in logarithms. There are monthly data releases from July 1986 ($V_F = 107 - 2/3$) to December 2010 ($V_L = 204$), which implies $m_v = 3$.¹⁰ *SP* 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. A final 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 analysis.

Before actually conducting the ROF analysis outlined in Section 2.2, let us give a short comment on the models we use for obtaining the forecasts, from which densities are constructed subsequently. We consider three models that contain high-frequency regressors and one autoregressive model that serves as a benchmark (Stark and Croushore, 2002). With respect to the former, two of the models temporally aggregate the high-frequency regressors prior to estimation, whereas the MI(xed) DA(ta) S(ampling), or MIDAS hereafter, regression model directly employs the high-frequency observations. For those three model specifications, we take into

⁷Note that the results are robust to the use of other kernels, e.g., the Parzen or Triangle kernel (see, for example, Greene, 2008).

⁸For this exercise we have extracted the series in February 2011

⁹In order to have the same number of working days per vintage we take the maximum amount of working days that is available in each month over the time period considered. This leads to 20 daily observations per month and, hence, 60 per quarter.

¹⁰The vintages do not match exactly for the three series. 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 *IPI*), 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 regressors. To facilitate the computation we create missing data releases by assigning missing values their respective values in the previous vintage. The alternative would have been to delete uncommon vintages.

account the possibility of cointegration between the variables by considering them with and without a cointegrating relationship (the former contains an 'a' in the label, the latter a 'b'). This amounts to 7 models, labeled *1a*, *1b*, *2a*, *2b*, *3a*, *3b* and 4 in Appendix A, where detailed descriptions of each specification are provided. Note that we abstract from other approaches such as factor models which present a popular choice when forecasting (or nowcasting) GNP growth rates (Giannone et al., 2008). Although the methodology presented in this paper can just as well be applied to any other model, also with a possibly large set of regressors as in Andreou et al. (2013), one has to obtain and clean a real-time data set for each regressor concerned, i.e., correct for possible mismatches or dealing with multiple vintages for one time period. This task may be very time-consuming or even impossible due to, e.g., non-existent data.

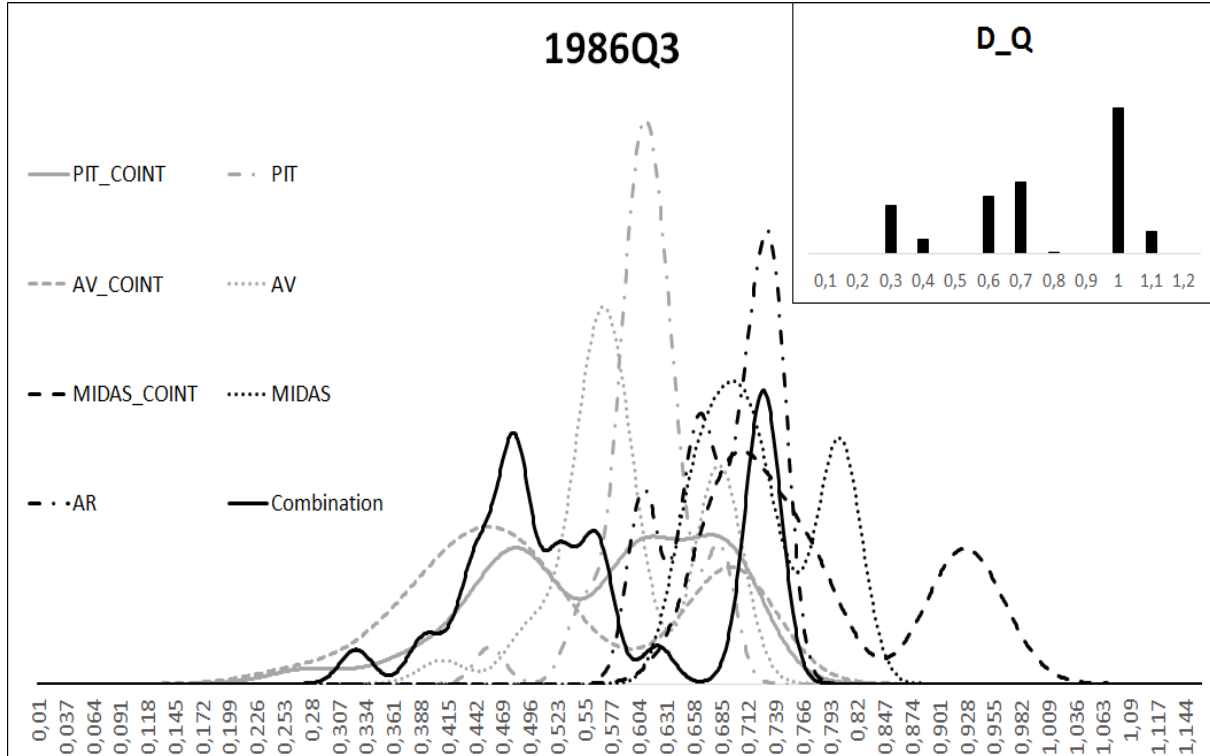
In order to analyze whether model comparisons are sensitive to the data release considered, let us conduct the aforementioned ROF analysis: We consider calendar dates $106 \leq t^* \leq 203$, corresponding to the range 1986Q2-2010Q3, and the historical series GNP_{t-1}^v , $t = 2, \dots, t^*$, $IPI_{t-i/3}^v$ with $t = 1, \dots, t^*$, $i = 0, 1, 2$, and $SP_{t-i/20}^v$ with $t = 1, \dots, t^*$, $i = 0, \dots, 19$ for a set of monthly vintages $v = t^* + 1/3, \dots, V_L$. For each model $i \in \{1a, 1b, 2a, 2b, 3a, 3b, 4\}$, a set of $V^* = 3(V_L - t^*)$ one-step-ahead point forecasts can be computed, i.e., $\hat{\mathbf{y}}_{t^*}^i = (\hat{y}_{t^*+1/3}^{t^*}, \dots, \hat{y}_{t^*+V_L}^{t^*})'$, which are used to construct densities of forecasts using (1).

Let us consider two individual dates $t^* = 107$ (1986Q3) and 128 (1991Q4) as illustrative examples.¹¹ Note that $V^* = 291$ and 228, respectively, for our example dates. Averaging (models *2a* and *2b*) is denoted by AV, Point-in-Time sampling (models *3a* and *3b*) by PIT and the same-vintage 'actuals' by D.Q. Whenever a long-run relationship is included, COINT is appearing in the label. The densities of forecasts produced by the respective seven models are plotted in Figures 2 and 3. Additionally, histograms of the same-vintage realizations are displayed in the top right corners. At this stage, we ask the reader to disregard the density displayed by a solid black line, which we get back to at a later stage in the paper.

Just as their corresponding realizations (see Figure 1), forecasts of the GNP growth rates differ substantially depending on which vintage we consider. This becomes apparent from the wide range of data the densities cover. Furthermore, the forecasts of each model result in quite a different density, i.e., the GNP growth rate forecasts differ not only with respect to the data vintage (something we call 'data uncertainty'), but also the model (model uncertainty) under consideration. In the case of 1986Q3, for example, PIT_COINT produces a rather wide and flat density whereas the one of PIT is concentrated on a smaller region. Similar observations can be made in Figure 3.

¹¹In comparison to Figure 1 the outcomes for 1996Q1 are not presented to save on space, but are available upon request.

Figure 2: Densities of Forecasts and Same-Vintage 'Actuals' for 1986Q3



Note: This figure graphs the densities of forecasts corresponding to the seven forecasting approaches under consideration (see Appendix A). Values of the GNP growth rate (multiplied by 100) are displayed on the x-axis. The density displayed by a solid black line corresponds to the combination density with weights computed as in (4). Corresponding same-vintage 'actuals' are depicted in the histogram in the top right corner.

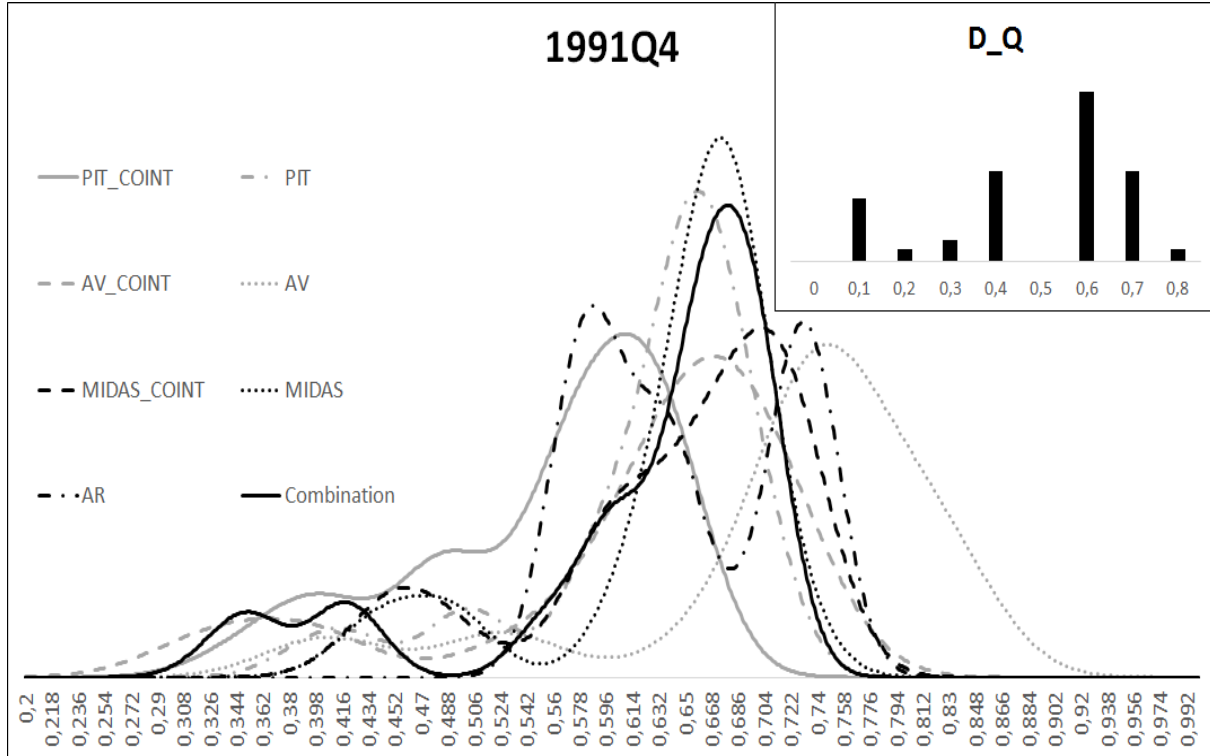
3 Sensitivity of Model 'Horse Races' to Vintages

3.1 Scoring Rules

To evaluate the densities of vintage-specific forecasts, we make use of scoring rules, an avenue often followed in this field of literature (see Amisano and Giacomini, 2007 or Diebold and Lopez, 1996). A scoring rule is a loss function whose value depends on the density, \hat{f}_{t^*} , and the vector of 'actuals', i.e., $\mathbf{y}_{t^*}^{\bar{V}} = (y_{t^*+1/m_v}^v, \dots, y_{t^*}^v)'$. Without going into too much detail, \bar{V} is not necessarily equal to V , i.e., we do not have to plug all same-vintage 'actuals' corresponding to $\hat{\mathbf{y}}_{t^*}$ into the scoring function to evaluate our densities. A scoring rule is hence denoted as $S(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{\bar{V}})$ and is constructed in such a way that a density mimicking the shape of the same-vintage 'actuals' better than another one is assigned a higher score.

To clarify, if the density for model A assigns a probability of (practically) zero to the occurrence of a particular same-vintage 'actual', i.e., if $\hat{f}_{t^*}^A(y_{t^*}^v) = 0$ for some v ($t^* + 1/m_v \leq v \leq \bar{V}$), whereas model B 's density assigns a positive probability to the same event, i.e.,

Figure 3: Densities of Forecasts and Same-Vintage 'Actuals' for 1991Q4



Note: See Figure 3

$\hat{f}_{t^*}^B(y_{t^*}^v) > 0$, then model B is assigned a higher score than model A for that particular same-vintage 'actual'. Indeed, $\hat{f}_{t^*}^A(y_{t^*}^v) = 0$ implies that the one-step-ahead forecasts of y_{t^*} generated by model A do not assign a positive probability to the event that the actual value of y_{t^*} equals $y_{t^*}^v$. Consequently, model A is not able to forecast that same-vintage 'actual' and receives a lower score than a model that places positive probability on that outcome to occur, i.e., model B . Doing this for each same-vintage realization under consideration yields a vector of scores for each model. We will, as Amisano and Giacomini (2007), employ the logarithmic scoring rule

$$S_l(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{\bar{V}}) = \log[\hat{f}_{t^*}(\mathbf{y}_{t^*}^{\bar{V}})]. \quad (2)$$

At this stage, let us argue why we opted for our same-vintage 'actuals' instead of, e.g., their final-vintage counterparts. If we had used one of the options described at the end of Section 2.2, the vector of 'actuals' would have consisted of only one element, e.g., $y_{t^*}^T$ with final-vintage 'actuals'. Employing same-vintage 'actuals' instead, the scoring vectors measure how well a model's forecasts resemble the revision process of a variable.

3.2 Combining Densities of Forecasts

It is well known in the literature for point (see, e.g., Timmermann, 2006) or density forecasting (Wallis, 2005) that combinations of forecasts may offer diversification gains making it attractive to combine the information present in the different forecasting models. Being equipped with densities produced by different models, a density combination approach is a natural avenue to follow (Wallis, 2005 or Aastveit et al., 2011).

Similar to many of the aforementioned studies we consider a so-called linear opinion pool, i.e., defining the combined density of forecasts as a linear combination of the individual densities:

$$\hat{f}_{t^*}^{Com}(y) = \sum_{i=1}^M \phi_i \hat{f}_{t^*}^i(y), \quad (3)$$

where M is the number of models to combine. When focusing on one date individually, we define the recursive log score weights of model $i = 1, \dots, M$ as

$$\phi_i = \frac{\exp(\bar{S}_{t^*}^i)}{\sum_{i=1}^M \exp(\bar{S}_{t^*}^i)}, \quad (4)$$

where,

$$\bar{S}_{t^*}^i = \frac{1}{\bar{V}^*} \sum_{v=t^*+1/m_v}^{\bar{V}} \log[\hat{f}_{t^*}^i(y_{t^*}^v)] = \frac{1}{\bar{V}^*} S_l(\hat{f}_{t^*}^i, \mathbf{y}_{t^*}^{\bar{V}})' \boldsymbol{\iota}_{\bar{V}^*},$$

where $\bar{V}^* = m_v(\bar{V} - t^*)$ and $\boldsymbol{\iota}_N$ denotes an $N \times 1$ vector of ones.

When interest lies not only in the weights for one date individually, but the models' ability to mimic the revision process over all or many consecutive dates in time, we propose to compute what we call revision-process-robust (RPR hereafter) weights. Note that these are adapted versions of the real-time optimal pool weights found in Amisano and Geweke (2013). As the name suggests, the latter can be interpreted as the optimal action a researcher can perform in real time. In our context, however, the RPR weights depend on the previously constructed densities of forecasts that may employ information of future data releases (if $V > T^*$), making the expression 'real-time' inappropriate. For each calendar date, the RPR weights summarize the extent to which the various models are able to resemble the revision process up until the date under consideration. In particular, the RPR weights at date t^* maximize the following criterion:

$$\mathbf{w}_{t^*}^{RPR} = \arg \max_{\mathbf{w}_{t^*}} \sum_{j=\lceil V_F - 1 \rceil}^{t^*} \log \left[\sum_{i=1}^M w_{i,t^*} \exp(\bar{S}_j^i) \right], \quad (5)$$

where \mathbf{w}_{t^*} is bound to be in the M -dimensional unit simplex. Note that the weights at any date t^* depend on the average scores of all previous dates (remember that $\lceil V_F - 1 \rceil$ is the first calendar date to consider). Consequently, they take into account how well the models resembled the revision process of previous dates. Stated differently, for each calendar date they find the optimal combination of forecasting models based on how they mimicked the revision process thus far.

3.3 Remark on Data Uncertainty

Until now we have almost exclusively used the term 'density of forecasts' and not 'forecast density'. The reason is, of course, that our densities are merely a nonparametric way to approximate the distribution of the one-step-ahead forecasts corresponding to each vintage. These are no density forecasts in the sense of Tay and Wallis (2000) in particular or the forecasting literature in general.

Consequently, we are only focusing on point forecasts itself and do not take parameter uncertainty into account. We, however, obtain a measure of data uncertainty, i.e., how much the revised observations differ from an underlying 'true' or final value. Let us introduce data uncertainty by setting $y_t = y_t^v - \mu_t^v$ with $\mu_t^v \sim N(0, \sigma_\mu^2)$ for all t . Note that we are not concerned with the news or noise issue (Mankiw and Shapiro, 1986 or Jacobs and van Norden, 2011), but merely claim that the true underlying value of y_t differs (potentially a lot) from the one observed at vintage v . Using our nonparametric distribution approximations we get an idea about the sensitivity of the point forecasts to different vintages and thereby a measure of data uncertainty.

An alternative approach would be to construct a forecast density for each vintage and subsequently combine these densities to deal with data uncertainty. This allows the researcher to analyze certain regions of the densities, not only their means as we do. We, however, leave such an approach for further research.

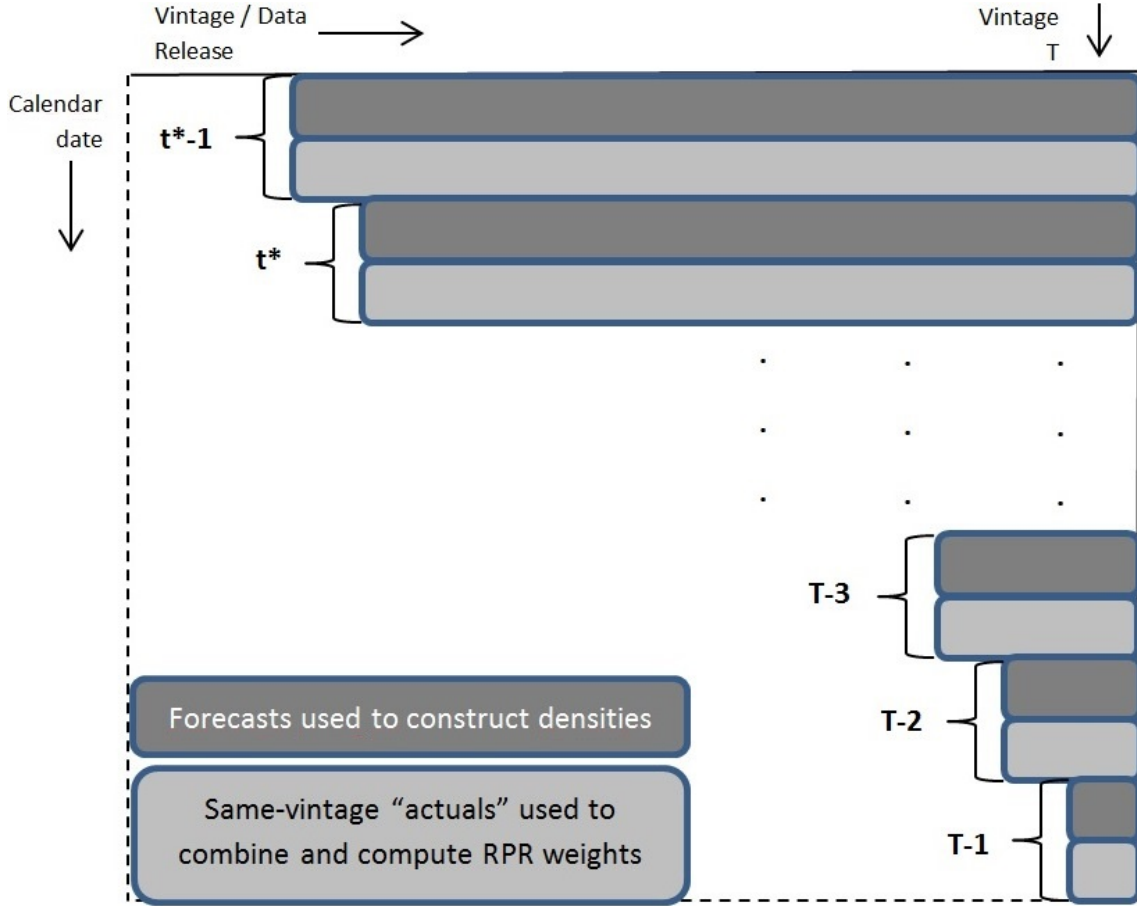
3.4 Application II

Let us return to our empirical analysis, which we started in Section 2.4. After having constructed densities of forecast for each model under consideration, we combine these densities using (3) and the weights in (4) or (5), depending on whether we look at, respectively, one calendar date individually or all of them consecutively. The scores used to evaluate the densities (and compute the corresponding weights) are based on all same-vintage 'actuals' that are at our disposal, i.e., $\bar{V} = V_L$ such that $S_l(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{\bar{V}}) = S_l(\hat{f}_{t^*}, \mathbf{y}_{t^*}^{V_L})$ in (2). Figure 4 illustrates the entire approach.

Before combining the model-specific densities for all calendar dates, let us focus on our two example dates, 1986Q3 and 1991Q1. Note that the weights in (4) are used to combine the densities in these cases, because we deal with individual calendar dates. The densities graphed by solid black lines in Figures 2 and 3 represent the corresponding combination densities. While each individual model-based density does a relatively poor job of mimicking the shape of the respective realizations distribution, the combination densities replicate their shapes quite well: In the case of 1986Q3 it mimics the tri-modality of the histograms and for 1991Q4 it detects the minimum at 0.005 as well as the maximum at roughly 0.0065 almost perfectly. As far as the weight composition is concerned, for $t^* = 107$ (1986Q3) the weights turn out to be 0.13, 0, 0.74, 0, 0.14, 0 and 0 for PIT_COINT, PIT, AV_COINT, AV, MIDAS_COINT, MIDAS and AR, respectively, whereas for $t^* = 128$ (1991Q4) they are 0.19, 0.02, 0.64, 0.13, 0.01, 0.01 and 0, respectively.

Note that the AR-model is assigned a zero weight for both dates. It turns out that, when

Figure 4: The ROF Approach illustrated



Note: The graph shows that for each date t^* under consideration a forecast is computed for each vintage available. Out of these forecasts (depicted as dark grey bars) densities are constructed using (1). These densities are evaluated using scores that are based on all same-vintage 'actuals' available (depicted as light grey bars). Note that the dark and light grey bars cover the same areas for each t^* visualizing that construction and evaluation of the densities of forecasts are based on all vintages at our disposal.

considering all calendar dates, the autoregressive model is very often outperformed by a majority of the competing approaches. Hence, adding explanatory variables seems to improve the forecasting performance across different vintages. This is mainly due to the weakness of the AR model not to incorporate mid-quarter information. As an example, *GNP* grew from 2000Q3 until 2001Q2 followed by a decline in 2001Q3. The autoregressive model, however, is not able to predict this decline whereas other models, taking into account mid-quarter information, are better able to do so. The relatively poor performance of the AR-model is, of course, likely to change for multi-step-ahead forecasts because regressors would have to be forecasted as well.

So far, we have merely looked at two individual dates. Let us inspect how the weight

distribution over our seven models evolves over all calendar dates by computing the RPR weights in (5). Figure 5 graphs the evolution of the RPR weights associated with our seven models.

It turns out that all models except AV possess non-zero RPR weights. It is striking that while the AR-model had a weight of up to 33% assigned in the beginning of the analysis, its weight quickly dropped to 5 to 10% where it remained until the end. The model with the largest average weight over the whole period is PIT_COINT (45%), followed by MIDAS_COINT (15%), MIDAS and AV_COINT (both 12%). Interestingly, MIDAS_COINT has a much larger average weight in the first half of the graph than in the second half, whereas the opposite is true for the MIDAS approach without a cointegrating term.

Note that the RPR weights graphed in Figure 5 enable us to rank the models under consideration. More importantly, such a ranking would take into account how well the each model's forecasts resembled the revision process of the variable of interest. In order to investigate whether model rankings are sensitive to the vintage chosen, we should compare the ranking implied by the RPR weights in Figure 5 with one that is based on final-vintage data only. To this end, we conduct the following forecasting exercise: For each model and using only data from the latest-available vintage, we compute 20 one-step-ahead out-of-sample forecasts of the U.S. growth rate for two different time periods, (1) 1995Q1-1999Q4 and (2) 2005Q4-2010Q3. Subsequently, the models are ranked based on their RMSEs. The results are summarized in Table 2.

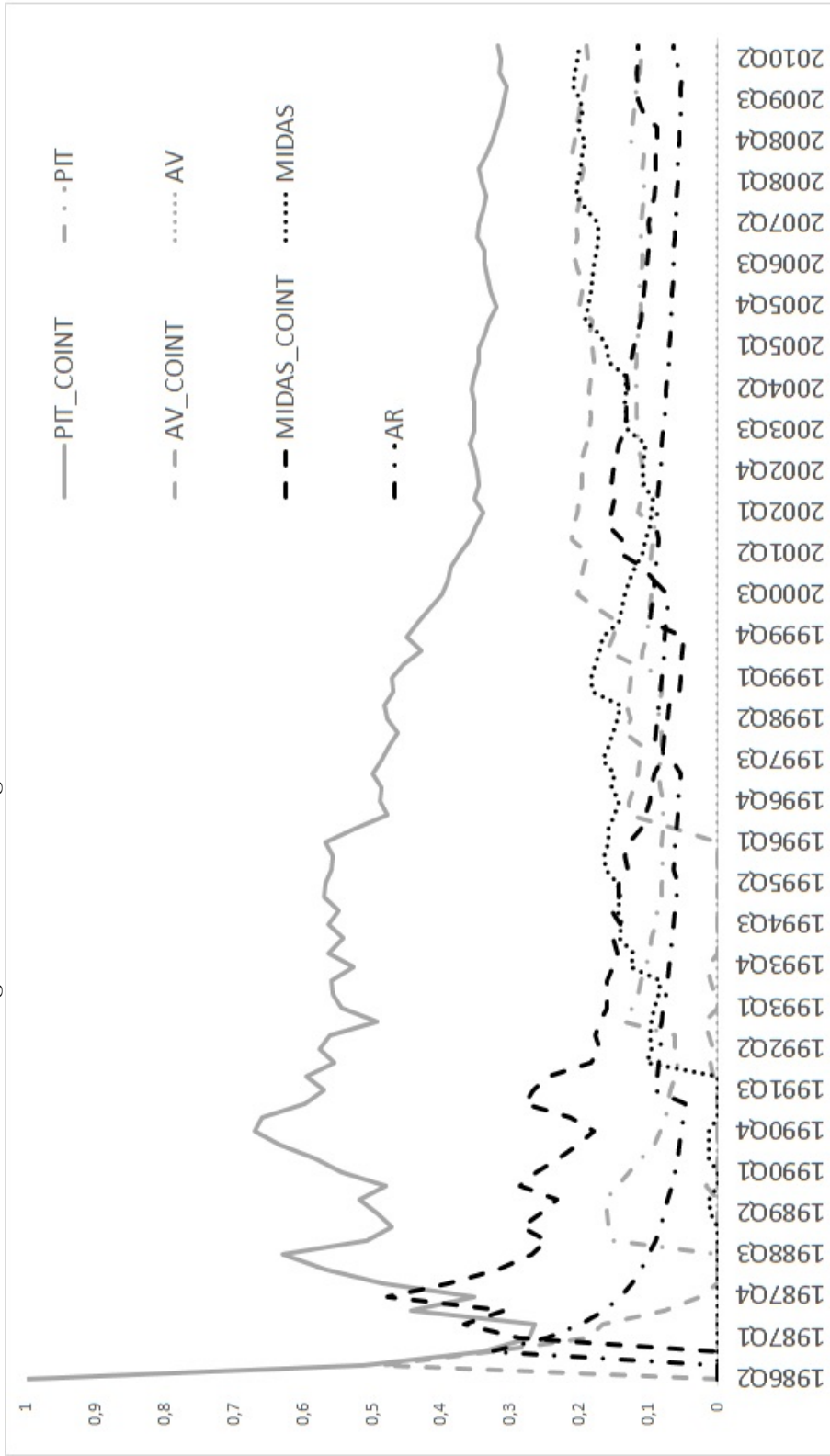
Table 2: RMSEs and Model Ranks using Final-Vintage Data

Model	(1) 1995Q1 - 1999Q4		(2) 2005Q4 - 2010Q3	
	RMSE	Rank	RMSE	Rank
PIT_COINT	0.0205	4	0.0219	1
PIT	0.0206	5	0.0230	4
AV_COINT	0.0233	7	0.0220	2
AV	0.0232	6	0.0229	3
MIDAS_COINT	0.0182	1	0.0275	6
MIDAS	0.0185	2	0.0275	6
AR	0.0193	3	0.0377	7

Note: The figures represent the RMSEs corresponding to a forecasting exercise conducted using only the final data release. 20 one-step-ahead out-of-sample forecasts of U.S. GNP growth were computed using seven different models for the time periods (1) 1995Q1-1999Q4 and (2) 2005Q4-2010Q3. Subsequently, the RMSEs and associated ranks were computed.

To obtain the ranking implied by the RPR weights in Figure 5, we rely on the average RPR weight of each model during the periods in question, (1) 1995Q1-1999Q4 and (2) 2005Q4-2010Q3. The outcomes are presented in Table 3. It turns out that several models appear to result in much better forecasts when only final-vintage data instead of all data releases are used, i.e., their ranking implied by the RPR weights is lower than it was in Table 2. Examples are MIDAS_COINT and AR for period (1) and AV for period (2). Conversely, some models get

Figure 5: RPR Weights for all Calendar Dates



Note: The figure displays the RPR weights assigned to the seven models for all the calendar dates in the sample. The different dates are represented on the horizontal axis, whereas the weight itself is plotted on the vertical axis.

assigned a deceptively low ranking based on latest-available data only, i.e., their average RPR weight suggests a higher ranking than it was in Table 2 (see PIT_COINT and AV_COINT for period (1) or MIDAS for period (2)).

Table 3: Model Ranks implied by RPR Weights

Model	(1) 1995Q1 - 1999Q4		(2) 2005Q4 - 2010Q3	
	<i>RPR weight</i>	Rank	<i>RPR weight</i>	Rank
PIT_COINT	0.498	1	0.329	1
PIT	0.086	5	0.113	4
AV_COINT	0.093	3	0.2	2
AV	0.001	7	0.001	7
MIDAS_COINT	0.092	4	0.103	5
MIDAS	0.159	2	0.194	3
AR	0.071	6	0.06	6

Note: The figures represent the average RPR weights (see Figure 5) for the time periods (1) 1995Q1-1999Q4 and (2) 2005Q4-2010Q3 and the associated ranks of seven different models.

Hence, model 'horse races' are sensitive to the data vintage employed. Tables 2 and 3 illustrate that the ranking of several models based on RMSEs computed using only final-vintage data differs from the ranking implied by the RPR weights, which are based on all data releases available to us. To be more precise, using only latest-available data, some models seem to possess a better (or worse) forecasting performance than when their forecast accuracy is evaluated also using past data releases. In this sense, our results extend the work of Stark and Croushore (2002) who concluded for autoregressive models that "measures of forecast error, such as root-mean-squared error or mean absolute error, can be deceptively lower when using latest-available data [...]" (p. 507).¹²

4 Incorporating Revision Process Information into Real-Time Studies

The results presented thus far do not provide guidance as to what a researcher should do in real time. Let us assume that we are at month T^* such that, coherent with Section 2.2, the most up-to-date data set is released at $v = T^*$. Now, we consider calendar dates t^* , where $(1986Q2 =)106 \leq t^* \leq \lceil T^* - 1 \rceil$, and again deal with the series $GNP_{t-1}^v, t = 2, \dots, t^*$,

¹²When considering more recent dates, the construction of our densities is based on only a few vintages, i.e., V^* is rather small. However, the RPR weights are based on the average scores of all previous time periods, and are thereby smoothing the weights corresponding to individual dates. In other words, the effect of basing the densities corresponding to more recent dates on only a small number of vintages is mitigated by constructing the RPR weights as cumulated weighted averages of the individual weights in (4). If one still suspects negative effects of the aforementioned decreased number of data releases, one could either forecast weights using a VAR model or forecast revisions as done in Clements and Galvão (2011).

$IPI_{t-i/3}^v$ with $t = 1, \dots, t^*, i = 0, 1, 2$, and $SP_{t-i/20}^v$ with $t = 1, \dots, t^*, i = 0, \dots, 19$ for vintages $v = t^* + 1/3, \dots, T^*$, i.e., $V = T^*$. Similar to before, we compute $V^* = 3(T^* - t^*)$ one-step-ahead forecasts for each model in order to construct densities of forecasts using (1). The scores used to evaluate and combine the model-specific densities of forecasts are, however, only based on maximally the first 12 (equivalent to 1 year) same-vintage 'actuals'. In other words, $\bar{V} = \min\{t^* + 4, T^*\}$ such that the vector of 'actuals' is written as

$$y_{\bar{V}} = \begin{cases} (y_{t^*+1/3}^{t^*+1/3}, \dots, y_{t^*+4}^{t^*+4})' & \text{if } t^* + 4 < T^* \\ (y_{t^*+1/3}^{t^*+1/3}, \dots, y_{t^*}^{T^*})' & \text{else.} \end{cases}$$

To clarify, we construct our densities of forecasts using as many data releases as possible in order to deal with a reliable kernel density estimation. However, when it comes to evaluating and combining the model-based densities we only check how well the models predict the first 12 releases of the variable of interest, i.e., when data is still revised a lot. Figure 6 illustrates the approach just outlined.

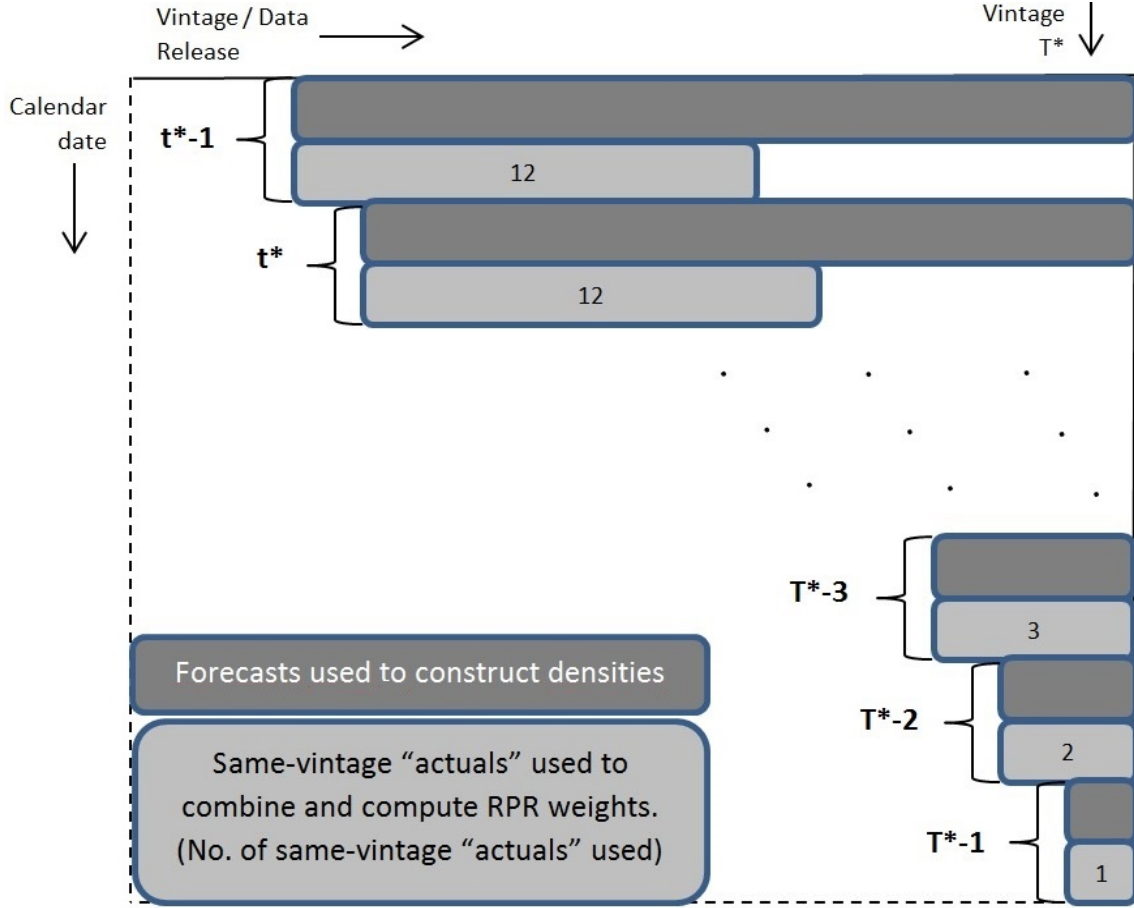
In essence, we redo the ROF analysis in the previous subsection restricted to the real-time data set available at T^* and to using at most 12 same-vintage 'actuals'. This implies that as time progresses we obtain a (possibly different) set of RPR weights, i.e., for each T^* we get a graph in the same fashion as the one in Figure 5. Suppose a researcher having obtained a new data release at moment T^* , which provides her or him with a first or revised estimate of $y_{\lceil T^* - 1 \rceil}$ (depending on whether T^* is the first month of quarter $\lceil T^* \rceil$ or not), aims at using the newly published data to now- or forecast the next 4 quarters of y , i.e., $\hat{y}_{\lceil T^* \rceil}^{T^*}$ (nowcast), $\hat{y}_{\lceil T^* + 1 \rceil}^{T^*}$, $\hat{y}_{\lceil T^* + 2 \rceil}^{T^*}$ and $\hat{y}_{\lceil T^* + 3 \rceil}^{T^*}$. The question arises how she or he can make use of the corresponding RPR weights in order to incorporate revision process information (up to month T^*) into these now- and forecasts.

We propose the following: As before, we estimate our models (1a)-(4)¹³ on the period from $t = 2$ to $\lceil T^* - 1 \rceil$. Note that we only employ vintage $v = T^*$ though. Subsequently, for each model we compute 4 one-step-ahead out-of-sample dynamic forecasts to obtain $\hat{y}_{\lceil T^* + h \rceil}^{T^*}$, $h = 0, 1, 2, 3$.¹⁴ For each h , these model-specific forecasts are then combined using the RPR weights corresponding to $\lceil T^* - 1 \rceil$. Indeed, as stated at the end of Section 3.2, the RPR weights summarize the extent to which our various models were able to resemble the revision process up until the date under consideration. Furthermore, as became evident from Figure 5, the RPR weights stabilize along time, which, together with the previous argument, makes it reasonable to use the latest-available RPR weights to combine those forecasts.

¹³Remember that models (1a)-(3b) are nested in equation (7) with $\alpha(L) = 1 - \alpha L$, $K_l = 1$ and two regressors (see Appendix A).

¹⁴Whenever necessary, we use a simple AR(1), respectively AR(m_j), model to forecast future (w.r.t. T^*) values of X . To be more precise, for Point-in-Time or Average sampled regressors an AR(1) is employed, i.e., $\hat{x}_t = \hat{\mu}^{(AV/PIT)} + \hat{\rho}^{(AV/PIT)} x_{t-1}$, whereas for MIDAS-restricted regressors an AR(m_j) is used, i.e., $\hat{x}_{t-k/m_j} = \hat{\mu}^{(MIDAS)} + \sum_{i=1}^{m_j} \hat{\rho}_i^{(MIDAS)} x_{t-(k+i)/m_j}$ for $k = 0, \dots, m_j - 1$. The parameters $\hat{\rho}^{(AV/PIT)}$ and $\hat{\rho}^{(MIDAS)}$ are computed on the estimation period $t = 2, \dots, \lceil T^* - 1 \rceil$ as well.

Figure 6: The Approach for Real-Time Studies illustrated



Note: The densities are evaluated using scores that are based on maximally the first 12 same-vintage 'actuals' (depicted as light grey bars). Hence, the evaluation and combination of model-specific densities is based on their forecasting performance with respect to the first 12 releases of the variable of interest, i.e., when data is revised a lot. For the rest, see Figure 4.

As an illustration, let us consider a researcher conducting the aforementioned real-time analysis from July 1987 ($T^* = V_F = 107 - 2/3$) until December 2009 ($T^* = 200$). In July 1987 she or he computes i) $\widehat{GNP}_{107}^{107-\frac{2}{3}}$, the first estimate of which becomes available 3 months later, ii) $\widehat{GNP}_{108}^{107-\frac{2}{3}}$, getting assigned a first estimate 6 months later, iii) $\widehat{GNP}_{109}^{107-\frac{2}{3}}$, in this sense a 9-month-ahead forecast and iv) $\widehat{GNP}_{110}^{107-\frac{2}{3}}$, a 12-month-ahead forecast. In August 1987, $\widehat{GNP}_{107}^{107-\frac{1}{3}}$, $\widehat{GNP}_{108}^{107-\frac{1}{3}}$, $\widehat{GNP}_{109}^{107-\frac{1}{3}}$ and $\widehat{GNP}_{110}^{107-\frac{1}{3}}$ constitute 2-, 5-, 8- and 11-month-ahead forecasts, respectively, whereas the corresponding now- and forecasts made in September 1987 are 1-, 4-, 7- and 10-month-ahead forecasts, respectively. As the example period consists

of 23.5 years, we obtain a 1-, ..., 12-month-ahead forecast for each of the 94 quarters under consideration. By comparing the now-/forecasts to their corresponding first releases, i.e., to $GNP_{[T^*+h]}^{[T^*+h]+\frac{1}{3}}$ for $h = 0, 1, 2, 3$, we can compute 12 RMSEs for the 1-, ..., 12-month-ahead forecasts. This allows us to investigate in what sense prediction accuracy of the first release depends on T^* and h .¹⁵ The corresponding 12 RMSEs are given in Table 4.

Table 4: RMSEs for 1- up to 12-month-ahead Forecasts

...months-ahead	RMSE
12-...	0.00578
11-...	0.00577
10-...	0.00576
9-...	0.00575
8-...	0.00574
7-...	0.00569
6-...	0.00568
5-...	0.00570
4-...	0.00545
3-...	0.00477
2-...	0.00457
1-...	0.00441

Note: The figures represent the RMSEs for the period July 1987 to December 2009 corresponding to forecasts of U.S. GNP growth made in real-time 1- up to 12-months ahead with respect to the first release of the respective figure. See the previous paragraph for details.

It becomes apparent that, with only one exception (from 6- to 5-months-ahead), the RMSEs are decreasing with the 'forecast horizon'. Note that the largest difference can be observed from 4- to 3-months-ahead, i.e., when turning from fore- to nowcasting. This implies the obvious conclusion that the closer a researcher is to the moment of the first release, the better her or his forecast. However, inspecting the forecast sensitivity to a particular vintage within a quarter, i.e., comparing 12-, 11- and 10-month-ahead forecasts, 9-, 8- and 7-month-ahead forecasts, and so on, reveals an interesting observation: Except for the 10-month-ahead case, forecasts made in the last month of a quarter, i.e., 1-, 4- and 7-months-ahead, are statistically significantly more accurate than those made in the first or second month of a quarter.¹⁶ This implies that in January 2010, for example, we should wait for the data release of March to become available in order to compute forecasts of GNP growth in 2010Q1-2010Q3.¹⁷

¹⁵In order to compute RMSEs for each of the forecasts under consideration, we neglected vintages beyond December 2009. In January 2010, for example, the 12-month-ahead forecast ($\widehat{GNP}_{204}^{201-\frac{2}{3}}$) would need to be compared to the figure published in January 2011 ($v = 204\frac{1}{3}$), a time period that is beyond our sample since the last vintage at our disposal is December 2010 ($V_L = 204$).

¹⁶As computed via respective Diebold and Mariano (1995) tests. Results available upon request.

¹⁷Note that we cannot formally compare, e.g., 3- and 4-month-ahead or 6- and 7-month-ahead forecasts, as

As mentioned before, for each T^* we obtain a (potentially) different set of RPR weights. With $T^* = 106\frac{1}{3}, 106\frac{2}{3}, \dots, 200$ in our example, we obtain 282 (amount of months in 23.5 years) graphs similar to the one in Figure 5. As an illustration of how the RPR weights change throughout the real-time analysis, Figure 7 displays the corresponding graphs made at six distant moments in time.

Note how the set of calendar dates, for which RPR weights are computed, extends as we move forward in time and how the final graph, i.e., the one corresponding to 2009M12, almost coincides with the one in Figure 5.¹⁸ Another interesting observation is that while the PIT model receives a relatively large weight in the top three panels, its cointegration-including counterpart, PIT_COINT, seems to take over in the bottom three panels.

Finally, when computing now- or forecasts in real time as outlined above, we proposed to employ the latest-available RPR weights at each corresponding moment in time. In other words, for the six example vintages in Figure 7 we would have used the weights corresponding to the right end points of each panel. For our real-time sample from 1987M7 ($T^* = V_F = 107 - 2/3$) to 2009M12 ($T^* = 200$) Figure 8 plots the corresponding latest-available RPR weights. It thus shows the evolution of the models' RPR weights that are ultimately used for now- or forecasting in real time.

Here, too, we can see how PIT_COINT replaces PIT as the dominant Point-in-Time sampling model after roughly half of the real-time periods have been considered. Furthermore, it seems that the latest-available RPR weights stabilize after roughly a third of the real-time sample with PIT_COINT and MIDAS claiming the largest weights (approximately 30% each).

5 Conclusion

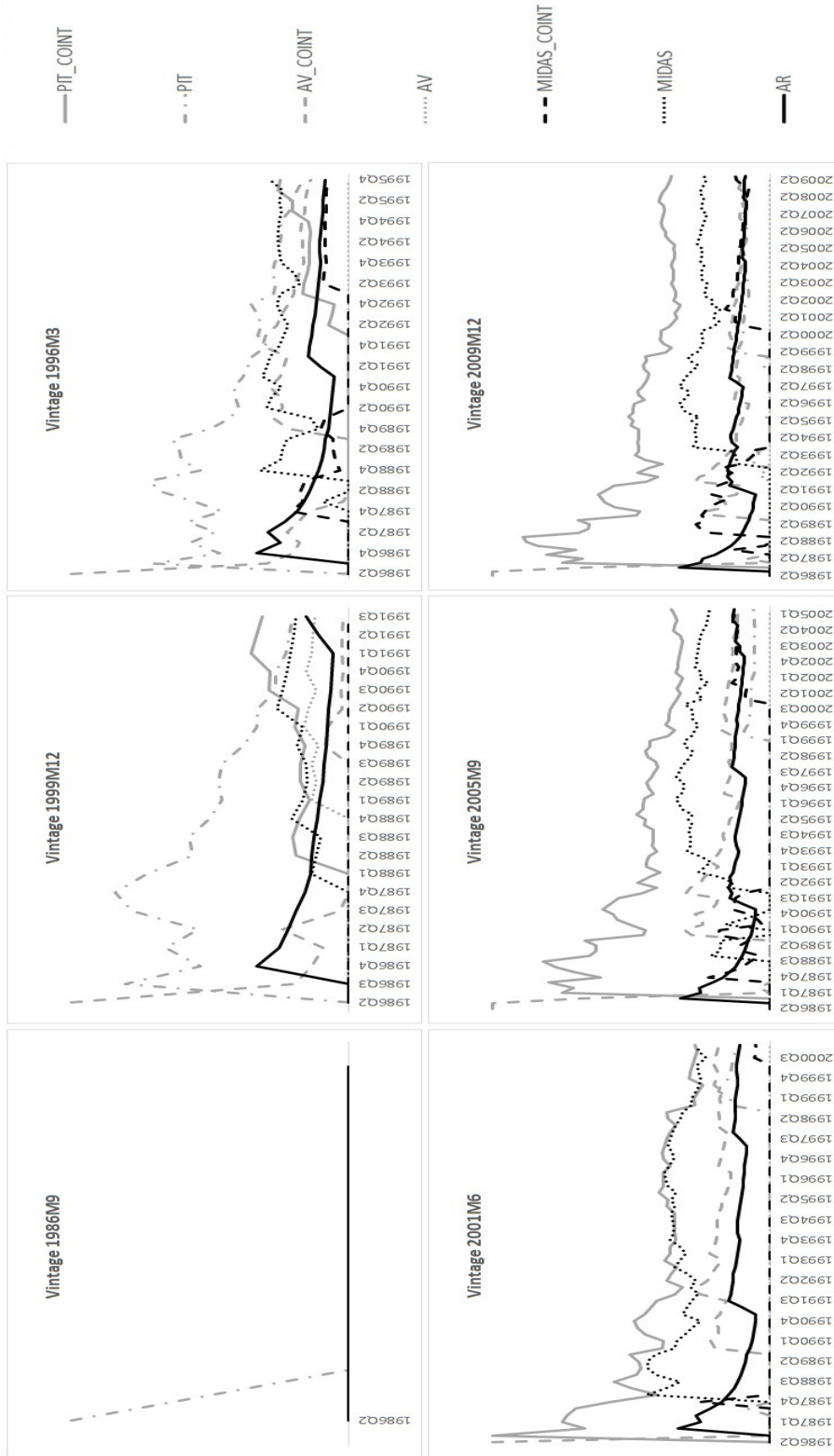
In this paper we combined the issues of working with real-time data sets and dealing with variables that are sampled at mixed frequencies. As such, the repeated observations forecasting approach of Stark and Croushore (2002), was extended to an ADL setting where the regressors are possibly sampled at higher frequencies than the regressand. Furthermore, we discussed how to use our approach in real-time analyses. The ideas were illustrated by means of an empirical analysis involving quarterly U.S. GNP growth.

As far as the extension of the ROF approach is concerned, the distribution of one-step-ahead forecasts, of each model, was non-parametrically approximated by a density function summarizing the extent to which forecasts of a particular calendar date differ across vintages. In other words, we considered and measured uncertainty surrounding the data. Combinations of these model-specific densities were subsequently constructed using what we called revision-process-robust weights, which summarize the extent to which the various models are able to mimic the revision process up until the date under consideration. The RPR weights implied a ranking of the different models under consideration, which was compared to usual RMSE-based

the set of first releases, to which we need to compare the forecasts, changes.

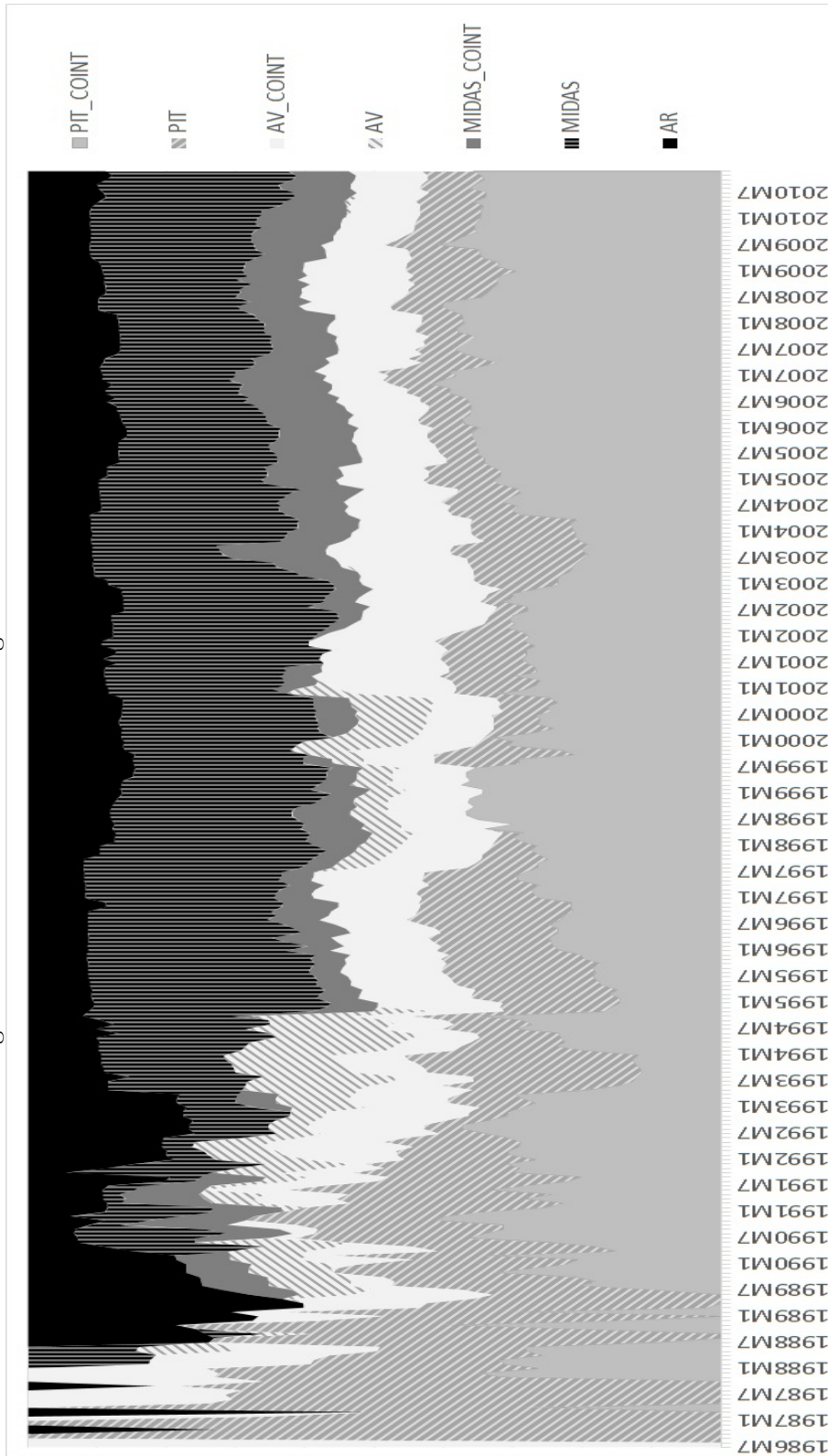
¹⁸Even at vintage V_L they would not coincide, though, since \bar{V} in Section 3.4 differs from \bar{V} in our real-time analysis.

Figure 7: RPR Weights at 6 Vintages in Real Time



Note: The figure displays the RPR weights assigned to the seven models at 6 different moments in time: 1986M9, 1991M12, 1996M3, 2001M6, 2005M9 and 2009M12. Depending on the moment of time at which the graph is produced the calendar dates, to which the RPR apply, differ. For example, in 1991M12 only calendar dates from 1986Q2 to 1991Q3 get weights assigned. Calendar dates are represented on the horizontal axis, whereas the weight itself is plotted on the vertical axis (ranging from 0 to 1).

Figure 8: Latest-available RPR Weights in Real Time



Note: The figure displays the latest-available RPR weights of our seven models for each vintage in our real-time sample period, i.e., for each $T^* \in [107 - 2/3, 200]$. At a given T^* (plotted on the horizontal axis), the vertical length of a shaded area corresponds to the respective model's latest-available RPR weight. Of course, the total vertical length equals 1 for each T^* .

rankings obtained using final-vintage data only. In this way it was investigated whether model rankings are sensitive to data releases.

Empirically, it was found that forecasts of U.S. GNP growth are sensitive to both the model chosen and the data release considered and that adding exogenous variables improves forecast accuracy. More importantly, model 'horse races' were sensitive to the data release employed, i.e., the model ranking based on RMSEs obtained using final-vintage data differed from the ranking implied by the RPR weights. To be more precise, using only latest-available data, some models seemed to possess a better (or worse) forecasting performance than when their forecast accuracy was evaluated also using past data releases.

As for the incorporation of revision process information into real-time studies, we proposed to conduct the ROF analysis from before (with emphasis on the models' ability to forecast early releases) up until the current moment in time. Assuming constancy of the RPR weights in the future, we used the latest-available RPR weights to combine out-of-sample forecasts of U.S. GNP growth. It turned out that RMSEs decrease with the 'forecast horizon' and that forecasts made in the last month of a quarter seem more accurate than the ones based on data from the first or second month. Finally, the latest-available RPR weights stabilized after roughly a third of our real-time sample with PIT_COINT and MIDAS as dominating models.

The preceding analysis could be extended along several lines. Most importantly, the set of regressors and models could be increased. Including variables of different and higher frequencies as well as a larger set of models might lead to new insights in several aspects, notably the combination approach, its associated weight structure and the use of higher-frequency vintages. Independent from the scale of the models and the variables' sampling frequencies, however, our analysis showed that model rankings are potentially very sensitive to the data release chosen. It also illustrates, though, that researchers can guard themselves against such critique by incorporating revision process information into their (real-time) studies. The computation of our proposed RPR weights is, however, dependent on the existence of a real-time data set, emphasizing the importance of making them available to empirical researchers.

References

- Aastveit, K. A., Gerdrup, K. R., Jore, A. S., and Thorsrud, L. A. (2011). Nowcasting gdp in real-time: A density combination approach. Technical report.
- Amisano, G. and Geweke, J. (2013). Prediction using several macroeconomic models. Technical report.
- Amisano, G. and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, 25:177–190.
- Andreou, E., Ghysels, E., and Kourtellis, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31:240–251.
- Clements, M. and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data. *Journal of Business & Economic Statistics*, 26:546–554.
- Clements, M. and Galvão, A. B. (2009). Forecasting u.s. output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics*, 24(7):1187–1206.
- Clements, M. P. and Galvão, A. B. (2011). Forecasting with vector autoregressive models of data vintages: U.s. output growth and inflation. *International Journal of Forecasting*, forthcoming.
- Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of Economic Literature*, 49(1):72–100.
- Croushore, D. and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1):111–130.
- Diebold, F. X. (2012). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests. Working Paper 18391, National Bureau of Economic Research.
- Diebold, F. X. and Lopez, J. A. (1996). *Forecast evaluation and combination*, volume 1 of *Handbook of Statistics*, chapter 14, pages 241–268. Amsterdam: North-Holland.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The midas touch: Mixed data sampling regression models. CIRANO Working Papers 2004s-20, CIRANO.

- Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.
- Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.
- Götz, T. B., Hecq, A., and Urbain, J.-P. (2013). *Testing for common cycles in non-stationary VARs with varied frequency data*, volume 32, pages 361–393. Emerald Group Publishing Limited.
- Götz, T. B., Hecq, A., and Urbain, J.-P. (2014). Forecasting mixed frequency time series with ecm-midas models. *Journal of Forecasting*, 33:198–213.
- Greene, W. H. (2008). *Econometric analysis*. Pearson Prentice Hall, 6th edition.
- Jacobs, J. P. and van Norden, S. (2011). Modeling data revisions: Measurement error and dynamics of 'true' values. *Journal of Econometrics*, 161(2):101–109.
- Jacobs, J. P. A. M., Sarferaz, S., Sturm, J.-E., and van Norden, S. (2013). Modeling Multivariate Data Revisions. CIRANO Working Papers 2013s-44, CIRANO.
- Mankiw, N. and Shapiro, M. (1986). News or noise: An analysis of gnp revisions. *Survey of Current Business*, 66:20–25.
- Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics*, 17(1):129–136.
- Miller, J. I. (2011). Conditionally efficient estimation of long-run relationships using mixed-frequency time series. Working Papers 1103, Department of Economics, University of Missouri.
- Stark, T. and Croushore, D. (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics*, 24(4):507–531.
- Tay, A. and Wallis, K. F. (2000). Density forecasting: A survey. Technical report.
- Timmermann, A. (2006). *Forecast Combinations*, volume 1 of *Handbook of Economic Forecasting*, chapter 4, pages 135–196. Elsevier.
- Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics*, 67(s1):983–994.

A Models

Below we list the models which we use for forecasting. As mentioned in the main body of the paper, the possibility of cointegration between the variables is taken into account by considering the models with and without a cointegrating relationship. Although we aim at forecasting GNP growth rates, it is well possible that *GNP* is cointegrated with one or more of the regressors. As shown in Götz et al. (2014), disregarding (including) a long-run relationship in the presence (absence) of cointegration leads to a considerably worsened forecasting performance. Hence, all models, except (4), i.e., the autoregressive one, are in an error-correction format in which a long-run term (z_{t-1}^v below) is either in- or excluded (labeled by 'a' and 'b', respectively).

- (1) **Restricted short-run (ECM-MIDAS):** 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., 2013). If y_t^v is a quarterly variable and x_{t-i/m_j}^v is daily, we might have to estimate over 50 parameters. In an (ECM-)MIDAS model we hyper-parameterize the polynomial lag structure yielding

$$\alpha(L)\Delta y_t^v = c + \delta z_{t-1}^v + \beta \sum_{i=0}^{K_h} w_{i+1}(\theta) \Delta^{(1/m_j)} x_{t-i/m_j}^v + u_t, \quad (6)$$

where $\alpha(L)$ is a lag polynomial allowing for autoregressive terms, z_{t-1}^v 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 β . Note that $\Delta^{(1/m_j)}$ represents the high-frequency difference operator, i.e., $\Delta^{(1/m_j)} x_{t-i/m_j}^v = x_{t-i/m_j}^v - x_{t-(i+1)/m_j}^v$, whereas Δ represents its low-frequency counterpart, i.e., $\Delta y_t^v = y_t^v - y_{t-1}^v$. The weights are based on an underlying function for which different specifications are proposed in the literature (see Ghysels et al., 2007 for a good overview). We employ the two-dimensional exponential Almon lag polynomial which is shown to be extremely flexible allowing the weight-determination to be completely data-driven.¹⁹

Equation (6) can be extended to include multiple regressors by adding separate weight functions for each one. If lags of the regressors beyond one low-frequency period are excluded, $K_h = m - 1$ is sufficient. If, however, high-frequency observations corresponding to, e.g., $t - 1$ enter the regression, two approaches are possible. First, one may simply set $K_h = 2m - 1$ and estimate one weight function per regressor. Second, separate weight functions for each set of m high-frequency observations (once corresponding to period t and once to period $t - 1$) may be estimated. The latter has the advantage of allowing more flexibility in the weights, but the disadvantage of a larger set of parameters that

¹⁹Another popular, and similarly flexible, choice is the Beta lag polynomial (see Ghysels et al., 2007 for details and more lag polynomial choices).

need to be estimated. Obviously, (6) changes to

$$\alpha(L)\Delta y_t^v = c + \delta z_{t-1}^v + \sum_{k=0}^{K_l} \beta_k \sum_{i=0}^{m_j-1} w_{i+1}^{(k)}(\theta) \Delta^{(1/m_j)} x_{t-k-i/m_j}^v + u_t, \quad (7)$$

where K_l denotes the number of low-frequency lag periods for x and where $w^{(0)}, \dots, w^{(K_l)}$ are the $K_l + 1$ distinct weight functions. In this paper we opt for the second approach, i.e., Equation (7), due its greater flexibility in determining the weights. The two-dimensional Almon lag polynomial then becomes

$$w_{i+1}^{(k)}(\theta) = w_{i+1}(\theta_1^{(k)}, \theta_2^{(k)}) = \frac{\exp(\theta_1^{(k)}(i+1) + \theta_2^{(k)}(i+1)^2)}{\sum_{i=0}^{m_j-1} \exp(\theta_1^{(k)}(i+1) + \theta_2^{(k)}(i+1)^2)}.$$

As far as the long-run term in the ECM-MIDAS model is concerned, Götzt et al. (2014) discuss the choice of which high-frequency observation to include in z_{t-1}^v . Actually, it does not make a significant difference in terms of forecasting performance whether one includes the end-of-period observation or an earlier high-frequency observation as long as the structure of short-run dynamics terms is adapted accordingly (see Götzt et al., 2014). This result should not come as a surprise given that cointegration is a long-run property (Marcellino, 1999). Due to this argument and for its simplicity, we use the end-of-period observations, x_t^v , to enter the disequilibrium error (as also done in Miller, 2011 or Götzt et al., 2014) whenever it is included in the model. For each vintage the cointegrating coefficient will be estimated by OLS implying that we follow the standard two-step Engle-Granger approach (Engle and Granger, 1987).

Note that ECM-MIDAS models including a cointegrating relationship are labeled (1a), whereas the ones without a long-run term are denoted by (1b).

- (2)-(3) **Averaging and Point-in-Time sampling:** Alternative to ECM-MIDAS models one can temporally aggregate the high-frequency series into low-frequency ones as discussed in Marcellino (1999) among others. Averaging, i.e., models (2a)-(2b), or Point-in-Time sampling, i.e., models (3a)-(3b), are the common aggregation schemes here. The former approach takes the average of the high-frequency observations in a certain t -period as the corresponding low-frequency observation, while the latter takes one specific high-frequency observation (often the last one). Both approaches are in fact special cases of the ECM-MIDAS model. Using the notation in (7) we have:

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

$$w_{i+1}^{(k)}(\theta) = w_{i+1} = \frac{1}{m_j} \quad \forall i = 1, \dots, m_j \quad (\text{Average sampling}).$$

- (4) **ARIMA(4, 1, 0):** As a benchmark we consider an *ARIMA*(4, 1, 0) as is done almost throughout the whole investigation in Stark and Croushore (2002).

We estimate models (1a)-(3b) with one low-frequency lag for the regressand and the two regressors, i.e., $\alpha(L) = 1 - \alpha L$ and $K_l = 1$ in Equation (7), which needs to be adjusted for the presence of two regressors with different frequencies, m_1 and m_2 .