



Forecasting with Mixed Frequencies

Michelle T. Armesto, Kristie M. Engemann, and [Michael T. Owyang](#)

A dilemma faced by forecasters is that data are not all sampled at the same frequency. Most macroeconomic data are sampled monthly (e.g., employment) or quarterly (e.g., GDP). Most financial variables (e.g., interest rates and asset prices), on the other hand, are sampled daily or even more frequently. The challenge is how to best use available data. To that end, the authors survey some common methods for dealing with mixed-frequency data. They show that, in some cases, simply averaging the higher-frequency data produces no discernible disadvantage. In other cases, however, explicitly modeling the flow of data (e.g., using mixed data sampling as in Ghysels, Santa-Clara, and Valkanov, 2004) may be more beneficial to the forecaster, especially if the forecaster is interested in constructing intra-period forecasts. (JEL C32)

Federal Reserve Bank of St. Louis *Review*, November/December 2010, 92(6), pp. 521-36.

Forecasting macroeconomic variables is an important task for central banks, financial firms, and any other entity whose outcome depends on business cycle conditions. Unfortunately, many important macroeconomic indicators are not sampled at the same frequency. For example, gross domestic product (GDP) data are sampled quarterly, employment and inflation data are sampled monthly, and most interest rate data are sampled daily. Forecasting models, however, generally require data to be of the same frequency. This presents a small, yet manageable, problem for the econometrician, for which several solutions are available.

In this article, we examine a few common solutions to the mixed-frequency problem. In most cases, forecasters time-aggregate higher-frequency data to match the sampling rate of lower-frequency data. For example, a forecaster may time-aggregate three monthly samples of

employment data into a single observation for each quarterly sample of GDP data. One way to do this is to take a simple average of the three monthly samples. The higher-frequency data would then be entered into the regression as a simple average along with the lower-frequency data.

Such simple averaging is the most common method of time-aggregating higher-frequency variables; however, in principle, one could use any (normalized) weighting function. For example, each intra-quarter observation could be assigned a different coefficient (henceforth, step weighting). While this may be tractable when mixing quarterly and monthly observations, other sampling frequencies may be problematic. With parsimony in mind, Ghysels, Santa-Clara, and Valkanov (2004) propose a general framework called mixed data sampling (MIDAS) to use when a forecaster wants to estimate a small number of hyperparameters relative to the sampling rate of the higher-frequency vari-

Michelle T. Armesto and Kristie M. Engemann are senior research associates and Michael T. Owyang is a research officer at the Federal Reserve Bank of St. Louis. The authors benefited from conversations with Eric Ghysels and Robert Rasche. The authors thank Arthur Sinko for sharing code.

© 2010, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

able (in particular, daily versus monthly sampling). Ghysels, Santa-Clara, and Valkanov (2004) employ (exogenously chosen) distributed lag polynomials as weighting functions.

MIDAS models have been used to forecast quarterly series using monthly or weekly data. For example, Clements and Galvão (2008) introduced a common factor to the MIDAS model with an autoregressive (AR) component. They found that their model provided better forecasts at short horizons—especially within-quarter horizons—than a benchmark AR or an AR distributed-lag model. Kuzin, Marcellino, and Schumacher (2009) used monthly series to forecast euro-area quarterly GDP. They compared the performance of the AR-MIDAS model of Clements and Galvão (2008) to a vector autoregression (VAR) and found that the AR-MIDAS model performed better near one-quarter horizons, while the VAR model performed better near three-quarter horizons. Galvão (2007) included a MIDAS framework in a smooth-transition autoregression to allow for changes in a higher-frequency variable's forecasting ability. Her model improved forecasts of quarterly GDP when using weekly short-term interest rate and stock returns data along with term spread data, sometimes up to horizons of two or three years.

Other studies have used daily or intra-daily data to forecast quarterly data. Tay (2006) used daily stock returns in three models to forecast quarterly GDP growth: an AR model and a MIDAS model, which both included higher-frequency data, and a benchmark model. He found that for the early 2000s, his MIDAS model outperformed his benchmark model by 20 to 30 percent, while his AR model using stock returns over a specified period performed even better. Ghysels, Santa-Clara, and Valkanov (2006) used daily and intra-daily stock returns in a MIDAS model to predict future stock-return volatility. Compared with their benchmark model, using high-frequency returns (especially the sum of 5-minute absolute returns data) improved the forecasts by up to 30 percent for horizons of up to four weeks. Ghysels and Wright (2009) included changes in daily interest rates in a MIDAS model to predict upcoming quarterly releases from the Survey of Professional Forecasters. Andreou, Ghysels, and Kourtellos

(2010a) found that incorporating daily factors (obtained from using financial data in a dynamic factor model) improved the forecasting ability of their MIDAS model for some horizons.

There are other methods for mixing frequencies. Although we will not examine them all in detail, we note a few. For example, one could treat the lower-frequency series as though the data existed but were missing, that is, conduct the forecasting regression at the higher frequency and use forecasted observations of the lower-frequency variable for dates with no actual observation. The obvious question, though, is how to construct the missing data. Fernández (1981), for example, suggests interpolation. Recently, Eraker et al. (2008) used similar methods in a Bayesian framework. One could also employ the Kalman filter to construct the missing data (e.g., Fulton, Bitmead, and Williamson, 2001) or construct factors (e.g., Giannone, Reichlin, and Small, 2008; Aruoba, Diebold, and Scotti, 2009; and Camacho and Perez-Quiros, 2010). Bai, Ghysels, and Wright (2010) discuss the link between the MIDAS regressions covered in this article and the Kalman filter.

In this article, we use simple time averaging, a step-weighting function, and exponential Almon polynomial MIDAS to forecast the following variables using the noted data as the predictor: (i) quarterly GDP growth, using its own lags and monthly employment growth rates, (ii) monthly inflation, using its own lags and daily interest rates, (iii) monthly industrial production (IP) growth, using its own lags and daily interest rates, and (iv) monthly employment growth, using its own lags and daily interest rates. These cases demonstrate how the three methods differ when the difference between the higher and lower sampling frequencies is increased. We then test these forecasts out-of-sample to provide a rough assessment of the performance of each method.

The balance of the paper is constructed as follows: The next section describes the data and forecasting environment and introduces the notation. The following section describes the three forecasting methods used. The subsequent two sections present results from the forecasting experiments: The first compares the performance of the three methods using end-of-period data;

the second evaluates the performance of MIDAS using intra-period data. The final section offers some concluding remarks.

THE FORECASTING ENVIRONMENT AND NOTATION

Before proceeding, a number of essential elements must be addressed: notation, forecast evaluation techniques, and the data.

The problem of mixed sampling frequencies is exemplified in Figures 1 and 2. Figure 1 shows quarterly GDP and monthly employment growth rates for the period 1980 to mid-2009: As is typical, the monthly employment observations fluctuate between the quarterly GDP observations. Figure 2 shows how daily federal funds rate observations similarly fluctuate between monthly consumer price index (CPI) inflation observations.

When comparing across modeling environments, it is important to use common notation. In the econometric procedures that follow, our objective is to forecast a lower-frequency variable, Y , sampled at periods denoted by time index t . Past realizations of the lower-frequency variable are denoted by the lag operator, L . For example, if Y_t is the monthly inflation rate, then the inflation rate one month prior would be the first lag of Y_t , $LY_t = Y_{t-1}$, two months prior would be $L^2Y_t = Y_{t-2}$, and so on.

In addition to lags of Y , we are interested in the information content of a higher-frequency variable, X , sampled m times between samples of Y (e.g., between $t-1$ and t).¹ L_{HF} denotes the lag operator for the higher-frequency variable. If X_t is the daily federal funds rate, then $L_{HF}X_t$ denotes the one-day-before- t realization of the federal funds rate (i.e., the last day of the previous month). If X_t is monthly employment growth used to forecast quarterly GDP, then $L_{HF}X_t$ denotes the employment growth rate of the last month of the previous quarter.

Figure 3 depicts the forecast timeline, which for simplicity shows one-period-ahead forecasts.

Generalization to longer horizons should be obvious. Assume that at time t we are interested in forecasting Y_{t+1} , the circled observation on the timeline. Standard forecasting experiments would use data available through time t ; this is depicted in the gray section of the timeline. We perform such end-of-period forecasting experiments using each of the three methods noted above. The blue section of the timeline depicts information that becomes available during the $t+1$ period (i.e., leads); this information may be relevant for forecasting Y_{t+1} . Using the MIDAS method, we perform intra-period forecasting experiments using both the data specified in the gray section and that in the blue section. For the end-of-period and intra-period forecasting experiments, we provide results from a rolling-window scheme (i.e., the in-sample estimation period is a fixed number of periods, T) and results from a recursive scheme (i.e., the in-sample estimation uses all available data up to period t). We compute root mean-squared errors (RMSEs) to compare across forecasts.

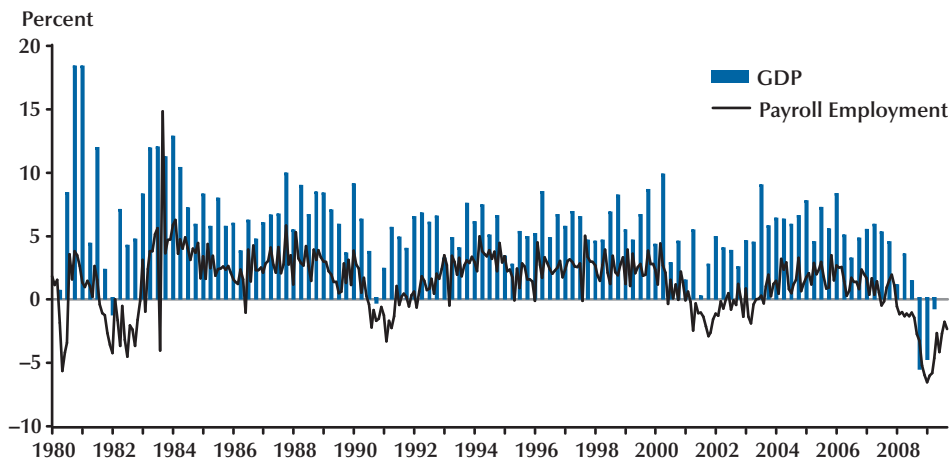
Data

To compare the three methods for time-aggregating higher-frequency data, we use two different sets of sampling rates (i.e., monthly data to forecast quarterly data and daily data to forecast monthly data). In all cases, we use lags of the forecasted variable in addition to a single predictor. Generalization to multiple predictors is straightforward. First, we compare forecasts of the quarterly GDP growth rate using the monthly payroll employment growth rate as the predictor. Then, we compare forecasts of the monthly CPI inflation growth rate, monthly IP growth rate, and monthly employment growth rate using daily predictors: the daily effective federal funds rate for CPI inflation and the daily term spread (which is the difference between the 10-year Treasury note yield and the 3-month Treasury bill yield) for IP and employment. Each vintage of real-time data is seasonally adjusted and thus uses a different set of seasonal factors.

In most forecasting experiments, one wishes to use data available at the time the forecaster would have constructed the forecasts, thus, prior

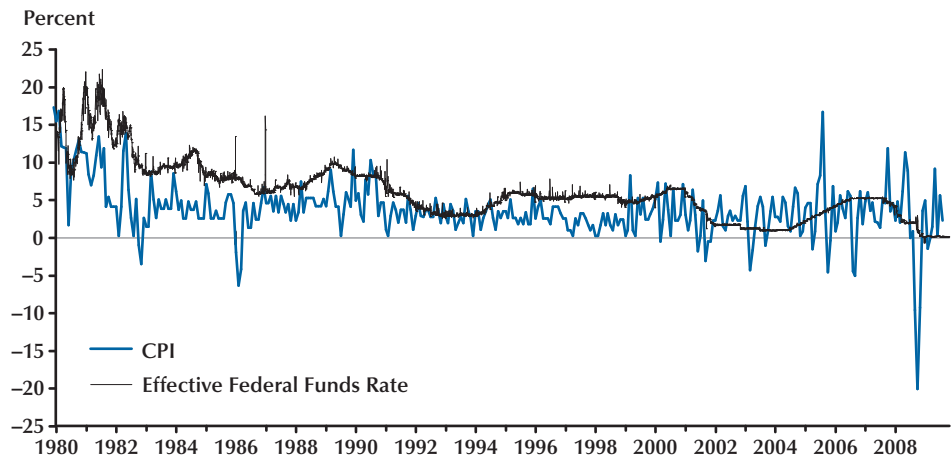
¹ The extension to include other exogenous variables sampled at the same frequency as Y is obvious and will not be explicitly explored here.

Figure 1
Quarterly GDP and Monthly Employment Growth Rates

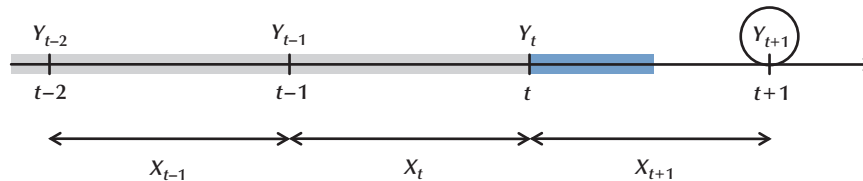


NOTE: Annualized log growth rates.

Figure 2
Monthly CPI Growth Rate and Daily Federal Funds Rate



NOTE: CPI is the annualized log growth rate.

Figure 3**Forecast Timeline**

NOTE: The timeline shows information available for forecasting Y_{t+1} .

to potential revisions (see, for example, Croushore, 2005 and 2006; and Croushore and Stark, 2001). We, therefore, use real-time data (detailed below) for the GDP growth rate, employment growth rate, CPI inflation rate, and IP growth rate. Interest rates are not revised, thus their initial data releases are used. We assume that the goal of the forecaster is to predict the true values of the variables. We also assume that the most recent data vintages of the variables are the true value and use these vintages to compute the forecast errors.²

For our forecasting experiments, the data we use are log growth rates of the seasonally adjusted annual rate of nominal GDP from the Bureau of Economic Analysis, seasonally adjusted nonfarm payroll employment and seasonally adjusted CPI from the Bureau of Labor Statistics, and seasonally adjusted IP from the Board of Governors of the Federal Reserve System. We also use interest rate data from the Board of Governors. The real-time data vintages (from Archival Federal Reserve Economic Data) used are December 1990 through September 2009. For the monthly forecasts using daily data, our initial in-sample estimations run from July 1975 to November 1990 and the out-of-sample forecasts begin in December 1990. For the GDP forecasts using employment data, our in-sample estimations run from the third quarter of 1975 to the fourth quarter of 1990 and the out-of-sample forecasts begin in the first quarter of 1991.

² This choice is made with some caveats. For example, IP is reindexed three times during the full sample period. This might introduce mild distortions in the RMSEs for the 1992, 1997, and 2002 periods.

TIME AGGREGATION

This section describes the three methods we use to time-aggregate higher-frequency data for use in forecasting lower-frequency variables.

Time Averaging

One solution to the problem of mixed sampling frequencies is to convert higher-frequency data to match the sampling rate of the lower-frequency data. The simplest method is to compute the simple average of the observations of X that occur between samples of the lower-frequency variable:

$$\bar{X}_t = \frac{1}{m} \sum_{k=1}^m L_{HF}^k X_t.$$

With the two variables Y_t and \bar{X}_t in the same time domain, our regression approach is simply

$$(1) \quad Y_t = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \sum_{j=1}^n \gamma_j L^j \bar{X}_t + \varepsilon_t,$$

where the γ_j s are the slope coefficient on the time-averaged X s. Notice that the third term in equation (1) employs the higher-frequency lag operator, indicating that we are using, for example, the prior j th month's average of daily X_t s.

Step Weighting

The previous method assumes the slope coefficients on each of the individual observations of X are equal. Alternatively, one could assume

that each of the slope coefficients for each k sampling of X are unique. This model, including one lag of the predictor X ($n = 1$), is

$$(2) \quad Y_t = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \sum_{k=1}^m \gamma_k L_{HF}^k X_t + \varepsilon_t,$$

where $\{\gamma_k\}_{k=1}^m$ represents a set of slope coefficients for all k .

This representation has two potential difficulties. First, the pattern of slope coefficients (and, thus, the weights on each of the lagged intermediate samplings) is unconstrained. One might have the prior belief, for example, that more weight should be given to the samples of X that are more contemporaneous to the observed Y .³ Second, as the sampling rate, m , increases, equation (2) leads to parameter proliferation. For example, for data sampled at a monthly frequency for use in a quarterly model with one lag, $m = 3$, and the number of parameters is manageable. In contrast, for data sampled at a daily frequency for use in a monthly model with one lag, assuming $m = 20$, for instance, means that 20 different slope coefficients must be estimated.

Thus, once the model is extended to multiple lags, the number of parameters could become quite large. The most general model is

$$(3) \quad Y_t = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \sum_{k=1}^{n \times m} \gamma_k L_{HF}^k X_t + \varepsilon_t,$$

which allows for up to n lower-frequency lags. This means we would have to estimate $n \times m = 4 \times 20 = 80$ parameters in the third term alone to forecast a monthly variable using four monthly lags of daily data. An alternative formulation preserves the within-month effects but allows for different effects across months:

$$(4) \quad Y_t = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \sum_{j=1}^n \Gamma_j L^j \sum_{k=1}^m \gamma_k L_{HF}^k X_t + \varepsilon_t.$$

For the same forecasting problem as above (i.e., using four monthly lags of daily data), equation (4)

would have 24 parameters in the third term. Although these are fewer parameters than needed for equation (3), this number is large compared with that needed for equation (1) (our time-averaging model) and could lead to overfitting. Corsi (2009, p. 181, footnote 11) provides an example of the use of step-weighting functions. Andreou, Ghysels, and Kourtellos (2010b) also discuss potential issues of step weighting (e.g., asymptotic biases and inefficiencies).

MIDAS

The time-averaging model is parsimonious but discards any information about the timing of innovations to higher-frequency data. The step-weighting model preserves the timing information but requires the user to estimate a potentially large number of parameters. To solve the problem of parameter proliferation while preserving some timing information, Ghysels, Santa-Clara, and Valkanov (2004) propose this MIDAS model:

$$(5) \quad Y_t = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \gamma \sum_{k=1}^m \Phi(k; \theta) L_{HF}^k X_t + \varepsilon_t,$$

where the function $\Phi(k; \theta)$ is a polynomial that determines the weights for temporal aggregation.

The weighting function, $\Phi(k; \theta)$, can have any number of functional forms; the desire here is to achieve flexibility while maintaining parsimony. Ghysels, Santa-Clara, and Valkanov (2004, 2005, and 2006) suggest, for example, a beta formulation:

$$(6) \quad \Phi(k; \theta_1, \theta_2) = \frac{f\left(\frac{k}{m}, \theta_1, \theta_2\right)}{\sum_{j=1}^m f\left(\frac{j}{m}, \theta_1, \theta_2\right)}$$

where

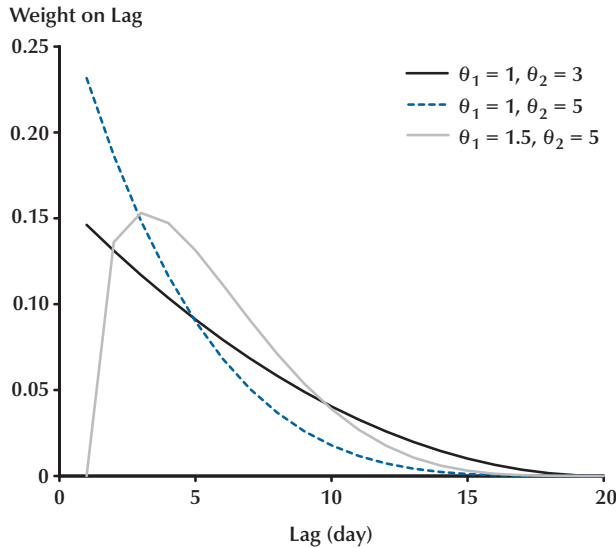
$$f(i, \theta_1, \theta_2) = \frac{i^{\theta_1-1} (1-i)^{\theta_2-1} \Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1) \Gamma(\theta_2)},$$

θ_1 and θ_2 are hyperparameters governing the shape of the weighting function, and

$$\Gamma(\theta_p) = \int_0^1 e^{-i} i^{\theta_p-1} di$$

³ This is similar in flavor to a number of priors used to estimate VARs, including the Minnesota prior and the Sims-Zha (1998) prior.

Figure 4
Beta Polynomial Weighting Function



is the standard gamma function. Figure 4 shows a few parameterizations of the beta polynomial weighting function. In particular, note that various parameterizations can obtain strictly decreasing or humped-shaped weighting functions. In addition, the rate of decay is governed by the parameterization. Simple time averaging is obtained when $\theta_1 = \theta_2 = 1$.

Ghysels, Santa-Clara, and Valkanov (2005) and Ghysels, Sinko, and Valkanov (2007) also suggest an exponential Almon specification:

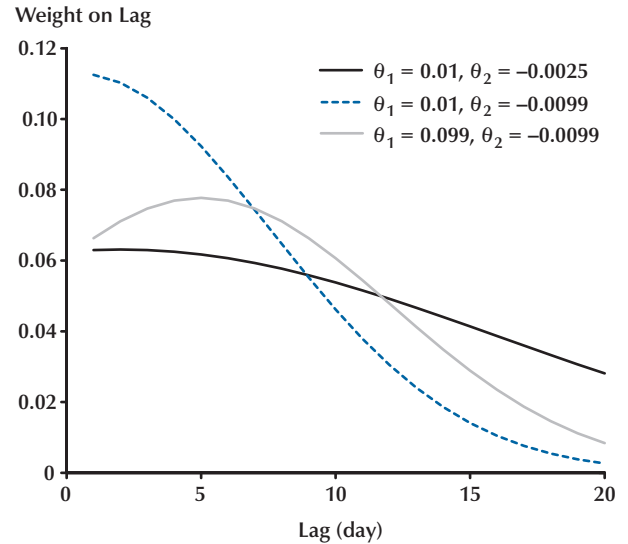
$$\Phi(k; \theta_1, \theta_2) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{j=1}^m \exp(\theta_1 j + \theta_2 j^2)}$$

In this case, simple time averaging is obtained when $\theta_1 = \theta_2 = 0$. Figure 5 shows various parameterizations of the exponential Almon polynomial weighting function.

We can generalize the MIDAS specification to multiple lags of the predictor X :

$$(7) \quad Y_t = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \gamma \sum_{k=1}^M \Phi(k; \boldsymbol{\theta}) L_{HF}^k X_t + \varepsilon_t,$$

Figure 5
Exponential Almon Polynomial Weighting Function



where $M = m \times n$. In this case, lags of the predictor are incorporated by expanding the weighting polynomial. Obviously, this restricts the manner in which the predictor influences the forecast; however, if we believe the influence of the predictor should decay monotonically after a certain period, equation (7) is a useful representation with the additional benefit of being extremely parsimonious.⁴

END-OF-PERIOD FORECASTING RESULTS

The preceding two sections propose a few solutions to the mixed-frequency problem. As with most forecasting problems, the efficacy of the model depends heavily on the nature of the forecasted data and the information available to the forecaster. Even across the models described here for time-aggregating higher-frequency data,

⁴ Although the two weighting schemes can produce many of the same shapes, we choose to present results using the exponential Almon polynomial weighting function.

one might expect each model’s efficacy to depend on, among other things, the difference in the sampling rate and the number of variables included. For example, in models that rely on a large number of explanatory variables, the problem of parameter proliferation might be more evident; thus, the more-parsimonious models, time averaging (equation (1)) and MIDAS (equation (5)), may be favored. When sampling frequencies vary widely, one might expect the step-weighting model (equation (2)) to suffer from parameter proliferation and the time-averaging model (equation (1)) to suffer from a poor approximation to the truth. We test the three models’ ability to forecast four macroeconomic variables using the noted data as the predictor: (i) quarterly GDP growth, using monthly employment growth data, (ii) monthly CPI inflation, using daily federal funds data, (iii) monthly IP growth, using daily term spread data, and (iv) monthly employment growth, using daily interest rate data.

Tables 1 and 2 show the results at various horizons from our end-of-period forecasting experiments using a rolling-window scheme and a recursive scheme, respectively. For these experiments, we use only data up through the end of the previous month (when forecasting IP or employment) or quarter (when forecasting GDP). At the shortest horizon, the MIDAS and time-averaging models are almost identical except for forecasting GDP, for which MIDAS slightly outperforms time averaging. In contrast, MIDAS outperforms step weighting in all cases, even more so when using the term spread as a predictor rather than the federal funds rate. At longer horizons, the models are essentially equivalent. Altering the lag length did not substantially change the results. Obviously, as the lag length increases, the viability of the step-weighting model breaks down. This is especially true for models forecasting monthly variables using daily data.

INTRA-PERIOD FORECASTING RESULTS

One advantage of MIDAS over time averaging and step weighting is that MIDAS can forecast within periods. Suppose we want to forecast CPI

inflation for time t . At the beginning of month t , we have information on CPI inflation for month $t-1$ and daily interest rate data for month $t-1$. With this data, we can construct forecasts for the four noted variables, as we did in the previous section. However, because information that may have predictive power—specifically, month- t daily interest rate data—comes in before the end of month t , we can update our forecast of period- t inflation daily throughout the month. Such an intra-period forecasting experiment would include both the gray and blue sections on the Figure 3 timeline, with the latter depicting the intra-period information.⁵

One way to perform this intra-period forecasting experiment is to simply include the new observations of the higher-frequency data in the regression. For the d th day of the month, the regression is

$$(8) \quad \hat{Y}_{td} = \alpha + \sum_{i=1}^p \beta_i L^i Y_t + \gamma^{LEAD} \sum_{j=m-d+1}^m \Phi(j; \theta^{LEAD}) L_{HF}^j X_{t+1} + \gamma^{LAG} \sum_{k=1}^m \Phi(k; \theta^{LAG}) L_{HF}^k X_t + \varepsilon_t,$$

which has some advantages if we believed there were some within-month calendar effects. Without the third term, equation (8) is identical to the MIDAS regression (7). The third term reflects the effect of the current month’s data.

Equation (8) is actually a set of day-dependent forecasting models—we would have a different regression for each day of the month. Another alternative is to make the following restrictions in equation (8):

$$(9) \quad \begin{aligned} \gamma^{LEAD} &= \gamma^{LAG}, \\ \theta^{LEAD} &= \theta^{LAG}. \end{aligned}$$

This means that forecasts do not differentiate between current- and past-month data—that is, we do not treat the new data as special. There are additional options; however, we do not explore them here.

⁵ The discussion here concerns the forecast of the $h = 0$ horizon of Y . Generalization to different horizons, again, should be obvious.

Table 1
RMSEs from End-of-Period Forecasts, Rolling Window

	Forecasted variable											
	GDP, <i>h</i> quarters ahead			CPI, <i>h</i> months ahead			IP, <i>h</i> months ahead			Employment, <i>h</i> months ahead		
	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5
Forecasting experiment												
1 Quarterly lag of <i>Y</i> ,												
3 Monthly lags of employment												
Model: average	0.5683	0.6307	0.7012									
Model: step weighting	0.5702	0.6457	0.7073									
Model: MIDAS	0.5480	0.6315	0.7042									
1 Monthly lag of <i>Y</i> , 30 daily lags of fed funds rate												
Model: average				0.2537	0.2771	0.2803						
Model: step weighting				0.2623	0.3129	0.2748						
Model: MIDAS				0.2536	0.2769	0.2801						
1 Monthly lag of <i>Y</i> , 22 daily lags of term spread												
Model: average							0.6688	0.6559	0.6589	0.1122	0.1175	0.1353
Model: step weighting							0.7357	0.6664	0.6586	0.1288	0.1247	0.1319
Model: MIDAS							0.6682	0.6590	0.6591	0.1126	0.1177	0.1348

Table 2
RMSEs from End-of-Period Forecasts, Recursive

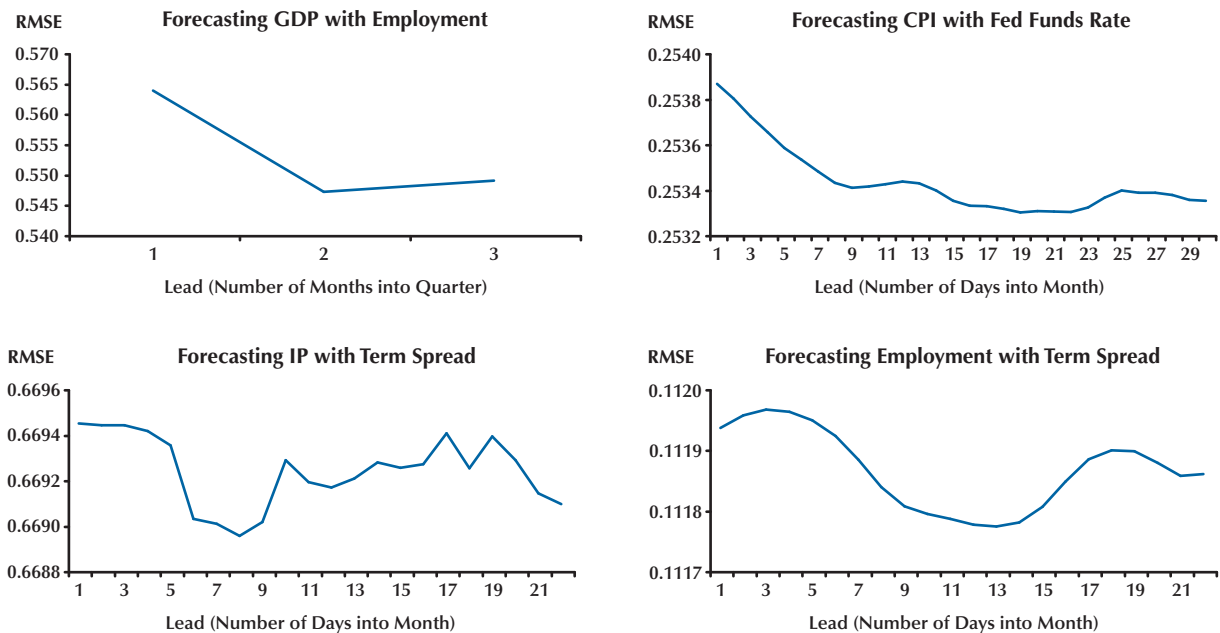
	Forecasted variable															
	GDP, <i>h</i> quarters ahead			CPI, <i>h</i> months ahead			IP, <i>h</i> months ahead			Employment, <i>h</i> months ahead						
	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	
Forecasting experiment																
1 Quarterly lag of <i>Y</i> ,																
3 Monthly lags of employment																
Model: average	0.5972	0.7090	0.7669													
Model: step weighting	0.5857	0.7416	0.7717													
Model: MIDAS	0.5669	0.6903	0.7635													
1 Monthly lag of <i>Y</i> , 30 daily lags of fed funds rate																
Model: average				0.2586	0.2987	0.3067										
Model: step weighting				0.2602	0.3017	0.2991										
Model: MIDAS				0.2577	0.2986	0.3086										
1 Monthly lag of <i>Y</i> , 22 daily lags of term spread																
Model: average				0.6588	0.6576	0.6679	0.6588	0.6576	0.6679	0.1208	0.1240	0.1456	0.6588	0.6576	0.6679	0.1208
Model: step weighting				0.6812	0.6745	0.6873	0.6812	0.6745	0.6873	0.1343	0.1371	0.1453	0.6812	0.6745	0.6873	0.1343
Model: MIDAS				0.6619	0.6588	0.6717	0.6619	0.6588	0.6717	0.1212	0.1232	0.1450	0.6619	0.6588	0.6717	0.1212

Table 3
RMSEs from MIDAS Intra-period Forecasts, Rolling Window

	Forecasted variable														
	GDP, h quarters ahead			CPI, h months ahead			IP, h months ahead			Employment, h months ahead					
	$h = 1$	$h = 2$	$h = 5$	$h = 1$	$h = 2$	$h = 5$	$h = 1$	$h = 2$	$h = 5$	$h = 1$	$h = 2$	$h = 5$	$h = 1$	$h = 2$	$h = 5$
Forecasting experiment															
1 Quarterly lag of Y_t															
3 Monthly lags of employment + d leads															
$d = 1$	0.5637	0.5995	0.6987												
$d = 2$	0.5470	0.5990	0.6917												
$d = 3$	0.5489	0.5585	0.6930												
1 Monthly lag of Y_t															
30 daily lags of fed funds rate + d leads															
$d = 10$				0.2534	0.2757	0.2790									
$d = 20$				0.2533	0.2753	0.2789									
$d = 30$				0.2533	0.2757	0.2786									
1 Monthly lag of Y_t , 22 daily lags of term spread + d leads															
$d = 5$							0.6693	0.6560	0.6625	0.1119	0.1177	0.1371			
$d = 15$							0.6693	0.6571	0.6627	0.1118	0.1178	0.1376			
$d = 22$							0.6691	0.6575	0.6627	0.1119	0.1179	0.1378			

Table 4
RMSEs from MIDAS Intra-period Forecasts, Recursive

	Forecasted variable														
	GDP, h quarters ahead			CPI, h months ahead			IP, h months ahead			Employment, h months ahead					
	$h = 1$	$h = 2$	$h = 5$	$h = 1$	$h = 2$	$h = 5$	$h = 1$	$h = 2$	$h = 5$	$h = 1$	$h = 2$	$h = 5$	$h = 1$	$h = 2$	$h = 5$
Forecasting experiment															
1 Quarterly lag of Y_t															
3 Monthly lags of employment + d leads															
$d = 1$	0.5918	0.6380	0.7681												
$d = 2$	0.5245	0.6375	0.7648												
$d = 3$	0.4931	0.5969	0.7628												
1 Monthly lag of Y_t															
30 daily lags of fed funds rate + d leads															
$d = 10$				0.2564	0.2939	0.3055									
$d = 20$				0.2560	0.2928	0.3057									
$d = 30$				0.2561	0.2930	0.3058									
1 Monthly lag of Y_t , 22 daily lags of term spread + d leads															
$d = 5$				0.6569	0.6576	0.6755	0.6569	0.6576	0.6755	0.6569	0.6576	0.6755	0.6569	0.6576	0.6755
$d = 15$				0.6565	0.6575	0.6750	0.6565	0.6575	0.6750	0.6565	0.6575	0.6750	0.6565	0.6575	0.6750
$d = 22$				0.6564	0.6575	0.6758	0.6564	0.6575	0.6758	0.6564	0.6575	0.6758	0.6564	0.6575	0.6758

Figure 6**RMSEs for Leads (One-Period-Ahead Forecasts, Rolling Window)**

Tables 3 and 4 show the results from our intra-period forecasting experiments (with restriction (9) imposed) using a rolling-window scheme and a recursive scheme, respectively. Compared with the results from Tables 1 and 2, those from Tables 3 and 4 show the benefit of including additional intra-period information. Roughly two-thirds of the RMSEs in Tables 3 and 4 are lower than those of the MIDAS forecasts in Tables 1 and 2. In Tables 3 and 4, for forecasting CPI inflation with the federal funds rate, the RMSE declines over at least the first two-thirds of the month. A similar result holds for forecasting GDP growth with intra-quarter employment growth. The gain is less apparent for forecasting real variables (IP and employment)—as opposed to nominal variables—using the daily term spread; here it appears the RMSEs rise when we take into account the term spread late in the month. This may be because information for real variables reacts more slowly to changes in monetary policy.

Figures 6 and 7 further demonstrate these results by showing the change in the RMSEs for

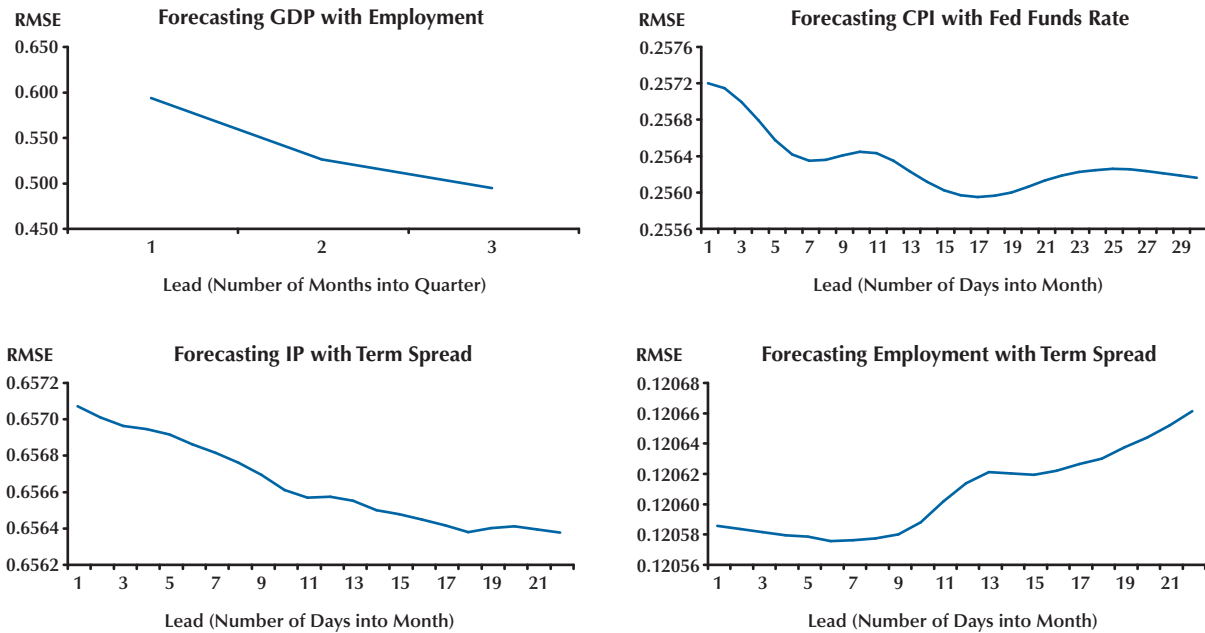
forecasts of the four variables over a quarter or month. The x-axis shows the number of months into the quarter or days into the month that the forecast is computed. In Figure 6, the RMSE for the CPI declines almost monotonically over the current month. This is not true, however, for IP and employment, where the RMSEs decline for early leads but then increase later in the month. For forecasts of one-quarter-ahead GDP, the RMSE across different leads initially declines but then increases slightly. In contrast, Figure 7 shows that the RMSEs for both IP and GDP decline over the given period, whereas the RMSE for employment increases almost monotonically over the month.

SUMMARY AND CONCLUSIONS

Forecasting is important for making policy and financial decisions. In some cases, however, forecasters are often confronted with the problem of mixing data frequencies. Macroeconomic data typically are sampled monthly (e.g., employment,

Figure 7

RMSEs for Leads (One-Period-Ahead Forecasts, Recursive)



CPI, IP) or quarterly (e.g., GDP); financial data (e.g., interest rates, asset prices) typically are sampled daily (or even more frequently). In this paper, we demonstrated some common approaches to the mixed-frequency problem. In particular, we demonstrated time aggregation, which transforms by means of summation—either weighted or unweighted—higher-frequency data to allow standard regression techniques to be used.

For a simple set of experiments, we found that the performances of different time-aggregation approaches vary—that is, there does not appear to be a golden rule. There may be trade-offs between parsimony (i.e., avoiding overfitting) and flexibility at different horizons and for different sets of data. We refer the reader to the extant literature for a more detailed discussion.⁶

⁶ Much of the code used here is available in the Matlab Toolbox for MIDAS regressions. The toolbox is described in Sinko, Sockin, and Ghysels (2010), and the code is available at www.unc.edu/~eghysels/Software_datasets.html.

REFERENCES

- Andreou, Elena; Ghysels, Eric and Kourtellos, Andros. "Should Macroeconomic Forecasters Use Daily Financial Data and How?" Working paper, February 2010a; [http://w4.stern.nyu.edu/emplibrary/Macro MIDAS Feb 22 2010 CEPR.pdf](http://w4.stern.nyu.edu/emplibrary/Macro%20MIDAS%20Feb%2022%202010%20CEPR.pdf).
- Andreou, Elena; Ghysels, Eric and Kourtellos, Andros. "Regression Models with Mixed Sampling Frequencies." *Journal of Econometrics*, October 2010b, 158(2), pp. 246-61.
- Aruoba, S. Borağan; Diebold, Francis X. and Scotti, Chiara. "Real-Time Measurement of Business Conditions." *Journal of Business and Economic Statistics*, October 2009, 27(4), pp. 417-27.
- Bai, Jennie; Ghysels, Eric and Wright, Jonathan H. "State Space Models and MIDAS Regressions." Working paper, January 2010; [www.unc.edu/~eghysels/papers/BGW Kalman vs MIDAS EG Jan 02.pdf](http://www.unc.edu/~eghysels/papers/BGW_Kalman_vs_MIDAS_EG_Jan_02.pdf).
- Camacho, Maximo and Perez-Quiros, Gabriel. "Introducing the Euro-Sting: Short-Term Indicator of Euro Area Growth." *Journal of Applied Econometrics*, June/July 2010, 25(4), pp. 663-94.
- Clements, Michael P. and Galvão, Ana Beatriz. "Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States." *Journal of Business and Economic Statistics*, October 2008, 26(4), pp. 546-54.
- Corsi, Fulvio. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics*, Spring 2009, 7(2), pp. 174-96.
- Croushore, Dean. "Do Consumer-Confidence Indexes Help Forecast Consumer Spending in Real Time?" *North American Journal of Economics and Finance*, December 2005, 16(3), pp. 435-50.
- Croushore, Dean. "Forecasting with Real-Time Macroeconomic Data," in Graham Elliott, Clive W.J. Granger, and Allan Timmermann, eds., *Handbook of Economic Forecasting*. Volume 1. Amsterdam: North-Holland, 2006, Chap. 17.
- Croushore, Dean and Stark, Tom. "A Real-Time Data Set for Macroeconomists." *Journal of Econometrics*, November 2001, 105(1), pp. 111-30.
- Eraker, Bjørn; Chiu, Ching Wai (Jeremy); Foerster, Andrew; Kim, Tae Bong and Seoane, Hernan. "Bayesian Mixed Frequency VAR's." Working paper, September 2008; www.marginalq.com/eraker/BMF.pdf.
- Fernández, Roque B. "A Methodological Note on the Estimation of Time Series." *Review of Economics and Statistics*, August 1981, 63(3), pp. 471-76.
- Fulton, Jennifer A.; Bitmead, Robert R. and Williamson, Robert C. "Smoothing Approaches to Reconstruction of Missing Data in Array Processing," in *Defence Applications of Signal Processing: Proceedings of the US/Australia Joint Workshop on Defence Applications of Signal Processing*. New York: Elsevier, 2001.
- Galvão, Ana Beatriz. "Changes in Predictive Ability with Mixed Frequency Data." Working Paper No. 595, Queen Mary, University of London, Department of Economics and Finance, May 2007; www.econ.qmul.ac.uk/papers/doc/wp595.pdf.
- Ghysels, Eric; Santa-Clara, Pedro and Valkanov, Rossen. "The MIDAS Touch: Mixed Data Sampling Regression Models." Working paper, June 2004; <http://docentes.fe.unl.pt/~psc/MIDAS.pdf>.
- Ghysels, Eric; Santa-Clara, Pedro and Valkanov, Rossen. "There Is a Risk-Return Trade-off After All." *Journal of Financial Economics*, June 2005, 76(3), pp. 509-48.
- Ghysels, Eric; Santa-Clara, Pedro and Valkanov, Rossen. "Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies." *Journal of Econometrics*, March/April 2006, 131(1/2), pp. 59-95.
- Ghysels, Eric; Sinko, Arthur and Valkanov, Rossen. "MIDAS Regressions: Further Results and New Directions." *Econometric Reviews*, 2007, 26(1), pp. 53-90.

Armesto, Engemann, Owyang

- Ghysels, Eric and Wright, Jonathan H. "Forecasting Professional Forecasters." *Journal of Business and Economic Statistics*, October 2009, 27(4), pp. 504-16.
- Giannone, Domenico; Reichlin, Lucrezia and Small, David. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics*, May 2008, 55(4), pp. 665-76.
- Kuzin, Vladimir; Marcellino, Massimiliano and Schumacher, Christian. "MIDAS versus Mixed-Frequency VAR: Nowcasting GDP in the Euro Area." Discussion Paper No. 07/2009, Deutsche Bundesbank, 2009; www.econstor.eu/bitstream/10419/27661/1/200907dkp.pdf.
- Sims, Christopher A. and Zha, Tao. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review*, November 1998, 39(4), pp. 949-68.
- Sinko, Arthur; Sockin, Michael and Ghysels, Eric. "Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regression Models." Unpublished manuscript, July 2010; www.unc.edu/~eghysels/papers/MIDAS_Usersguide_Version1.pdf.
- Tay, Anthony S. "Mixing Frequencies: Stock Returns as a Predictor of Real Output Growth." Working Paper No. 34-2006, Singapore Management University, Economics and Statistics Working Paper Series, December 2006; https://mercury.smu.edu.sg/rsrchpubupload/7469/Tay_2006.pdf.