LUIS M. GÓMEZ-ZAMUDIO
Bank of Mexico

RAÚL IBARRA
Bank of Mexico

# Are Daily Financial Data Useful for Forecasting GDP? Evidence from Mexico

**ABSTRACT** This article evaluates the use of financial data sampled at high frequencies to improve short-term forecasts of quarterly GDP for Mexico. The model uses both quarterly and daily sampling frequencies while remaining parsimonious. In particular, the mixed data sampling (MIDAS) regression model is employed to deal with the multi-frequency problem. To preserve parsimony, factor analysis and forecast combination techniques are used to summarize the information contained in a data set containing 392 daily financial series. Our findings suggest that the MIDAS model incorporating daily financial data leads to improvements in quarterly forecasts of GDP growth over traditional models that either rely only on quarterly macroeconomic data or average daily frequency data. The evidence suggests that this methodology improves the forecasts for the Mexican GDP notwithstanding its higher volatility relative to that of developed countries. Furthermore, we explore the ability of the MIDAS model to provide forecast updates for GDP growth (nowcasting).

*JEL Classifications:* C22, C53, E37

*Keywords:* GDP forecasting, mixed frequency data, daily financial data, nowcasting

Forecasting influences the economy as a whole, as individuals and policymakers rely on predictions of macroeconomic variables to make decisions. Consequently, these predictions must provide a good approximation of the realizations of the variable of interest. In turn, the accuracy of the forecasts depends on the information set and the forecasting model.

Financial data, such as stock prices and interest rates, contain potentially useful information for making predictions due to their forward-looking nature. Exploiting this type of data, however, presents some challenges. First, financial information is sampled at a much higher frequency than macroeconomic variables such as gross domestic product (GDP). These macroeconomic variables

**ACKNOWLEDGMENTS** The authors thank Bernardo Guimaraes, Juan R. Hernández, Jorge Herrera, and Gonzalo Rangel for valuable comments. José A. Jurado and Andrea Miranda provided excellent research assistance. Support provided by CONACYT is gratefully acknowledged.

are typically available on a quarterly basis, whereas many financial variables are sampled on a daily basis. The standard approach to using this information to make forecasts is to average the high-frequency financial data in the quarter, that is, to employ a flat aggregation weighting scheme, to be able to estimate a regression with quarterly data. This method might not be optimal, however: for instance, if more recent data are more informative, they should receive a higher weight than earlier data. A simple linear regression would require estimating a large number of parameters, thus leading to high estimation uncertainty. One possible way to overcome this difficulty is to use the mixed data sampling (MIDAS) approach.[1] The MIDAS approach consists of regressions that allow the forecast variable and the regressors to be sampled at different frequencies, using distributed lag polynomials to achieve parsimony. This family of models has been used in recent literature to improve the accuracy of predictions of quarterly GDP with monthly indicators.[2] Additionally, Andreou, Ghysels, and Kourtellos use the MIDAS model to forecast GDP growth with financial data.[3] These studies conclude that the use of mixed-frequency data improves forecast accuracy.

A second challenge is how to incorporate all the available information in such a way that the model remains parsimonious. In this regard, some methods are potentially useful to deal with large sets of financial variables, such as factor models and forecast combinations, and a wide variety of model parameterization options considerably reduce the number of estimated coefficients. Factor models are useful to summarize the information content of large data sets with a few common factors.[4] Forecast combinations improve accuracy over individual forecasts by exploiting information from a set of models rather than relying on a single model.[5] In this paper, we employ factor models and forecast combinations as complementary approaches. That is, we use forecast combinations of models estimated with different factors extracted from the group of financial variables.

The MIDAS approach has not been previously applied to developing economies to forecast GDP. The volatility of economic variables in these countries tends to be high, which affects forecast accuracy. In this paper, we investigate whether the proposed methodologies lead to improvements in short-term

1. Ghysels, Santa-Clara, and Valkanov (2004); Ghysels, Sinko, and Valkanov (2007).
2. For example, Clements and Galvão (2008); Marcellino and Schumacher (2010).
3. Andreou, Ghysels, and Kourtellos (2013).
4. Stock and Watson (2002).
5. Timmermann (2006).

forecasting of the Mexican GDP growth rate. For this purpose, we obtained a large set of 392 financial variables from Bloomberg. These variables can be grouped in the following categories: commodities, equities, corporate risk, foreign exchange, and fixed income. This data set constitutes the main information source for our analysis. Because of the large number of variables, we use factor analysis to summarize all the information. These factors then provide the basis for estimating the MIDAS model and obtaining forecasts for different specifications at horizons of one and four quarters ahead. We then compare the performance of the MIDAS models with traditional benchmark models that only use quarterly macroeconomic data.[6] We also use forecast combinations to further improve accuracy. Finally, we present the GDP forecasts from a MIDAS regression model using a monthly data set of macroeconomic variables, following Marcellino and Schumacher.[7] This enables us to assess the performance of daily financial variables vis-à-vis monthly variables.

Our most important result is that the inclusion of daily financial data and the use of the MIDAS regression model to forecast quarterly GDP growth does improve accuracy over traditional models for Mexico. Furthermore, in line with existing literature, we find that forecast combinations are effective at improving the predictive ability of a set of models. We conclude that the methodologies described herein are successful at incorporating additional information while preserving parsimony.

Our paper also provides statistical comparisons of the forecasting ability of the MIDAS model. First, we investigate whether the MIDAS model that incorporates daily financial data leads to improvements for quarterly forecasts of GDP growth over traditional models that rely only on quarterly macroeconomic data. Second, we compare the MIDAS model against a flat aggregation weighting scheme. Finally, we explore the ability of the MIDAS model to provide forecast updates of GDP growth using recent information (nowcasting). Our results show that the model with financial data and quarterly macroeconomic data outperforms a model that only employs quarterly macroeconomic variables and that the MIDAS model outperforms the flat aggregation

6. Alternative methods of using high-frequency data to predict quarterly GDP growth include bridge models (Baffigi, Golinelli, and Parigi, 2004), state-space models (Mariano and Murasawa, 2003), and factor models (Giannone, Reichlin, and Small, 2008). While bridge models and state-space models rely on small sets of variables, factor models exploit large data sets by summarizing the information into a few common factors. Our paper focuses exclusively on MIDAS models, although comparisons of forecasts from MIDAS models with some of these methods would clearly be of interest for future research.

7. Marcellino and Schumacher (2010).

scheme in terms of accuracy. The MIDAS model is useful to provide updates of GDP growth, although the forecasts with leads seem to have a similar predictive accuracy compared to the short-run forecasts without leads.

The rest of the paper is organized in the following way. The next section introduces the MIDAS regression model, factor analysis, and forecast combination. An overview of the data set is then presented, followed by a discussion of the results. A final section concludes the paper. Supplemental results are provided in the appendix.

## Methodology: The MIDAS Model

To illustrate the MIDAS model, we focus on two of the variables used in the study: Mexican GDP growth as the dependent variable and the Goldman Sachs Commodity Index (GSCI) of silver as the independent variable. GDP growth is sampled quarterly, while the GSCI index is sampled daily.

We define $Y_t^Q = GDP_t$ and $X_{m,t}^D = GSCI_t$, where $Q$ stands for quarterly, $D$ stands for daily, and $m$ is the number of trading days in a quarter. Using this notation, a prediction of the GDP growth rate $h$ periods in the future with the model proposed by Ghysels, Santa-Clara, and Valkanov and Ghysels, Sinko, and Valkanov has the following form:[8]

$$Y_{t+h}^{Q,h} = \mu^h + \sum_{j=0}^{p_Y^Q-1} \rho_{j+1}^h Y_{t-j}^Q + \beta^h \sum_{j=0}^{q_X^D-1} \sum_{i=0}^{m-1} w_{i+j*m}^{\theta^h} X_{m-i,t-j}^D + u_{t+h}^h.$$

This model has a constant, the traditional autoregressive (AR) terms with $p_Y^Q$ quarterly lags of the dependent variable (denoted by $\rho_{j+1}^h$), and a term that incorporates $q_X^D$ times $m$ daily lags for the independent variable. The term multiplying the daily variable $w_{i+j*m}^{\theta^h}$ deserves special attention. This term is the weighting scheme that will reduce the number of parameters to estimate and lead to a more parsimonious model.

Ghysels, Sinko, and Valkanov describe five weighting schemes that significantly reduce the number of parameters to estimate.[9] First, the U-MIDAS is an unrestricted version of the model in the sense that every high-frequency lag has its own coefficient to estimate. It can be useful when $m$ is small. A desirable

8. Ghysels, Santa-Clara, and Valkanov (2004); Ghysels, Sinko, and Valkanov (2007).
9. Ghysels, Sinko, and Valkanov (2007).

characteristic of these weights is that they can be estimated using traditional ordinary least squares.

Second, the normalized beta probability function has the following form, consisting of three parameters:

$$w_i(\theta_1, \theta_2, \theta_3) = \frac{a_i^{\theta_1 - 1}(1 - a_i)^{\theta_2 - 1}}{\sum_{i=1}^{N} a_i^{\theta_1 - 1}(1 - a_i)^{\theta_2 - 1}} + \theta_3,$$

where $a_i = (i - 1)/(N - 1)$, with $i = 1, 2, \ldots, N$. This scheme can be made more parsimonious by restricting the first parameter to be unitary or the third parameter to be zero (or both).[10] If all of these parameters are unrestricted, this weighting scheme is called non-zero beta. $N$ denotes the number of high-frequency lags used in the regression.

Third, the normalized exponential Almon lag polynomial consists of two parameters, represented as

$$w_i(\theta_1, \theta_2) = \frac{\exp(\theta_1 i + \theta_2 i^2)}{\sum_{i=1}^{m} \exp(\theta_1 i + \theta_2 i^2)}.$$

As with the previous weighting scheme, the second parameter can be restricted to be zero.

Fourth, the Almon lag polynomial is unable to identify the parameter β. Therefore,

$$\beta w_i(\theta_0, \ldots, \theta_P) = \sum_{p=0}^{P} \theta_p i^p.$$

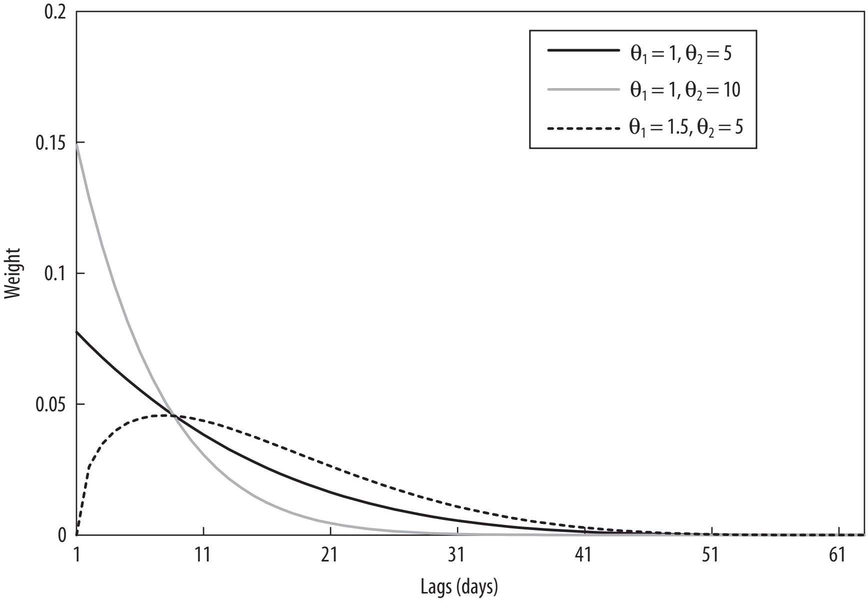
The order of the polynomial $P$ is chosen by the researcher.

Fifth, the step functions are also unable to identify β:

$$\beta w_i(\theta_0, \ldots, \theta_P) = \theta_1 I_{i \in [a_0, a_1]} + \sum_{p=2}^{P} \theta_p I_{i \in [a_{p-1}, a_p]}.$$

10. The beta function described above follows from Ghysels (2015) and approximates the beta function described in Galvão (2013) as $\beta(\theta_1, \theta_2) = [a_i^{\theta_1 - 1}(1 - a_i)^{\theta_2 - 1}\, \Gamma(\theta_1 + \theta_2)]/[\Gamma(\theta_1)\Gamma(\theta_2)]$, where Γ is the gamma function.

**FIGURE 1. Beta Polynomial Weighting Function**[a]




a. The figure plots the weights on the first sixty-three lags of the beta polynomial function for different values of the parameters.

with $a_0 = 1 < a_1 < \cdots < a_p = m$. $I$ is an indicator function with a value of one whenever $i$ lies within the specified interval and is zero otherwise.

With the exception of the U-MIDAS and the Almon lag polynomial, the above schemes are estimated by nonlinear least squares. As described in Ghysels, Sinko, and Valkanov, the exponential Almon lag and the beta probability functions are flexible enough to accommodate various shapes, such as slow-declining, fast-declining, or hump-shaped patterns.[11] In contrast, the unrestricted MIDAS and the step-function schemes impose less structure on the function. Thus, these schemes can result in nonmonotonic shapes, possibly associated with mean-reverting effects of high-frequency variables on the dependent variable. The advantage of those schemes is that they can be estimated through OLS, but they require a larger number of parameters to estimate. As shown in our results, we find that the beta function seems to perform better in terms of forecasting accuracy. Figure 1 shows various shapes

11. Ghysels, Sinko, and Valkanov (2007).

of the beta weighting function for several values of the parameters. The rate of decay is governed by the values of the parameters.

The more traditional way of using high-frequency data is to make an average, which is called a flat aggregation scheme. In our case, that would mean averaging the GSCI daily index for each quarter, that is, assigning the same weight to all the lags in a quarter. Although this scheme is widely used in the literature, it may not be optimal for time series that exhibit memory decay. In contrast, the MIDAS regression allows us to choose the optimal shape of the weights.

### *Factor Models*

Following Stock and Watson, we use factor models to condense the information of a large number of variables into a few factors.[12] Stock and Watson find that factor models are useful to improve the forecasts of key macroeconomic variables, such as output and inflation. The goal is to obtain a small set of factors that explains an important part of the variation in the entire set of variables. Formally, suppose there is a large set of variables $\mathbf{X}$ that will be used for forecasting. This set contains $N$ variables with $T$ observations each. It is possible that $N > T$. The goal is to find a set of factors $\mathbf{F}$ and a set of parameters $\mathbf{\Lambda}$ that best explain $\mathbf{X}$. The factor model can be written as:

$$\mathbf{X}_t = \mathbf{\Lambda}\mathbf{F}_t + \mathbf{e}_t$$

where $\mathbf{e}_t$ are idiosyncratic disturbances with limited cross-sectional and temporal dependence. Another way to look at a factor is to think of it as an unobservable variable that explains an important part of the variation of the observed variables.

To estimate the factors, Stock and Watson propose using the method of principal components, which consists of minimizing the following expression:[13]

$$V\left(\tilde{\mathbf{F}}, \tilde{\mathbf{\Lambda}}\right) = (NT)^{-1} \sum_i \sum_t \left(\mathbf{X}_{it} - \tilde{\boldsymbol{\lambda}}_\iota \tilde{\mathbf{F}}_t\right)^2,$$

where $\tilde{\boldsymbol{\lambda}}_\iota$ is the $i$th row of $\tilde{\mathbf{\Lambda}}$. Most of the literature has focused on extracting factors at low frequencies, such as quarterly or monthly data. Following this approach, we extract factors from a large set of daily financial variables. Once

12. Stock and Watson (2002).
13. Stock and Watson (2002).

the factors are estimated, they are incorporated into the MIDAS regression as a high-frequency variable. For instance, if we use the factor that explains the largest variation of the entire set of financial variables, denoted by $\mathbf{F}^1$, as the high-frequency regressor, our MIDAS regression model can be written as follows:

$$Y^{Q,h}_{t+h} = \mu^h + \sum_{j=0}^{p^Q_Y-1} \rho^h_{j+1} Y^Q_{t-j} + \beta^h \sum_{j=0}^{q^D_X-1} \sum_{i=0}^{m=1} w^{\theta^h}_{i+j*m} \mathbf{F}^1_{m-i,t-j} + u^h_{t+h}.$$

In our case, the first factor accounts for 23 percent of the variability of the 392 daily time series used. The first five factors explain 42.7 percent of underlying variation. The next section presents more details about the data set.

Following Marcellino, Stock, and Watson, we standardize the series before obtaining the factors, by subtracting their means and dividing by their standard deviations.[14] This is necessary because a wide variety of series are employed and they differ in their units of measurement. Two approaches can be used to estimate the factors, namely, the static and the dynamic methods. In this paper, we employ the static method. According to Stock and Watson, the static method is both parsimonious and robust to having temporal instability in the model, as long as the instability is relatively small and idiosyncratic.[15]

### *Forecast Combinations*

To employ the information contained in several of the estimated factors without increasing the number of parameters in the model, we use forecast combination methods. In this way, we can include the information contained in an important number of explanatory variables. By preserving parsimony, we achieve lower parameter uncertainty, thus improving forecasting accuracy. Furthermore, the use of forecast combinations allows us to construct forecasts from a relatively large number of possible parameterizations of the MIDAS model. Thus, forecast combinations deal with the problem of model uncertainty by using information from alternative models instead of focusing on a single model.[16]

The literature finds, as a general result, that forecast combinations improve forecast accuracy.[17] Following Andreou, Ghysels, and Kourtellos, we present a

14. Marcellino, Stock, and Watson (2003).
15. Stock and Watson (2002, 2008).
16. For a survey of forecast combination methods, see Timmermann (2006).
17. Timmermann (2006).

few combinations that improve the root-mean-squared forecast error (RMSFE) of the individual predictions.[18] Formally,

$$\hat{Y}^{Q,h}_{CM,t+h} = \sum_{i=1}^{n} w^{h}_{i,j} \hat{Y}^{Q,h}_{i,t+h}.$$

Thus, a forecast combination $\hat{Y}_{CM,t+h}$ can be interpreted as a weighted average of the $n$ forecasts $\hat{Y}_{i,t+h}$ for the horizon $h$ of $n$ models. Again, an important decision is the selection of the weighting scheme. For this purpose, we need to think in terms of a loss function. Formally, a combination of $n$ forecasts is preferred to a single forecast if

$$E\left[\mathcal{L}\left(\hat{\mathrm{Y}}^{Q,h}_{i,t+h}, Y_{t+h}\right)\right] > \min_{c(\cdot)} E\left[\mathcal{L}\left(C\left(\hat{\mathrm{Y}}^{Q,h}_{1,t+h}, \hat{\mathrm{Y}}^{Q,h}_{2,t+h}, \ldots, \hat{\mathrm{Y}}^{Q,h}_{n,t+h}\right), Y_{t+h}\right)\right],$$

for $i = \{1, 2, \ldots, n\}$.

In the inequality above, $\mathcal{L}$ is a loss function that relates the forecast and observed values. Intuitively, the loss function is expected to grow as the forecast value drifts further from the actual value. $C$, on the other hand, is the combination function that relates the individual forecasts. Thus, we would like to select a function $C$ that minimizes the expected loss, and the forecast combination would be preferred if the expected value of the loss function for that combination is smaller than each of the expected losses for each of the individual forecasts.

Given the previous assumptions, the solution is a linear combination of individual forecasts. To finish this derivation, let us denote by $\hat{\mathbf{Y}}^{Q,h}_{t+h}$ a vector containing all individual forecasts and by $\mathbf{w}^{h}_{t+h}$ a vector of parameters. Then, the combination function can be rewritten as $C(\hat{\mathbf{Y}}^{Q,h}_{t+h}; \mathbf{w}^{h}_{t+h})$. The last step requires defining a loss function. We use the mean-squared forecast error (MSFE), which has been found to provide the highest improvement in forecasts.[19] Thus, the MSFE weights are selected by analyzing the historical forecasting performance of the model and assigning each forecast a weight inversely proportional to its MSFE.

## *Nowcasting*

The MIDAS models have the ability to incorporate recent information to improve the forecasts. Suppose that current-quarter GDP growth needs to

18. Andreou, Ghysels, and Kourtellos (2013).
19. Andreou, Ghysels, and Kourtellos (2013).

be predicted. If we are one month into the current quarter (that is, at the end of January, April, July, or October), we will have about twenty-one trading days (one month) of daily data to forecast quarterly economic growth. Using the information to date to forecast the next value of a variable of interest is called nowcasting.

Formally, the MIDAS model is augmented with leads in the following way:

$$Y_{t+h}^{Q,h} = \mu^h + \sum_{j=0}^{p_Y^Q-1} \rho_{j+1}^h Y_{t-j}^Q + \beta^h \left[ \sum_{i=(3-J_x)*\frac{m}{3}}^{m-1} w_{i-m}^{\theta^h} X_{m-i,t+1}^D \sum_{j=0}^{q_X^D-1} \sum_{i=0}^{m-1} w_{i+j*m}^{\theta^h} X_{m-i,t-j}^D \right] + u_{t+h}^h.$$

The new term has two noticeable aspects. First, the subindex $t + 1$ for the financial variable $X^D$ implies that the forecasting equation includes high-frequency information generated during the present quarter. The other important thing to notice is the values of $i$ and $J_X$. If $m = 63$, then there are sixty-three trading days in a quarter. If the first month of the quarter has just finished, there are twenty-one days of data available, so $J_X = 1$ needs to be selected to obtain the appropriate limits of the sum.

Traditional nowcasting involves state-space models, potentially implying a large number of parameters and measurement equations. In contrast, the MIDAS approach provides a parsimonious framework to deal with a large number of high-frequency predictors.

## *Forecast Evaluation*

To compare the forecasting ability of alternative models, we use the Diebold-Mariano test.[20] That is, we test for the null hypothesis that two different models have the same forecasting ability. To that end, we define a quadratic forecast loss function for model $i$ as $g(u_{i,t}) = u_{i,t}^2$. Under the null hypothesis, both models have equal forecasting ability, that is,

$$H_0 : g(u_{1,t}) = g(u_{2,t}).$$

Diebold and Mariano first define the difference between the loss functions for two alternative models as $d_t = g(u_{1,t}) - g(u_{2,t})$.[21] They then propose the following test statistic:

20. Diebold and Mariano (1995).
21. Diebold and Mariano (1995).

$$DM = \frac{\bar{d}}{\sqrt{\mathrm{var}(d)}},$$

where $\bar{d}$ is the sample mean of $d_t$ and $\sqrt{\mathrm{var}(\bar{d})}$ is defined as

$$\sqrt{\mathrm{var}(d)} = \frac{\gamma_0 + 2\gamma_1 + \cdots + 2\gamma_q}{H-1}.$$

$H$ is the number of forecast periods and $\gamma_j = \mathrm{cov}(d_t, d_{t-j})$. The statistic has a $t$-student distribution with $H-1$ degrees of freedom. The $p$ values shown later in the paper are derived from a regression with robust errors of $d_t$ on a constant, where we test whether the constant is statistically significant.

### *Alternative Models*

To analyze the relative performance of the MIDAS model, we estimate the following alternative models: an autoregression (AR), a random walk (RW), a vector autoregression (VAR), a Bayesian vector autoregression (BVAR) model, and a dynamic stochastic general equilibrium (DSGE) model. We also compare our results to the Survey of Professional Forecasters (SPF). These models and the survey are widely used by both central banks and the empirical literature as benchmarks for GDP forecasting.[22] The AR and RW models contain seasonal dummy variables. We used the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to choose the order of the AR model, which resulted in one autoregressive lag.

VAR models represent a systematic way to capture the dynamics and comovements of a set of time series without restricting for a specific functional form. They have been particularly useful for forecasting purposes since the influential paper by Sims.[23] The VAR model can be written as follows:

$$\mathbf{Y}_t = \mathbf{A}_0 + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{Y}_{t-i} + \boldsymbol{\varepsilon}_t,$$

where $\mathbf{Y}_t$ is the vector of variables being forecast, $\mathbf{A}_i$ are the matrices of coefficients to estimate, and $\boldsymbol{\varepsilon}_t$ is a vector of residuals. The variables included

22. Chauvet and Potter (2013).
23. Sims (1980).

in the VAR model are the GDP growth rate, quarterly inflation rate, interest rate. and U.S. GDP growth rate.[24] To determine the number of lags $p$, we use the AIC and set the maximum number of lags to four. The model can also contain seasonal dummy variables that are not included in the equation above for simplicity.

A limitation of VAR models is that they often imply a large number of parameters to estimate, resulting in a loss of degrees of freedom and thus in inefficient estimates and lower forecasting performance. To deal with this limitation, we estimate a Bayesian VAR (BVAR) model.[25] The idea is to use an informative prior to shrink the unrestricted VAR model toward a parsimonious naïve benchmark, thereby reducing parameter uncertainty and improving forecasting accuracy. Previous studies find that BVAR models have a good forecasting performance compared to conventional macroeconomic models for different countries and periods.[26]

A BVAR model requires specifying the mean and standard deviation of the prior distribution of the parameters. We follow the Minnesota prior, in which each variable follows a random walk around a deterministic component.[27] If the model is specified in first differences, this prior specification shrinks all of the elements of $\mathbf{A}_i$ toward zero. This implies that each variable depends mainly on its own first lag. In addition, the Minnesota prior incorporates the belief that more recent lags should provide more reliable information than more distant ones and that own lags explain more of the variation of a given variable than lags of other variables in the equation. The prior beliefs are imposed by setting the following moments for the prior distribution of the parameters:

$$E\left[\left(\mathbf{A}_k\right)_{ij}\right]=0, \quad V\left[\left(\mathbf{A}_k\right)_{ij}\right]=\begin{cases}\dfrac{\lambda^2}{k^{2\tau}}, & j=i\\ \left(\dfrac{\lambda^2\gamma^2}{k^{2\tau}}\right)\left(\dfrac{\sigma_i^2}{\sigma_j^2}\right), & \text{otherwise.}\end{cases}$$

Thus, the Minnesota prior can be described by three hyperparameters: the overall tightness parameter, $\lambda$; the relative cross-lags parameter, $\gamma$; and the

24. Herrera Hernández (2004) and Capistrán and López-Moctezuma (2010) find that U.S. GDP is useful for improving Mexican GDP forecasts in a VAR framework.

25. Litterman (1986); Doan, Litterman, and Sims (1984).

26. See, for example, Litterman (1986); McNees (1986); Artis and Zhang (1990); Bańbura, Giannone, and Reichlin (2010).

27. Litterman (1986).

decay parameter, $\tau$. Changes in these parameters imply changes in the variance of the prior distribution. The overall tightness parameter $\lambda$ indicates the tightness of the random-walk restriction, or the relative weight of the prior distribution with respect to the information contained in the data. For $\lambda = 0$, the data do not influence the estimates. As $\lambda \rightarrow \infty$, the posterior estimates converge to the OLS estimates. The parameter $\gamma < 1$ indicates the extent to which the lags of other variables are less informative than own lags. The parameter $\tau \geq 0$ captures the extent to which more recent lags contain more information than more distant ones. Thus, the factor $1/k^{2\tau}$ represents the rate at which prior variance decreases with increasing lag length. Additionally, $\sigma_i^2/\sigma_j^2$ accounts for the different scale and variability of the series; $\sigma_i$ and $\sigma_j$ are estimated as the standard errors of a univariate AR regression for each variable. Finally, we use a noninformative (diffuse) prior for the deterministic variables. The BVAR model is estimated using Theil's mixed estimation method.[28]

The hyperparameters are chosen based on forecasting performance. In particular, we estimate the BVAR model for the combinations resulting from setting the following parameters: $\lambda = \{0.1,0.2\}$, $\gamma = \{0.3,0.5\}$, $\tau = 1$, and the number of lags $p = \{1, 2, 3, 4\}$.[29] From these sixteen combinations of hyperparameters, we select the combination that minimizes the RMSFE in a pseudo out-of-sample forecasting exercise.

To provide further evidence of the forecasting accuracy of the MIDAS model, our forecasts are also compared with those of the Survey of Professional Forecasters, which is maintained by the Bank of Mexico. Capistrán and López-Moctezuma find that the forecasts from this survey outperform forecasts from traditional univariate and multivariate time series models.[30] There are about thirty survey participants, including financial, consulting, and academic institutions. We use the consensus forecast for the GDP growth rate, defined as the mean across forecasters. For the forecasting period used in this paper, the data are only available at the one-quarter-ahead horizon.

We also generate the GDP forecasts using a DSGE model.[31] Because DSGE models allow the researcher to represent an economy based on microfoundations, they are widely used by central banks as a tool for policy analysis

28. Theil and Goldberger (1961).

29. These values for the hyperparameters are used in the literature (for example, Dua and Ray, 1995; LeSage, 1999, pp. 154–59; Canova, 2007).

30. Capistrán and López-Moctezuma (2010). Their paper provides an in-depth description of this survey.

31. The model is from Rubaszek and Skrzypczyński (2008).

and forecasting.[32] In this paper, we use a benchmark DSGE model that has proven useful for forecasting GDP in the literature.[33] The economy consists of households that maximize lifetime utility, firms that maximize profits, and a monetary authority that cares about price and output stability. The model is subject to demand, productivity, and monetary shocks. It includes three core equations: a dynamic IS curve, a forward-looking Phillips curve, and a monetary policy rule, which determine the path for output, prices, and interest rates.

The utility function of the representative household is an increasing function of consumption and a decreasing function of labor. The households receive wages from labor, dividends from owned firms, and interest from bonds. There are perfectly competitive final-goods producers that use a continuum of differentiated intermediate goods as inputs, taking their price as given. The intermediate-goods producers operate under monopolistic competitive conditions and hire labor as the only input. Nominal rigidities are introduced following Calvo's sticky-price framework.[34] Finally, the central bank follows a Taylor rule in which the interest rate responds to changes in inflation and the output gap. The model is characterized by the following system of equations:

$$\hat{h}_t = -\frac{1}{\sigma}\left(\hat{R}_t - \hat{\Pi}_{t+1} + \hat{\varepsilon}^D_{t+1} - \hat{\varepsilon}^D_t\right) + \hat{h}_{t+1};$$

$$\hat{\Pi}_t = \frac{\delta}{1+\beta\delta}\hat{\Pi}_{t-1} + \frac{\beta}{1+\beta\delta}\hat{\Pi}_{t+1} + \frac{(1-\beta\xi)(1-\xi)}{(1+\beta\delta)\xi}\widehat{mc}_t;$$

$$\hat{R}_t = \gamma\hat{R}_{t-1} + (1-\gamma)\left[\gamma_\pi\hat{\Pi}_t + \gamma_{\Delta y}\left(\hat{y}_t - \hat{y}_{t-1}\right)\right] + \eta^M_t;$$

$$\hat{\varepsilon}^i_t = \rho^i\hat{\varepsilon}^i_{t-1} + \eta^i_t, \quad i = \{D, S\}.$$

The first equation is the dynamic IS curve. The second equation represents the dynamic Phillips curve, while the third equation presents the interest rate rule followed by the monetary authority. The laws of motion for the demand and supply shocks are represented in the last equation. A hat above a

32. For evidence on the forecasting accuracy of DSGE models, see Smets and Wouters (2004); Adolfson, Lindé, and Villani (2007); Liu, Gupta, and Schaling (2009); Rubaszek and Skrzypczyński (2008); Edge, Kiley, and Laforte (2010); del Negro and Schorfheide (2013).

33. Rubaszek and Skrzypczyński (2008); del Negro and Schorfheide (2013).

34. Calvo (1983).

variable indicates a deviation from its steady-state value. In the equations, $y_t$ is detrended output, $h_t$ is habits, $R_t$ is the nominal interest rate, $\Pi_t$ is the inflation rate, $mc_t$ is the marginal cost of ouput, $\sigma$ is the inverse of the intertemporal elasticity of substitution, $\lambda$ is the habit formation parameter, $\xi$ measures the degree of price stickiness, $\delta$ is the degree of price adjustment, $\beta$ is the discount factor, $\gamma$ and $\rho^i$ are persistence parameters, and $\eta^i_t$ are independent and identically distributed (i.i.d.) white noise disturbances for $i = \{D, S, M\}$. Habit formation is given by $(1 - \lambda)\, \hat{h}_t = \hat{y}_t - \lambda \hat{y}_{t-1}$. Marginal costs can be written as $\widehat{mc}_t = \sigma \hat{h}_t + \varphi \hat{L}_t - \hat{\varepsilon}^S_t$, where labor supply is given by $\hat{L}_t = \hat{y}_t - \hat{\varepsilon}^S_t$ and $\varphi$ is the elasticity of labor supply.

The system is transformed into a state-space representation, where the observable variables are expressed in terms of the model variables. The model is estimated using a Bayesian approach. The assumptions about the prior distributions of the parameters follow those of Rubaszek and Skrzypczyński, which in turn are consistent with those of Ireland; Smets and Wouters; and Del Negro and others.[35] The priors are combined with the conditional density of the observables to obtain the posterior distribution. The posterior densities are estimated with the Markov chain Monte Carlo algorithm for each quarter of the evaluation sample, using 25,000 draws. Table A1 in the appendix shows the prior and posterior distribution of the parameters.[36]

## Data

We use three databases in our analysis at different sampling frequencies: daily, monthly, and quarterly. The daily database is divided into five different categories of financial information: commodities (166 series), equities (94 series), foreign exchange (27 series), corporate risk (53 series), and fixed income (52 series). As previously stated, the dependent variable is Mexican GDP. Most of these daily financial series are considered to be good predictors of output growth.[37] The initial estimation period is 1999:1 to 2009:4, and the forecasting period is 2009:4 + $h$ to 2013:4. Although the sample is relatively small for nonlinear least squares estimation, Bai, Ghysels, and Wright provide

35. Rubaszek and Skrzypczyński (2008); Ireland (2004); Smets and Wouters (2004); del Negro and others (2007).

36. We checked the sensitivity of the forecasting results on the prior distributions of the parameters. The results are essentially unchanged.

37. Andreou, Ghysels, and Kourtellos (2013).

evidence based on Monte Carlo simulations showing that the forecasting performance of MIDAS regression models may not be affected.[38]

The Mexican GDP time series starts in 1993. The estimation period is effectively shorter, however, because several financial variables are only available from 1999 onward. Although it might be a short period for forecasting purposes, it allows for the inclusion of useful daily information. Moreover, we use a sample period during which Mexico has exclusively followed a floating exchange regime and exclude the 1995 economic crisis from the estimation period, which could affect our estimations.

The constructed database is primarily a subset of the time series suggested by Andreou, Ghysels, and Kourtellos, which has been shown to provide good predictive content for U.S. GDP.[39] We made a number of adjustments to the data set to reflect the Mexican data. First, the twenty-eight-day Mexican Federal Treasury Certificates (CETES) rate is included in the fixed-income group. The interest rate is the primary monetary policy instrument for Mexico. While data for other CETES maturities are not available for the study period, U.S. bonds and bills should partially compensate for this lack of information. Second, the foreign exchange rates are expressed in terms of Mexican pesos. Third, in terms of equity, we use two indexes of the Mexican Stock Exchange, namely, the IPC (the main stock index) and the INMEX (a market capitalization index). Finally, we excluded some of the financial variables specific to Mexico that could be relevant to forecast GDP, such as bonds and commodities, as they are unavailable for the entire study period. All the financial information was retrieved from Bloomberg.[40]

Following Marcellino, Stock, and Watson, we transformed some of the series because they were nonstationary.[41] We tested the hypothesis of unit roots by means of an augmented Dickey Fuller (ADF) test with twelve lags. Nonstationary series were transformed to first log-differences. Then, to ensure stationarity, we tested the transformed series for unit roots using the ADF test. In general, we transform commodity prices, stock prices, and exchange rates into daily returns (that is, first log differences). Interest rates for U.S. corporate bonds are transformed to first differences. Domestic interest rates are found to be stationary in levels. The forecasting variable, that is, the GDP

38. Bai, Ghysels, and Wright (2013).
39. Andreou, Ghysels, and Kourtellos (2013).
40. A detailed description of the series used is available from the authors on request.
41. Marcellino, Stock, and Watson (2003).

growth rate, is not seasonally adjusted. Therefore, regressions are estimated using seasonal dummy variables.

Another important set of information included in our regressions is the quarterly macroeconomic data. This set comprises twenty macroeconomic variables whose high explanatory and predictive power for GDP is documented in the literature.[42] This set contains information such as price indexes, international trade variables, inflation rates, and economic activity indexes for Mexico and the United States. Some of these variables are available on a monthly basis. To transform these variables into quarterly data, monthly data are averaged for every quarter.

In addition to the daily financial variables, a monthly macroeconomic data set is used as the high-frequency data for the MIDAS regression. This set consists of eighteen variables, such as price indexes, economic activity indexes for Mexico, and the U.S. consumer price index (CPI). The same procedure is followed to preserve parsimony: that is, a set of factors is estimated, and different forecasts using each factor are combined to obtain the final forecast. The Mexican data were obtained from the National Institute of Statistics, Geography, and Information (INEGI) and the Bank of Mexico (the Mexican central bank). U.S. data were obtained from the Federal Reserve Bank of St. Louis.

## Results

Before we present our results for the forecasting exercise, a few points require further clarification. First, a recursive window is used for all the model specifications and horizons, unless otherwise stated. For instance, consider forecast $i$, with $i = 0, 1, \ldots, n - 1$, where $n$ is the number of one-step-ahead forecasts. Then, the start date of the estimation is fixed at 1999:1, whereas the end date changes with each forecast value, which is 2009:(4 + $i$). Thus, the model is estimated each time the window changes, and the forecasts are computed one-step-ahead. This window grows with each forecast point, as it includes the next observed value. The recursive window is expected to improve the forecasts over a fixed estimation window, since each new estimation includes more recent information.

Second, our forecasting exercise is not in real time. Given that GDP is subject to revisions (as are other macroeconomic variables used as regressors),

42. Andreou, Ghysels, and Kourtellos (2013).

the data actually available at a particular quarter may differ from the final values that will be released by statistical offices. Ideally, we would perform a real-time forecasting exercise by using the vintages of data that were actually available to the forecasters, but real-time data for Mexico are unavailable. We therefore use revised data in our estimations. However, our models are still comparable in the forecasting evaluation exercise, since all of them use the same information.

As discussed in the introduction, the MIDAS model is capable of nowcasting, using current information. Quarterly GDP figures are published with a lag of two months, on average. However, given that every month has about twenty-one trading days and that all the daily information used is available at the end of the day, as is the case with most of the financial variables, it is theoretically feasible to obtain sixty-three progressive forecasts for a quarter's GDP. Nowcasting provides a way to incorporate new information into the forecast, up to the publishing of an official value for the variable of interest. By the time the first-quarter GDP has been published, two months of information will be available on the second quarter. The model allows the user to use this extra information to forecast the second (or current) quarter, well before the official quarterly GDP figure is published. Even more, because the model can incorporate data up to the publishing date, then up to five months of useful information could be used to estimate and forecast second-quarter GDP, assuming a publishing lag of two months.

We use the Akaike and Bayesian information criteria (AIC and BIC) to select the number of lags for both the autoregressive terms and the high-frequency terms. In our preferred forecasting framework, we use the information from five factors. In particular, we follow a similar approach to Andreou, Ghysels, and Kourtellos and use a forecast combination from the five models estimated with each of the five factors extracted from the entire set of financial variables.[43] That is, we use both factor models and forecast combinations to deal with the large set of financial variables.

We use the beta function, as it presented the lowest RMSFE in most cases. The variance of the RMSFE of this weighting scheme is also smaller. The tests to identify the best models were implemented using a maximum of five lags of the dependent variable and one to six lags of the independent factor ($q_X^D$). Given that the number of trading days in a quarter is $m = 63$, the maximum number of daily lags is $63 \times 6 = 378$. The model selection was done

43. Andreou, Ghysels, and Kourtellos (2013).

**TABLE 1. RMSFE Comparison for Models with No Leads**[a]

| Model | Beta weighting scheme | h = 1 RMSFE | h = 1 RMSFE as % of AR | h = 4 RMSFE | h = 4 RMSFE as % of AR |
|---|---|---|---|---|---|
| Alternative model | | | | | |
| AR | | 1.1348 | 1.0000 | 1.1136 | 1.0000 |
| RW | | 1.2890 | 1.1359 | 3.2330 | 2.9032 |
| VAR | | 1.1156 | 0.9831 | 1.2112 | 1.0877 |
| BVAR | | 0.9285 | 0.8182 | 0.9899 | 0.8889 |
| SPF | | 0.9688 | 0.8537 | | |
| DSGE | | 1.1378 | 1.0026 | 1.0255 | 0.9209 |
| MIDAS model | | | | | |
| All variables F1 | $\beta\,(p=2, q=6)$ | 1.6978 | 1.4961 | 1.8168 | 1.6315 |
| Commodities F1 | $\beta\,(p=1, q=1)$ | 1.5170 | 1.3368 | 1.3902 | 1.2483 |
| Equities F1 | $\beta\,(p=3, q=5)$ | 1.4217 | 1.2528 | 1.5319 | 1.3756 |
| Corporate risk F1 | $\beta\,(p=1, q=2)$ | 1.4375 | 1.2667 | 1.5536 | 1.3951 |
| Foreign exchange F1 | $\beta\,(p=1, q=1)$ | 1.0367 | 0.9135 | 1.0429 | 0.9365 |
| Fixed income F1 | $\beta\,(p=1, q=5)$ | 1.9653 | 1.7319 | 1.9792 | 1.7773 |
| Forecast combinations | | | | | |
| Factors 1 to 5 | $\beta$ Best AIC/BIC | 1.0453 | 0.9211 | 1.1936 | 1.0718 |

a. The table shows the root mean square forecast error (RMSFE) for GDP horizons of $h = 1$ and $h = 4$ for the sample 1999:1–2013:4. Estimation period: 1999:4+$h$ to 2009:4. Forecasting period: 2009:4+$h$ to 2013:4. The RMSFEs are also presented as a percentage of the AR. First, the forecasts are estimated for each of the alternative models described in the paper. Second, the table shows the results for the MIDAS model using the first daily factor of the 392 financial variables. Then, the forecasts are also estimated using the first factor (F1) of each group of financial variables. Finally, a forecast combination based on the first five factors is presented. A recursive window is used for all estimations.

following the BIC. As explained before, regardless of the high-frequency lags specified, the model estimates only two parameters for the beta weighting scheme.

## *Forecasting Results*

Table 1 presents forecasts for different specifications estimated for two different forecasting horizons: one-quarter-ahead ($h = 1$) and one-year-ahead ($h = 4$). Out of the alternative benchmark models, the BVAR and the SPF have the best forecasting performance at the one-quarter-ahead horizon. The BVAR and the DSGE models have the best forecasting performance at the four-quarters-ahead horizon. As expected, the RW model has the highest RMSFE. The forecasting accuracy of the BVAR model is consistent with previous studies for different countries.[44] Similarly, our result about the predictive

44. For example, Artis and Zhang (1990); Bańbura, Giannone, and Reichlin (2010).

accuracy of the DSGE model at the one-year horizon is also consistent with previous literature.[45]

The table also presents the relative RMSFE of the MIDAS model with respect to the benchmark AR model. The RMSFE of the MIDAS model that employs the first factor is outperformed by the benchmark models. A possible explanation is that the benchmark models contain macroeconomic variables with good predictive content for GDP forecasting that are not contained in the MIDAS model. In the last part of this subsection, we present an exercise that incorporates macroeconomic variables into the MIDAS model to provide evidence of the forecasting ability of this methodology and the use of high-frequency data.

Factor estimation is also applied to each group of financial variables. From this decomposition, five factors are extracted, one for each of the five groups of financial variables. Table 1 shows the forecasting results with the first factor of each group. We use the beta weighting scheme and select the number of lags using Akaike and Bayesian information criteria. We only include the first factor in each regression because the variables in each group are highly interrelated. Even though this is a parsimonious weighting specification, the predictive power for all variable groups, except for exchange rates, does not seem to improve over the benchmark models. In other words, the uncertainty associated with parameter estimation for these specifications outweighs the additional predictive power incorporated through the individual sets of financial series. The performance of the exchange rates in forecasting GDP could be explained in part by Mexico's status as a small open economy. On the other hand, the other sets of variables might be less related to Mexican GDP dynamics than they are for U.S. GDP. In particular, the variables included in the corporate risk and fixed income groups focus mainly on the U.S. economy. Although they include some variables specific to Mexico, such as interest rates, these do not seem to provide sufficient information to predict Mexican GDP growth by themselves. Equities might also present a similar problem.

Forecasting accuracy thus does not improve when the individual groups of financial variables are included in the model. However, when all variables are included together and the factors contain mixed information, they are clearly successful at improving the forecasting accuracy of the model. The last section of the table presents a forecast combination based on the RMSFE

45. Del Negro and Schorfheide (2013).

**TABLE 2. RMSFE Comparisons of Alternative Models, Non-Seasonally Adjusted GDP[a]**

| | $h=1$ | | $h=4$ | |
|---|---|---|---|---|
| *Model and data used* | *RMSFE* | *RMSFE as % of AR* | *RMSFE* | *RMSFE as % of AR* |
| Traditional models | | | | |
| AR | 1.1348 | 1 | 1.1136 | 1 |
| RW | 1.2899 | 1.1367 | 3.2331 | 2.9033 |
| VAR | 1.1156 | 0.9831 | 1.2112 | 1.0877 |
| BVAR | 0.9285 | 0.8182 | 0.9899 | 0.8889 |
| SPF | 0.9687 | 0.8537 | | |
| DSGE | 1.1378 | 1.0026 | 1.0255 | 0.9209 |
| Quarterly macroeconomic data | | | | |
| Factor AR | 0.7407 | 0.6527 | 0.7308 | 0.6563 |
| Monthly + quarterly macroeconomic data | | | | |
| $\beta$ ($p=4, q=3$) | 0.9671 | 0.8522 | 1.1151 | 1.0014 |
| Combined MIDAS | 0.9827 | 0.8659 | 1.1415 | 1.0250 |
| Financial data | | | | |
| Flat aggregation | 1.8181 | 1.6021 | 1.5537 | 1.3952 |
| $\beta$ ($p=2, q=6$) | 1.6978 | 1.4961 | 1.8168 | 1.6315 |
| Combined MIDAS | 1.0453 | 0.9211 | 1.1936 | 1.0718 |
| Financial data + quarterly macroeconomic data | | | | |
| Flat aggregation | 0.7159 | 0.6309 | 0.6682 | 0.6000 |
| $\beta$ ($p=2, q=1$) | 0.4709 | 0.4150 | 0.5131 | 0.4608 |
| Combined MIDAS | 0.4614 | 0.4066 | 0.5003 | 0.4492 |

a. The table shows the root mean square forecast error (RMSFE) for GDP horizons of $h = 1$ and $h = 4$ for the sample 1999:1–2013:4. Estimation period: 1999:4+$h$ to 2009:4. Forecasting period: 2009:4+$h$ to 2013:4. The RMSFEs are also presented as a percentage of the AR. The five MIDAS forecasts estimated from each of the daily factors are combined to obtain the combined MIDAS. A recursive window is used for all forecasts.

using five MIDAS specifications, one for each of the first five factors. Each of these models is optimal in the AIC-BIC sense, but for different factors. As expected, the combination yields a lower RMSFE. This improvement can be explained by the fact that it considers the information contained in each factor. The results suggest that forecast combinations improve the accuracy of different information sets.

The goal of the final part of this section is to investigate whether introducing daily financial data into a MIDAS regression framework is useful for forecasting GDP beyond macroeconomic data. We also compare the forecasting accuracy of the MIDAS model with the traditional models that take a simple average of daily financial data, that is, a flat aggregation scheme.

Table 2 contains a summary of the RMSFE for several models. The factor autoregressive (FAR) model incorporates quarterly macroeconomic data in the AR model using a factor model as in Stock and Watson.[46] In particular, we extract three quarterly macroeconomic factors from the database of twenty quarterly macroeconomic series described in the data section. As a result, the first three factors explain nearly 76 percent of the overall variation of the twenty quarterly macroeconomic series.[47] These estimated factors augment the benchmark AR model to obtain the FAR models. A second family of MIDAS models uses the same set of monthly macroeconomic variables that were averaged using a flat aggregating scheme as the high frequency data. That is, they use both monthly and quarterly data. The flat aggregation models take an average of the values for all trading days of the daily financial assets within the quarter to obtain a single value per quarter. The combined MIDAS models use a combination of the MSFE of five MIDAS specifications: one for each of the first five factors. Finally, the financial data models incorporate the information contained in the 392 daily financial series.

As before, the RMSFE for different specifications is presented in table 2. The results show that adding quarterly data to the AR model improves forecasting accuracy in terms of the RMSFE at both horizons. In particular, the factor model that includes quarterly macroeconomic data outperforms the AR, VAR, BVAR, and SPF forecasts. That is, quarterly macroeconomic data such as consumption, investment, trade, inflation, and foreign macroeconomic variables seem to provide important information to predict future GDP. The monthly macroeconomic data also improve forecast accuracy at the one-quarter-ahead horizon, but the gains are smaller than for quarterly macroeconomic data.

Adding financial data also improves forecasting accuracy, especially for a combined MIDAS mode. The results further suggest that the gains from the inclusion of financial data, in terms of RMSFE, are larger under a MIDAS regression scheme than under a flat aggregation scheme.

The last part of table 2 shows the results of adding the financial data to the specifications that include quarterly macroeconomic data. The results illustrate that adding daily financial data to a MIDAS regression scheme improves forecasting accuracy over a traditional model that contains only quarterly macroeconomic data at both forecast horizons. Finally, forecast combinations

46. Stock and Watson (2002).

47. Ibarra (2012) finds that for the case of Mexico, the estimated factors from a broad set of macroeconomic variables for the period 1992–2009 are highly related to relevant subsets of key macroeconomic variables, such as output and inflation. That is, the estimated factors seem to be informative and interpretable from an economic point of view. Our results are consistent with those findings.

**TABLE 3. Tests of Equal Predictive Ability**[a]

| Model | *h* = 1 DM | *h* = 4 DM |
|---|---|---|
| Financial data versus AR | | |
| Combined MIDAS | 0.2198 | 0.2000 |
| Financial data versus Quarterly macroeconomic data | | |
| Combined MIDAS | 0.1830 | 0.0920 |
| Financial + macroeconomic data versus Quarterly macroeconomic data | | |
| Combined MIDAS | 0.0023 | 0.5010 |
| Financial + macroeconomic data versus Monthly macroeconomic + quarterly macroeconomic data | | |
| Combined MIDAS | 0.0108 | 0.0590 |
| MIDAS versus Flat aggregation | | |
| Financial data | 0.1178 | 0.5560 |
| Financial + macroeconomic data | 0.0000 | 0.0270 |

a. The table reports *p* values of a test for the null hypothesis that the models shown in the left column have equal predictive ability. The comparison is based on a Diebold-Mariano test. Sample period: 1999:1 to 2013:4. Estimation period: 1999:4+*h* to 2009:4. Forecasting period: 2009:4+*h* to 2013:4. A recursive window is used for all forecasts.

of MIDAS regression models based on different groups of financial variables improve forecasting accuracy.

In sum, the combined MIDAS model with financial data has, in general, a lower RMSFE than the benchmark models. However, the macroeconomic regressors help to improve forecasting accuracy in both models. This is not surprising, as they are highly correlated with GDP. Finally, the MIDAS regression approach that incorporates daily financial variables outperforms the flat aggregation scheme.[48]

The tests for equal forecasting ability for different models can be found in table 3. The table shows the *p* values obtained from the Diebold-Mariano test as described earlier. The results show that the null hypothesis of equal forecasting accuracy between the benchmark AR model and the AR model augmented with financial data cannot be rejected. Similarly, the null hypothesis of equal forecasting accuracy between the MIDAS model with financial data

48. We also conducted the forecasting exercise using seasonally adjusted data. The results are presented in table A2 in the appendix. The conclusions are similar to those using data that are not seasonally adjusted. At the one-quarter-ahead horizon, the use of the MIDAS approach and the inclusion of the daily financial variables improve the forecasting accuracy over traditional models that use only quarterly macroeconomic data. However, at the four-quarters-ahead horizon, the DSGE shows a similar predictive ability to the MIDAS model, which is in line with previous literature suggesting that DSGE models perform well at one-year horizons (del Negro and Schorfheide, 2013).

**TABLE 4. RMSFE Comparisons of Alternative Models, Non-Seasonally Adjusted GDP**[a]

| | *h* = 1 | | *h* = 4 | |
|---|---|---|---|---|
| *Data and model used* | *RMSFE* | *RMSFE as % of AR* | *RMSFE* | *RMSFE as % of AR* |
| Financial data | | | | |
| Flat aggregation | 1.3790 | 1.2150 | 1.3190 | 1.1621 |
| Combined MIDAS | 0.9370 | 0.8256 | 1.0338 | 0.9109 |
| Financial data + macroeconomic data | | | | |
| Flat aggregation | 0.6891 | 0.6072 | 0.7130 | 0.6282 |
| Combined MIDAS | 0.4886 | 0.4305 | 0.5301 | 0.4670 |

a. The table shows the root mean square forecast error (RMSFE) for GDP horizons of $h = 1$ and $h = 4$ for the sample 1999:1–2013:4. Estimation period: 1999:4+$h$ to 2009:4. Forecasting period: 2009:4+$h$ to 2013:4. The RMSFEs are also presented as a percentage of the AR. The five MIDAS forecasts estimated from each of the daily factors are combined to obtain the combined MIDAS. A recursive window is used for all forecasts.

and the AR model with quarterly data cannot be rejected at the conventional significance levels.

However, according to the Diebold-Mariano test, the MIDAS model with financial and quarterly data outperforms the model with only quarterly macroeconomic data. That is, the forecasting gains of adding financial data through a combined MIDAS regression model over the traditional approach of using only macroeconomic data are statistically significant at the 5 percent level at the one-quarter-ahead horizon. Moreover, the MIDAS model that includes financial variables is superior to the MIDAS model that includes monthly variables. Although the MIDAS model with financial data and the flat aggregation scheme have similar predictive ability, the MIDAS model that includes quarterly data outperforms the flat aggregation scheme that includes quarterly data. This result suggests that financial factors need to be used alongside macroeconomic variables to extract their full forecasting potential.

In short, using quarterly macroeconomic data and financial data through a MIDAS regression model improves forecasting ability over traditional models that only include macroeconomic data. The results suggest that the inclusion of financial data provides the model with useful information to forecast GDP. Furthermore, the forecast gains of the MIDAS model over the flat aggregation scheme are significant at the conventional levels. We conclude that MIDAS is superior to a simple flat aggregation scheme.

## *MIDAS Forecasts with Leads*

Table 4 shows the results for predicting GDP at horizons of one and four quarters ahead using information one month further into the quarter. This

**TABLE 5. Tests of Equal Predictive Ability**[a]

| | $h=1$ | $h=4$ |
|---|---|---|
| *Model* | *DM* | *DM* |
| Nowcasting versus forecasting | | |
| Flat aggregation (financial data) | 0.0810 | 0.2080 |
| Combined MIDAS (financial data) | 0.2117 | 0.0767 |
| Flat (financial + macroeconomic data) | 0.4110 | 0.2350 |
| Combined MIDAS (financial + macroeconomic data) | 0.1841 | 0.3610 |
| Flat aggregation versus MIDAS | | |
| Financial data | 0.0760 | 0.2294 |
| Financial + macroeconomic data | 0.0001 | 0.0229 |

a. The table reports *p* values of a test for the null hypothesis that the models shown in the left column have equal predictive ability. The comparison is based on a Diebold-Mariano test. Sample period: 1999:1 to 2013:4. Estimation period: 1999:4+*h* to 2009:4. Forecasting period: 2009:4+*h* to 2013:4. A recursive window is used for all forecasts.

exercise illustrates the use of the MIDAS approach for nowcasting, as current quarter information is introduced to provide updates of quarterly GDP growth. As before, the daily financial variables within a MIDAS approach lead to important gains over the benchmark model, especially when the quarterly macroeconomic data are also included. The RMSFEs from the nowcasting exercise are similar to the forecasts shown in table 2 for most of the specifications.

Finally, table 5 shows the Diebold-Mariano test for the nowcasting exercise. From this table, we cannot reject the null hypothesis that nowcasting and forecasting with MIDAS have similar predictive ability. That is, the information contained in the current month does not seem to improve the predictive accuracy for GDP growth.[49] Importantly, we find that MIDAS is statistically superior in its predictive ability over the flat aggregation scheme when both models contain quarterly macroeconomic information.

Most of our results for $h = 1$ are consistent with those obtained by Andreou, Ghysels, and Kourtellos.[50] For $h = 4$, however, Andreou, Ghysels, and Kourtellos find statistically significant differences in predictive power that favor the MIDAS model, whereas our results are more in line with those of Marcellino

49. There are two potential explanations for this result. First, improvements for future GDP predictions could be higher than for current GDP due to the forward-looking nature of financial data. Second, the financial variables in our database could have more predictive power for updating U.S. GDP than Mexican GDP.

50. Andreou, Ghysels, and Kourtellos (2013).

and Schumacher and Arnesto, Engerman, and Owyang.[51] The latter conclude that the forecasting gains of the MIDAS approach over alternative methodologies that employ high-frequency information are smaller for long horizons.

## Conclusion

Following the methodology proposed by Ghysels, Santa-Clara, and Valkanov and by Ghysels, Sinko, and Valkanov, we have estimated a MIDAS model that incorporates a large data set of daily financial variables using factors.[52] We then used this model to generate out-of-sample forecasts for Mexican GDP for horizons of one and four quarters ahead. The results show that the use of this methodology and the inclusion of daily financial variables improve the forecasting accuracy over traditional models that use quarterly macroeconomic data. The MIDAS framework helps to circumvent the problems initially found when dealing with data at different sampling frequencies, while remaining parsimonious. To deal with large sets of financial variables, we use factor analysis and forecast combinations.

The model comparisons favor the use of the MIDAS approach against flat aggregation. In addition, we find that a MIDAS model has a better forecasting performance than an AR model augmented with factors based on macroeconomic variables. In a nowcasting exercise, the results favor the MIDAS model over flat aggregation. However, the MIDAS model with leads seems to have a similar predictive accuracy as the MIDAS model without leads.

We conclude that this methodology improves forecasts even in an emerging economy that displays higher volatility. To improve or extend this work, a different set of financial variables that are more directly related to the Mexican economy could be employed. The unavailability of historic data on those useful variables might no longer be a limitation in a few years. This methodology could also be used to predict other monthly or quarterly macroeconomic variables, such as unemployment or inflation. Finally, as we consider only Mexican data, it would be useful to evaluate whether the MIDAS model that includes daily financial data is also successful at predicting GDP for other developing countries. We leave those extensions for further research.

51. Marcellino and Schumacher (2010); Arnesto, Engerman, and Owyang (2010).
52. Ghysels, Santa-Clara, and Valkanov (2004); Ghysels, Sinko, and Valkanov (2007).

## Appendix: Supplemental Results

**TABLE A1. Prior Distribution and Recursive Mode of Posterior of the DSGE Model Parameters**

| | Prior distribution | | | | Recursive mode of posterior | | | |
|---|---|---|---|---|---|---|---|---|
| *Variable* | *Parameter* | *Type* | *Mean* | *Std. dev. or DF*[a] | *Minimum* | *Median* | *Mean* | *Maximum* |
| Habit formation | $\lambda$ | Beta | 0.70 | 0.10 | 0.56 | 0.61 | 0.61 | 0.67 |
| Elasticity of substitution | $\sigma$ | Normal | 1.00 | 0.38 | 0.68 | 0.70 | 0.72 | 0.81 |
| Labor supply elasticity | $\varphi$ | Normal | 2.00 | 0.75 | 1.43 | 1.53 | 1.54 | 1.71 |
| Price indexation | $\delta$ | Beta | 0.75 | 0.15 | 0.30 | 0.33 | 0.34 | 0.40 |
| Calvo prices | $\xi$ | Beta | 0.75 | 0.15 | 0.64 | 0.70 | 0.70 | 0.76 |
| Interest rate smoothing | $\gamma$ | Beta | 0.80 | 0.10 | 0.70 | 0.71 | 0.71 | 0.72 |
| Inflation response | $\gamma_{\Pi}$ | Normal | 1.70 | 0.10 | 1.68 | 1.71 | 1.71 | 1.73 |
| Output growth response | $\gamma_{\Delta y}$ | Normal | 0.15 | 0.05 | 0.15 | 0.15 | 0.15 | 0.16 |
| Steady-state inflation | | Normal | 0.75 | 0.15 | 0.92 | 0.93 | 0.93 | 0.94 |
| Steady-state output growth | | Normal | 0.75 | 0.15 | 0.98 | 0.98 | 0.98 | 0.99 |
| Steady-state nominal interest rate | | Normal | 1.50 | 0.15 | 0.89 | 0.95 | 0.95 | 0.98 |
| Supply shock persistence | $\rho^{S}$ | Beta | 0.85 | 0.10 | 0.63 | 0.66 | 0.66 | 0.68 |
| Demand shock persistence | $\rho^{D}$ | Beta | 0.85 | 0.10 | 1.41 | 1.43 | 1.43 | 1.46 |
| Supply shock STD | $\sigma^{S}$ | Inv.gamma | 0.40 | 2.00[a] | 1.48 | 1.53 | 1.55 | 1.70 |
| Demand shock STD | $\sigma^{D}$ | Inv.gamma | 0.20 | 2.00[a] | 2.96 | 2.98 | 3.01 | 3.25 |
| Monetary shock STD | $\sigma^{M}$ | Inv.gamma | 0.10 | 2.00[a] | 0.26 | 0.27 | 0.27 | 0.29 |

a. For the inverse gamma function, degrees of freedom are indicated instead of the standard deviation.

**TABLE A2. RMSFE Comparisons of Alternative Models Using Seasonally Adjusted GDP[a]**

| | $h=1$ | | $h=4$ | |
|---|---|---|---|---|
| *Model and data used* | *RMSFE* | *RMSFE as % of AR* | *RMSFE* | *RMSFE as % of AR* |
| Alternative models | | | | |
| AR | 0.7162 | 1 | 0.6030 | 1 |
| RW | 0.7869 | 1.0987 | 0.9555 | 1.5845 |
| VAR | 0.7246 | 1.0117 | 0.6810 | 1.1294 |
| BVAR | 0.6303 | 0.8801 | 0.6187 | 1.0261 |
| SPF | 1.0043 | 1.4022 | | |
| DSGE | 0.7520 | 1.0500 | 0.5833 | 0.9673 |
| Quarterly macroeconomic data | | | | |
| Factor AR | 0.6691 | 0.9342 | 0.6253 | 1.0368 |
| Financial data | | | | |
| Flat aggregation | 0.8392 | 1.1718 | 0.6359 | 1.0545 |
| $\beta$ ($p=1, q=3$) | 0.8434 | 1.1775 | 0.8690 | 1.4409 |
| Combined MIDAS | 0.6772 | 0.9455 | 0.7317 | 1.2133 |
| Financial data + quarterly macroeconomic data | | | | |
| Flat aggregation | 0.6294 | 0.8788 | 0.5921 | 0.9818 |
| $\beta$ (p = 1, q = 3) | 0.6768 | 0.9450 | 0.7185 | 1.1915 |
| Combined MIDAS | 0.6047 | 0.8442 | 0.6096 | 1.0109 |

a. The table shows the root mean square forecast error (RMSFE) for GDP horizons of $h = 1$ and $h = 4$ for the sample 1999:1–2013:4. Estimation period: 1999:4+$h$ to 2009:4. Forecasting period: 2009:4+$h$ to 2013:4. The RMSFEs are also presented as a percentage of the AR. The five MIDAS forecasts estimated from each of the daily factors are combined to obtain the combined MIDAS. A recursive window is used for all forecasts.

## References


Adolfson, Malin, Jesper Lindé, and Mattias Villani. 2007. "Forecasting Performance of an Open Economy DSGE Model." *Econometric Reviews* 26(2–4): 289–328.

Andreou, Elena, Eric Ghysels, and Andros Kourtellos. 2013. "Should Macroeconomic Forecasters Use Daily Financial Data and How?" *Journal of Business and Economic Statistics* 31(2): 240–51.

Arnesto, Michelle T., Kristie M. Engerman, and Michael T. Owyang. 2010. "Forecasting with Mixed Frequencies." *Federal Reserve Bank of St. Louis Review* 92(6): 521–36.

Artis, Michael J., and Wenda Zhang. 1990. "BVAR Forecasts for the G-7." *International Journal of Forecasting* 6(3): 349–62.

Baffigi, Alberto, Roberto Golinelli, and Giuseppe Parigi. 2004. "Bridge Models to Forecast the Euro Area GDP." *International Journal of Forecasting* 20(3) 447–60.

Bai, Jennie, Eric Ghysels, and Jonathan H. Wright. 2013. "State Space Models and MIDAS Regressions." *Econometric Reviews* 32(7): 779–813.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin. 2010. "Large Bayesian Vector Auto Regressions." *Journal of Applied Econometrics* 25(1): 71–92.

Calvo, Guillermo A. 1983. "Staggered Prices in a Utility-Maximizing Framework." *Journal of Monetary Economics* 12(3): 383–98.

Canova, Fabio. 2007. *Methods for Applied Macroeconomic Research,* vol. 13. Princeton University Press.

Capistrán, Carlos, and Gabriel López-Moctezuma. 2010. "Las expectativas macroeconómicas de los especialistas: una evaluación de pronósticos de corto plazo en México." *El Trimestre Económico* 77(306): 275–312.

Chauvet, Marcelle, and Simon Potter. 2013. "Forecasting Output." In *Handbook of Economic Forecasting,* vol. 2, edited by Graham Elliott, Clive W. J. Granger, and Allan Timmermann, chap. 3. Amsterdam: Elsevier.

Clements, Michael P., and Ana B. Galvão. 2008. "Macroeconomic Forecasting with Mixed-Frequency Data." *Journal of Business and Economic Statistics* 26(4): 546–54.

Del Negro, M. D., and Schorfheide, F. 2013. "DSGE Model-Based Forecasting." In G. Elliot and A. Timmerman, eds., *Handbook of Economic Forecasting*, vol. 2, 57–140, Amsterdam: Elsevier.

Del Negro, Marco, and others. 2007. "On the Fit of New Keynesian Models." *Journal of Business and Economic Statistics* 25(2): 123–43.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics* 13(3): 253–265.

Doan, Thomas, Robert B. Litterman, and Christopher A. Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews* 3(1): 1–100.

Dua, Pami, and Subhash C. Ray. 1995. "A BVAR Model for the Connecticut Economy." *Journal of Forecasting* 14(3): 167–80.

Edge, Rochelle M., Michael T. Kiley, and Jean-Philippe Laforte. 2010. "A Comparison of Forecast Performance between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model." *Journal of Applied Econometrics* 25(4): 720–54.

Galvão, Ana B. 2013. "Changes in Predictive Ability with Mixed Frequency Data." *International Journal of Forecasting* 29(3): 395–410.

Ghysels, Eric. 2015. "MATLAB Toolbox for Mixed Sampling Frequency Data Analysis Using MIDAS Regression Models." University of North Carolina at Chapel Hill.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2004. "The MIDAS Touch: Mixed Data Sampling Regression Models." Working Paper 2004s-20. Montreal: CIRANO.

Ghysels, Eric, Arthur Sinko, and Rossen Valkanov. 2007. "MIDAS Regressions: Further Results and New Directions." *Econometric Reviews* 26(1): 53–90.

Giannone, Domenico, Lucrezia Reichlin, and David H. Small. 2008. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics* 55(4): 665–76.

Herrera Hernández, Jorge. 2004. "Business Cycles in Mexico and the United States: Do They Share Common Movements?" *Journal of Applied Economics* 7(2): 303–23.

Ibarra, Raúl. 2012. "Do Disaggregated CPI Data Improve the Accuracy of Inflation Forecasts?" *Economic Modelling* 29(4): 1305–13.

Ireland, Peter N. 2004. "Technology Shocks in the New Keynesian Model." *Review of Economics and Statistics* 86(4): 923–36.

LeSage, James P. 1999. "Applied Econometrics Using MATLAB." University of Toronto, Department of Economics.

Litterman, Robert B. 1986. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business and Economic Statistics* 4(1): 25–38.

Liu, Guangling D., Rangan Gupta, and Eric Schaling. 2009. "A New-Keynesian DSGE Model for Forecasting the South African Economy." *Journal of Forecasting* 28(5): 387–404.

Marcellino, Massimiliano, and Christian Schumacher. 2010. "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP." *Oxford Bulletin of Economics and Statistics* 72(4): 518–50.

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson. 2003. "Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-Wide Information." *European Economic Review* 47(1): 1–18.

Mariano, Roberto S., and Yasutomo Murasawa. 2003. "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series." *Journal of Applied Econometrics* 18(4): 427–43.

McNees, Stephen K. 1986. "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts." *Journal of Business and Economic Statistics* 4(1): 5–15.

Rubaszek, Michal, and Pawel Skrzypczyński. 2008. "On the Forecasting Performance of a Small-Scale DSGE Model." *International Journal of Forecasting* 24(3): 498–512.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48(1): 1–48.

Smets, Frank, and Rafael Wouters. 2004. "Forecasting with a Bayesian DSGE Model: An Application to the Euro Area." *JCMS: Journal of Common Market Studies* 42(4): 841–67.

Stock, James H., and Mark W. Watson. 2002. "Forecasting Using Principal Components from a Large Number of Predictors." *Journal of the American Statistical Association* 97(460): 1167–79.

———. 2008. "Forecasting in Dynamic Factor Models Subject to Structural Instability." In *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry,* eds. J. Castle and N. Shephard. Oxford University Press.

Theil, Henri, and Arthur S. Goldberger. 1961. "On Pure and Mixed Statistical Estimation in Economics." *International Economic Review* 2(1): 65–78.

Timmermann, Allan. 2006. "Forecast Combinations." In *Handbook of Economic Forecasting,* vol. 1, edited by Graham Elliott, Clive W. J. Granger, and Allan Timmermann, pp. 136–96. Amsterdam: Elsevier-North.