UPPSALA UNIVERSITY - DEPARTMENT OF STATISTICS

MIDAS

Forecasting quarterly GDP using higher-frequency data

Authors: Hanna Lindgren and Victor Nilsson

Supervisor: Lars Forsberg

January 12, 2015

We forecast US GDP sampled quarterly over horizons ranging from one quarter to three years. Using AR-MIDAS models we study three lag polynomials: the Almon lag, the exponential Almon lag and the beta lag, and nine macroeconomic variables, sampled weekly or monthly. Our benchmark model is an AR(1) and we compare forecast errors using RMSE. In all instances the AR-MIDAS achieves lower forecast errors compared to the benchmark model. The predictor sampled weekly generally performs better compared to other predictors, which are sampled monthly.

KEYWORDS: MIDAS, GDP, forecasting, mixed-frequency data

Contents

1 Introduction

Time series regression models traditionally use data where all variables are sampled at the same frequency. In situations with mixed frequencies, temporal aggregation is commonly applied to equate the variables, with the implication of information loss when higher-frequency variables are transformed to lower-frequency variables. In 2004, Ghysels, Santa-Clara and Valkanov introduced Mixed Data Sampling regressions (henceforth MIDAS) which allow the regressand and the regressors to be sampled at different frequencies. Their work focused primarily on volatility predictions, but has also proven to be useful for macroeconomic modeling. One advantage of MIDAS is that a lag polynomial is used to weight the lags of the explanatory variable. The lag polynomial only requires a few parameters to be estimated, resulting in a parsimonious model.

Since technological improvement has facilitated financial time series to be recorded at higher frequencies, research has focused on how to take advantage of the additional data. Ghysels, Santa-Clara and Valkanov (2005) examine the relationship between risk and expected returns. MIDAS-regressions allow them to use monthly returns to proxy expected returns, while using daily squared returns to estimate conditional variance. Using the mixed frequencies rather than monthly data for both variables, they obtain a better estimate of the conditional variance and conclude that there is a trade-off between risk and returns.

After MIDAS models were proven to be useful in forecasting volatility, macro economists explored whether the approach could be successfully adapted to model and forecast macroeconomic variables. Clements and Galvão (2008) include an autoregressive term into the MIDAS framework and compare the forecasts against the quarterly AR(1) and autoregressive distributed lags (ADL) forecasts. They show that AR-MIDAS result in sizeable reductions in root mean square errors for short horizon forecasts when monthly data on industrial production and capacity utilization are available in the current quarter. In another paper, Clements and Galvão (2009) conclude that AR-MIDAS outperforms the AR-model when combining leading indicators and exploiting current-quarter monthly data. Kuzin, Marcellino and Schumacher (2011) compare MIDAS with a mixed-frequency VAR model (MF-VAR) and conclude that the models work as complements since MIDAS performs better for short horizons up to four to five months, while MF-VAR performs better on horizons up to nine months.

The purpose of this study is to evaluate the forecasting performance of MIDAS models. We compare three lag polynomials: the Almon lag, the exponential Almon lag and the beta lag and use a number of macroeconomic variables that are sampled weekly or monthly. We forecast US GDP over short and long horizons and compare the results to an AR(1). Root mean square error is used to measure forecast accuracy.

The thesis is organized in the following way: section two will cover the theoretical background providing an explanation of the MIDAS and AR-MIDAS models and describe the Almon lag, the exponential Almon lag and the beta lag polynomials. Section three covers the data and the estimation method; section four presents and discusses the results. Finally, section five concludes.

2 Theoretical background

2.1 MIDAS

In the MIDAS approach we regress a dependent variable of lower frequency on a lagged independent higher-frequency variable. Using notations similar to Ghysels, Sinko and Valkanov (2007), we write a simple one regressor MIDAS model, forecasting h periods ahead as

$$y_{t+h} = \beta_0 + \beta_1 B\left(L^{1/m}; \theta\right) x_t^{(m)} + \varepsilon_{t+h}, \tag{1}$$

where y_t is the regressand and $x_t^{(m)}$ is the regressor, sampled at frequency m. If the regressand and regressor are sampled at the same frequency, then $m = 1$, whereas if e.g. the regressand is sampled quarterly while the regressor is sampled monthly, $m = 3$.

The polynomial lag operator $B(L^{1/m}; \theta)$ is working on $x_t^{(m)}$, weighting each lagged observation, where

$$B\left(L^{1/m}; \theta\right) = \sum_{k=1}^{K} B(k; \boldsymbol{\theta}) L^{k/m}, \tag{2}$$

while k is the number of lags and $L^{1/m}$ is a lag operator such that $L^{1/m} x_t^{(m)} = x_{t-1/m}^{(m)}$. For $B(k; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters to be estimated, we use three distributed lag polynomials as weighting functions, restricting the lag coefficients to lie on a polynomial function. This imposes smoothness on, yet allows for flexibility in, the shape of the lag weight distribution. As the lag polynomials only require three to four parameters to be estimated, parsimony is achieved.

The first lag polynomial we use is the Almon lag which was named after its developer Shirley Almon (1965), where the weight on each lag k is calculated as

$$B(k;\boldsymbol{\theta}) = \sum_{q=0}^{Q} \theta_q k^q, \tag{3}$$

where Q denotes the order of the polynomial. We use a third order specification, $Q = 3$, thus $\boldsymbol{\theta} = [\theta_0, \theta_1, \theta_2, \theta_3]$.

The second lag polynomial is the exponential Almon lag, inspired by Almon's work and introduced by Ghysels et al. (2007). The exponential Almon lag polynomial in its general form is given by

$$B(k,\boldsymbol{\theta}) = \frac{e^{\left(\theta_1 k^1 + \ldots + \theta_Q k^Q\right)}}{\sum_{k=1}^{m} e^{\left(\theta_1 k^1 + \ldots + \theta_Q k^Q\right)}}, \tag{4}$$

where Q denotes the order of the polynomial. In line with Ghysels et al. (2007), we use the functional form of two parameters, $\boldsymbol{\theta} = [\theta_1, \theta_2]$, and can write the exponential Almon lag as

$$B(k,\theta_1,\theta_2) = \frac{e^{\left(\theta_1 k + \theta_2 k^2\right)}}{\sum_{k=1}^{m} e^{\left(\theta_1 k + \theta_2 k^2\right)}}. \tag{5}$$

The beta lag polynomial (ibid.) has two parameters and is given by

$$f(k,\theta_1,\theta_2) = \frac{f\left(\frac{k}{K},\theta_1,\theta_2\right)}{\sum_{k=1}^{K} f\left(\frac{k}{K},\theta_1,\theta_2\right)}, \tag{6}$$

where

$$f\left(\frac{k}{K},\theta_1,\theta_2\right) = \frac{\left(\frac{k}{K}\right)^{\theta_1 - 1}\left(1 - \frac{k}{K}\right)^{\theta_2 - 1}\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}, \tag{7}$$

and

$$\Gamma(\theta_p) = \int_0^{\infty} e^{-\frac{k}{K}} \left(\frac{k}{K}\right)^{\theta_p - 1} d\,\frac{k}{K}. \tag{8}$$

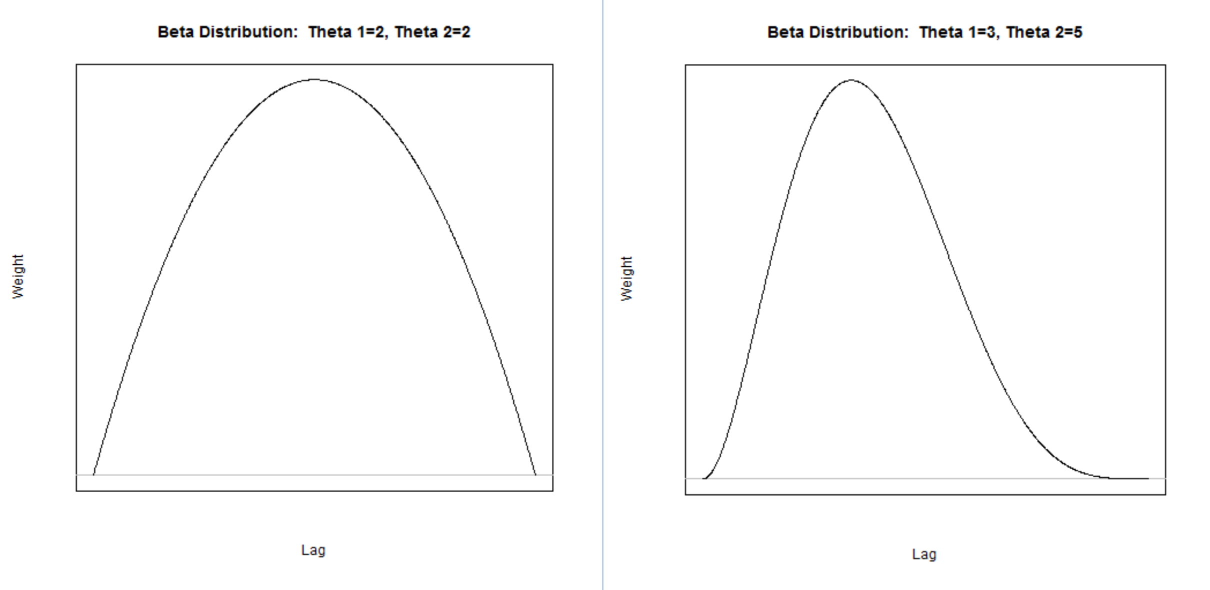
For both the exponential Almon lag and the beta lag specifications, we have four parameters to estimate in equation (1); θ_1, θ_2, the intercept β_0 and the slope coefficient β_1. The weights of these polynomials sum up to unity due to the expressions in the denominators. For the Almon lag the weights do not sum up to unity and consequently no β_1 is estimated. As we use a third order Almon polynomial, we estimate five parameters: $\theta_0,\ \theta_1,\ \theta_2,\ \theta_3$ and β_0. Another noteworthy difference between the Almon lag and the other two polynomials is that for the exponential Almon lag and the beta lag, the lag selection is purely data driven; once the functional form is specified, the rate of decline determines the number of lags included in the regression. This means that one can choose too few, but not too many, lags. For the Almon lag however, lag length selection is not data driven and therefore of greater importance. Regardless, we use 25 lags, i.e. $k = 25$, for all model specifications. All parameters are estimated using non-linear least squares.

For clearness we present the more expanded form of equation (1), taking the exponential Almon lag in equation (5) as an example:

$$y_{t+h} = \beta_0 + \beta_1 \sum_{k=0}^{K} \frac{e^{(\theta_1 k + \theta_2 k^2)}}{\sum_{k=1}^{m} e^{(\theta_1 k + \theta_2 k^2)}} L^{k/m} x_t^{(m)} + \varepsilon_{t+h}. \tag{9}$$

Figure 1 below illustrates two of the various shapes the beta weighting function can take, showing the flexibility though only two parameters are estimated.

Figure 1: Two shapes of the beta lag polynomial



The figure to the left shows the shape of the beta lag distribution when $\boldsymbol{\theta_1 = 2}$ and $\boldsymbol{\theta_2 = 2}$, and the figure to the right shows the shape of the beta lag distribution when $\boldsymbol{\theta_1 = 3}$ and $\boldsymbol{\theta_2 = 5}$.

2.2 AR-MIDAS

As we are interested in forecasting GDP we want to include autoregressive dynamics into the MIDAS model. Simply introducing an AR part into regression (1) we obtain

$$y_{t+h} = \beta_0 + \lambda y_t + \beta_1 B\left(L^{1/m};\theta\right) x_t^{(m)} + \varepsilon_{t+h}. \tag{10}$$

Ghysels et al. (2007) note that this implies that the response of the dependent variable will be "seasonal", regardless of whether the independent variable $x_t^{(m)}$ exhibits seasonal variation. Clements and Galvão (2008, 2009) solve this by adding the autoregressive dynamics as a common factor

$$y_{t+h} = \beta_0 + \lambda y_t + \beta_1 B\left(L^{1/m};\theta\right)\left(1 - \lambda L^h\right) x_t^{(m)} + \varepsilon_{t+h}. \tag{11}$$

Also in this model all parameters, including λ, are estimated using non-linear least squares. As we are comparing an AR(1)-model with an AR(1)-MIDAS, we can interpret the results as whether adding a high-frequency regressor improves the forecasting performance.

3 Data and estimation

3.1 Data

In the following section our dataset is described. All data is gathered from FRED, the Federal Reserve Economic Data program administered by the Federal Reserve Bank of St. Louis. Table 1 below lists all data series used, the frequency at which they are sampled and a short description of what each series measures. Note that civilian labor force is not used directly to forecast GDP, only to adjust other explanatory variables.

Table 1: Data description

Data series	Frequency	Description
Civilian Labor Force	Monthly	Number of people in the labor force. Individuals under the age of 16 are excluded.
Core CPI	Monthly	The Consumer Price Index excluding food and energy prices.
Real GDP	Quarterly	Real Gross Domestic Product. The inflation-adjusted value of all goods and services produced in the U.S.
Initial Claims	Weekly	The number of claims by individuals seeking state unemployment benefits.
ISM Manufacturing: PMI Composite Index©	Monthly	An index measuring whether the manufacturing sector is expanding or declining. Values range from 0 to 100, where 50 is neutral, lower values indicate a contraction and higher values indicate an expanding sector.
Total nonfarm payroll	Monthly	The number of workers in the U.S. economy. The following are excluded: proprietors, private household employees, unpaid volunteers, farm employees, and the unincorporated self-employed. Covers approx. 80 % of workers contributing to GDP.
Unemployment	Monthly	The share of labor force participants who are unemployed.
For more detailed descriptions of the data, we refer the reader to FRED found at http://research.stlouisfed.org/fred2/		

Following Armesto, Engemann, and Owyang (2010), we use log-difference transformation for the GDP series and multiply the obtained value by 100. Each observation can then be interpreted as the percentage change in GDP from the previous quarter. We are also left with a stationary series while the original GDP series has a unit root. Full results from the unit root tests are presented in the appendix. The same transformation method is also used for the other explanatory variables except the ISM Manufacturing Index which is used both log-differenced and untransformed.

In times of low economic activity some individuals might choose to study, or stop looking for employment. This means leaving the labor force and not being counted as unemployed. The unemployment rate therefore tends to underestimate the number of people not working during bad economic conditions. The opposite occurs at times of high economic activity. People who previously were not part of the labor force begin looking for work, become part of the labor force thus adding to the number of unemployed. These reasons motivate the use of total non-farm payroll, as an alternative to the unemployment rate. As the series measures the number of people employed we avoid the inconsistencies of the unemployment measure.

We create an adjusted initial claims series, dividing the original series by the labor force size. This leaves us with a series measuring the share of the labor force losing employment in a given week. We use both the adjusted and non-adjusted initial claims series to forecast GDP. As the initial claims series is the only one sampled weekly we are interested in whether any difference in forecasting performance is due to the higher frequency or due to initial claims being a better predictor. Therefore we create an initial claims series of monthly frequency by calculating the monthly average. The time aggregation is done for both the labor force adjusted and non-adjusted initial claims series.

3.2 Estimation

The AR-MIDAS model estimation and forecasting is done using Matlab R2014a together with the Matlab toolbox for MIDAS regressions written by Hang Qian and made available on Ghysels' homepage[1]. Our dataset consists of observations from January 1st 1967, to July 1st 2014. We use the period from April 1st 1975 up to January 1st 2010, as in-sample window and estimate the model. The remainder of the dataset is used for forecasting, using the rolling window method. This means that after each prediction we move the estimation window forward one step, while preserving the length of the estimation window. The model is re-estimated and a new forecast is made, the estimation window is moved one step forward and the process is repeated[2].

[1] We use version 1.1 of the toolbox, available through http://www.unc.edu/~eghysels/

[2] As the model is reestimated before every forecast, the amount of estimated parameters becomes large; therefore we do not present all the results. However, to illustrate the forecasting process, we plot the theta estimates of a one horizon forecast using weekly initial claims adjusted for labor force, using the beta lag polynomial. The plots are presented in Figure A1 in the appendix.

The choice of estimation window length requires a trade-off. We are trying to approximate the data generating process, henceforth DGP, behind GDP. A longer window means more observations available, leading to a more precise estimate. However, we might expect the DGP to change over time. A longer estimation window increases the risk of approximating multiple DGPs with a single model. Assuming no additional change in the DGP, the out of sample window used for forecasting comes from the most recent DGP. If the model used for these forecasts is estimated over multiple DGPs, forecasting performance is likely to suffer.

To compare the forecasting performance of different model specifications we use the root mean square error measure, henceforth RMSE. Each forecast made is compared to the realized value, and the squared difference is saved. As is shown in equation (12) below; the squared errors are summed, we calculate the mean and take the root to obtain the RMSE:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}. \tag{12}$$

Using lag polynomials to determine the lag weights in the AR-MIDAS means that the number of parameters estimated do not differ as we alter the number of lags included or the explanatory variables used. Recalling section 2, the exponential Almon and beta lag polynomials require estimating four parameters, whereas the Almon lag polynomial requires five parameters. As the number of parameters estimated are equal, or almost equal, we can compare the forecast performance of different explanatory variables and lag polynomials.

The forecasting results obtained from the AR-MIDAS models are compared to an autoregressive model with one lag. The AR(1)-model is, like the AR-MIDAS models, estimated using Matlab R2014a. We use the same estimation window and forecasting method as for AR-MIDAS.

4 Results

The forecast results are presented in Table 2 and Table 3 below. We divide the RMSE obtained from the AR-MIDAS forecasts by the RMSE obtained from the AR(1) forecasts and present the ratio. A value under one implies a lower forecast error for AR-MIDAS compared to the AR(1). Table 2 shows the forecasting results over one and two-quarter horizons, while Table 3 shows the results over one, two and three-year horizons[3].

The main finding is that all values are below one, thus over all horizons the AR-MIDAS performs better than the benchmark model in our sample. Looking at the shorter horizon forecasts in Table 2, the lowest forecast error over one-quarter horizon is obtained using unemployment as predictor and the exponential Almon polynomial. Over two quarters the best forecast is achieved by adjusted initial claims together with the Almon polynomial. Overall, for one and two quarters, initial claims performs better compared to other predictors. In the majority of cases, initial claims sampled weekly performs better than the monthly average, which suggests that the informational advantage of the higher frequency improves forecasting performance.

The beta and exponential Almon polynomials perform quite consistently, whereas the performance of the Almon polynomial seems to be somewhat more erratic (see e.g. the difference in RMSE between the lag polynomials for the predictor unemployment). This issue should probably be expected. As described in the theory section, the Almon polynomial is sensitive to lag length selection. It is most likely that the optimal number of lags varies between different predictors and horizons. However, we use the same number of lags for all estimations and therefore it is not surprising that the performance of the Almon polynomial varies.

[3] RMSE for all models and specifications are presented in Table A3-A5 in the appendix.

Table 2: 1 and 2-quarter GDP forecasts						
Horizon / **Predictor**	1 quarter			2 quarters		
	Almon	Beta	Exp Almon	Almon	Beta	Exp Almon
Weekly initial claims	0.6767	0.6900	**0.6538**	**0.6871**	0.6972	0.6991
Weekly initial claims adj.	**0.6727**	0.6861	0.6749	**0.6174**	0.6324	0.6355
Monthly initial claims	0.8026	0.7134	**0.6845**	0.7893	**0.6871**	0.6892
Monthly initial claims adj.	0.8034	0.7067	**0.6799**	0.7949	**0.6858**	0.6959
Monthly CPI	**0.7010**	0.7295	0.7384	**0.6717**	0.7129	0.7071
Monthly ISM log diff	0.7758	0.8015	**0.6871**	0.7280	0.7109	**0.7010**
Monthly ISM level	0.7750	0.7827	**0.7072**	0.7469	0.7777	**0.6892**
Monthly payroll	0.7908	0.6970	**0.6963**	0.7699	0.6789	**0.6685**
Monthly unemployment	0.8425	0.6765	**0.6515**	0.8271	**0.6913**	0.7023

The table shows $\boldsymbol{RMSE_{AR-MIDAS}/RMSE_{AR(1)}}$ for the different horizons. All model specifications use 25 lags. All values under one implies that AR-MIDAS results are better than AR(1). For each predictor and horizon, the result of the best performing lag polynomial is emphasized. The results for initial claims are presented both for the labor force adjusted and unadjusted predictor. Monthly initial claims is the monthly average of weekly initial claims.

In Table 3 below we see that inflation is the best predictor over both one and two-year horizons, followed by initial claims. Also over longer horizons forecast errors are lower for initial claims when sampled weekly. Once again the beta and exponential Almon polynomial performances are comparable while the Almon polynomial performance is less consistent. For the three-year forecasts, adjusted initial claims sampled weekly gives the best result, regardless of the weighting polynomial. However the Almon achieves the best. Inflation is still performing well, being the second best predictor. Compared to forecasts over shorter horizons, the three year performance of the Almon polynomial is more consistent. It is still the case though that both the highest and lowest forecast errors are attained using the Almon polynomial. This suggests that the beta lag and, in particular, the exponential Almon lag specifications give more adequate forecast errors and that lag specification is an issue for all the forecast horizons covered in this study.

Payroll and unemployment perform quite similarly. Our fears that unemployment would suffer as a predictor due to the reasons described earlier were either misguided, or our proposed solution was inadequate. The relatively good performance of initial claims raises the question of whether this is due to good general predicting performance or good performance during our forecast window. The out of sample period used for forecasting extends over a period when the US economy recovers from the worst financial crisis since the depression of the 1930's. We would caution against applying findings from this period to more normal times.

Table 3: 1, 2 and 3-year GDP forecasts

Predictor / **Horizon**	1 year			2 years			3 years		
	Almon	Beta	Exp Almon	Almon	Beta	Exp Almon	Almon	Beta	Exp Almon
Weekly initial claims	0.7322	0.6621	**0.6585**	**0.6728**	0.6753	0.6887	0.7355	0.7024	**0.6774**
Weekly initial claims adj.	0.7287	0.6598	**0.6571**	**0.6714**	0.6740	0.6880	**0.6457**	0.6578	0.6567
Monthly initial claims	0.7771	0.6708	**0.6599**	0.7579	0.7694	**0.6835**	0.7469	**0.7228**	0.7260
Monthly initial claims adj.	0.7845	0.6687	**0.6571**	0.7649	0.7534	**0.6829**	0.7532	0.7292	**0.7235**
Monthly CPI	**0.6521**	0.6846	0.6944	0.6789	**0.6507**	0.6755	**0.6529**	0.6600	0.6752
Monthly ISM log diff	0.7101	**0.6963**	0.6995	0.6943	**0.6659**	0.6825	**0.6966**	0.7050	0.7149
Monthly ISM level	0.7293	0.7339	**0.6730**	0.7272	0.7314	**0.7178**	0.7233	**0.7164**	0.7362
Monthly payroll	0.7589	0.7584	**0.7008**	0.7588	0.7850	**0.6975**	**0.7626**	0.7721	0.7720
Monthly unemployment	0.7991	0.7681	**0.6710**	0.7926	0.7522	**0.7011**	0.7974	0.7577	**0.7258**

The table shows $\boldsymbol{RMSE_{AR-MIDAS}/RMSE_{AR(1)}}$ for the different horizons. All model specifications use 25 lags. All values under one implies that AR-MIDAS results are better than AR(1). For each predictor and horizon, the result of the best performing lag polynomial is emphasized. The results for initial claims are presented both for the labor force adjusted and unadjusted predictor. Monthly initial claims is the monthly average of weekly initial claims.

5 Conclusion

This thesis compares the forecasting performance of AR-MIDAS, using an AR(1) as benchmark model. For the AR-MIDAS models we use nine predictors, three lag polynomials and forecast quarterly US GDP over five horizons. Across all horizons, all AR-MIDAS specifications yield lower forecast errors, compared to the benchmark model. These findings are in line with previous research. The beta and exponential Almon polynomials perform similarly in terms of forecasting performance. The exponential Almon polynomial achieves slightly lower forecast errors compared to the beta polynomial in a majority of cases, but no test is conducted to test whether the differences are statistically significant. The performance of the Almon polynomial exhibits larger variation. This is most likely due to Almon lag performance being sensitive to the number of lags included, and that the optimal number of lags varying between predictors and/or horizons.

Weekly sampled initial claims, adjusted or not, is consistently performing well from short to long horizons. This indicates that there is an informational advantage in using higher-frequency data over both short and longer horizons. Unemployment is one of the better predictors over short horizons but its relative performance drops when forecasts are made over longer horizons. The opposite applies for inflation.

The consistent performance of the AR-MIDAS, together with previous findings, leads us to believe that the addition of higher-frequency predictors aids forecasting performance. For further research, we suggest examining whether the use of multiple higher-frequency explanatory variables could further improve forecasting performance.

6 References

Almon, S. (1965), "The Distributed Lag Between Capital Appropriations and Expenditures", Econometrica 33: 178–196.

Armesto, M., Engemann K. & Owyang, M., (2010), ”Forecasting with mixed frequencies”, Review, Federal Reserve Bank of St. Louis, issue Nov, pages 521-536.

CBOE White Paper, available at http://www.cboe.com/micro/vix/vixwhite.pdf, accessed November 18, 2014.

Clements M. and Galvão A. (2008), "Macroeconomic forecasting with mixed frequency data: forecasting output growth in the United States", Journal of Business and Economic Statistics 26: 546–554

Clements, M. and Galvão, A. (2009), "Forecasting US Output Growth Using Leading Indicators: An Appraisal Using MIDAS Models", Journal of Applied Econometrics 24:7 1187–1206

FRED, Federal Reserve Bank of St. Louis https://research.stlouisfed.org/fred2/, January 10, 2015.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2004), "The MIDAS Touch: Mixed Data Sampling Regression Models”, CIRANO Working Papers 2004s-20, CIRANO

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005), ”There is a risk-return trade-off after all”. Journal of Financial Economics 76, 509–548. doi:10.1016/j.jfineco.2004.03.008

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006), “Predicting volatility: getting the most out of return data sampled at different frequencies”, Journal of Econometrics, 131, 59–95.

Ghysels, E., Sinko, A. and Valkanov, R. (2007), ”MIDAS Regressions: Further Results and New Directions”. Econometric Reviews 26, 53–90. doi:10.1080/07474930600972467

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

MIDAS regression, Matlab toolbox v.1.1, available at http://www.unc.edu/~eghysels/, accessed on November 12, 2014.

7 Appendix

7.1 Unit root tests

Table A1: Raw GDP series	
pValue	0.9990
Stat	11.6372
Critical Value	-1.9423
N	191
Coefficient	1.0061
Standard Error	5.2228e-04
Cov	2.7277e-07
μ	9.5860e+03
σ	3.7050e+03
DWStat	1.2624
SSR	2.6095e+09
SSE	1.0214e+06
SST	2.6105e+09
MSE	5.4040e+03
RMSE	73.5118
RSq	0.9996
aRSq	0.9996
LL	1.0856e+03
AIC	2.1732e+03
BIC	2.1765e+03
HQC	2.1745e+03
The null hypothesis of unit root presence is not rejected. The GDP series is unit root.	

Table A2: Log-diffed GDP series	
pValue	0.001
Stat	-6.8165
Critical Value	-1.9423
N	190
Coefficient	0.6016
Standard Error	0.0584
Cov	0.0034
μ	0.6971
σ	0.8242
DWStat	2.3543
SSR	61.1930
SSE	140.4125
SST	201.6055
MSE	0.7469
RMSE	0.8642
RSq	0.3035
aRSq	0.3035
LL	-240.0988
AIC	482.1977
BIC	485.4394
HQC	483.5110
The null hypothesis of unit root presence is rejected. The series is stationary.	

7.2 RMSE for AR-MIDAS and AR(1)

Table A3: RMSE for 1 and 2-quarter GDP forecasts

Horizon / **Predictor**	1 quarter			2 quarters		
	Almon	Beta	Exp Almon	Almon	Beta	Exp Almon
Weekly initial claims adj.	0.5329	0.5435	0.5346	0.5011	0.5133	0.5158
Weekly initial claims	0.5360	0.5466	0.5179	0.5577	0.5659	0.5674
Monthly CPI	0.5553	0.5779	0.5849	0.5452	0.5786	0.5739
Monthly initial claims	0.6358	0.5651	0.5422	0.6406	0.5577	0.5594
Monthly initial claims adj.	0.6364	0.5598	0.5386	0.6452	0.5566	0.5648
Monthly ISM % change	0.6145	0.6349	0.5443	0.5909	0.5770	0.5690
Monthly ISM level	0.6139	0.6200	0.5602	0.6062	0.6312	0.5594
Monthly payroll	0.6264	0.5521	0.5516	0.6249	0.5510	0.5426
Monthly unemployment	0.6674	0.5359	0.5161	0.6713	0.5611	0.5700
AR(1)	0.7921			0.8117		

The table shows RMSEs for AR-MIDAS and AR(1) for the different horizons. All model specifications use 25 lags. The results for initial claims are presented both for the labor force adjusted and unadjusted predictor. Monthly initial claims is the monthly average of weekly initial claims.

Table A4: RMSE for 1 and 2-year GDP forecasts

Horizon / **Predictor**	1 year			2 years		
	Almon	Beta	Exp Almon	Almon	Beta	Exp Almon
Weekly initial claims adj.	0.6048	0.5476	0.5453	0.5622	0.5644	0.5761
Weekly initial claims	0.6077	0.5495	0.5465	0.5634	0.5655	0.5767
Monthly CPI	0.5412	0.5682	0.5763	0.5685	0.5449	0.5657
Monthly initial claims	0.6449	0.5567	0.5477	0.6347	0.6443	0.5724
Monthly initial claims adj.	0.6511	0.5550	0.5453	0.6405	0.6309	0.5719
Monthly ISM % change	0.5893	0.5779	0.5805	0.5814	0.5576	0.5715
Monthly ISM level	0.6053	0.6091	0.5585	0.609	0.6125	0.6011
Monthly payroll	0.6298	0.6294	0.5816	0.6354	0.6574	0.5841
Monthly unemployment	0.6632	0.6375	0.5569	0.6637	0.6299	0.5871
AR(1)	0.8299			0.8374		

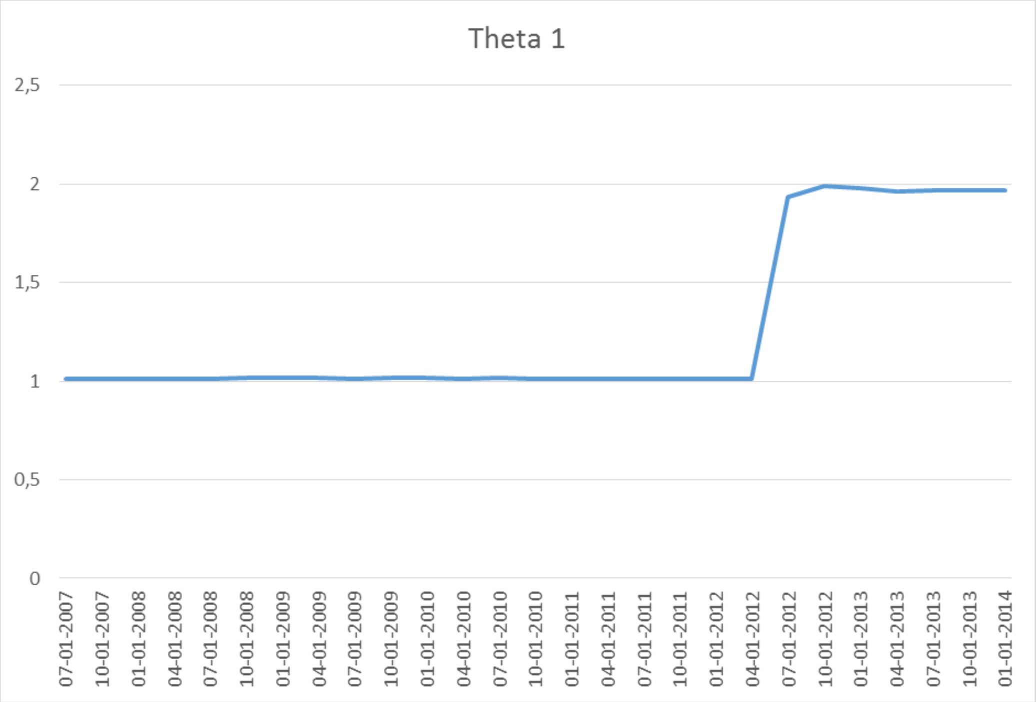
The table shows RMSEs for AR-MIDAS and AR(1) for the different horizons. All model specifications use 25 lags. The results for initial claims are presented both for the labor force adjusted and unadjusted predictor. Monthly initial claims is the monthly average of weekly initial claims.

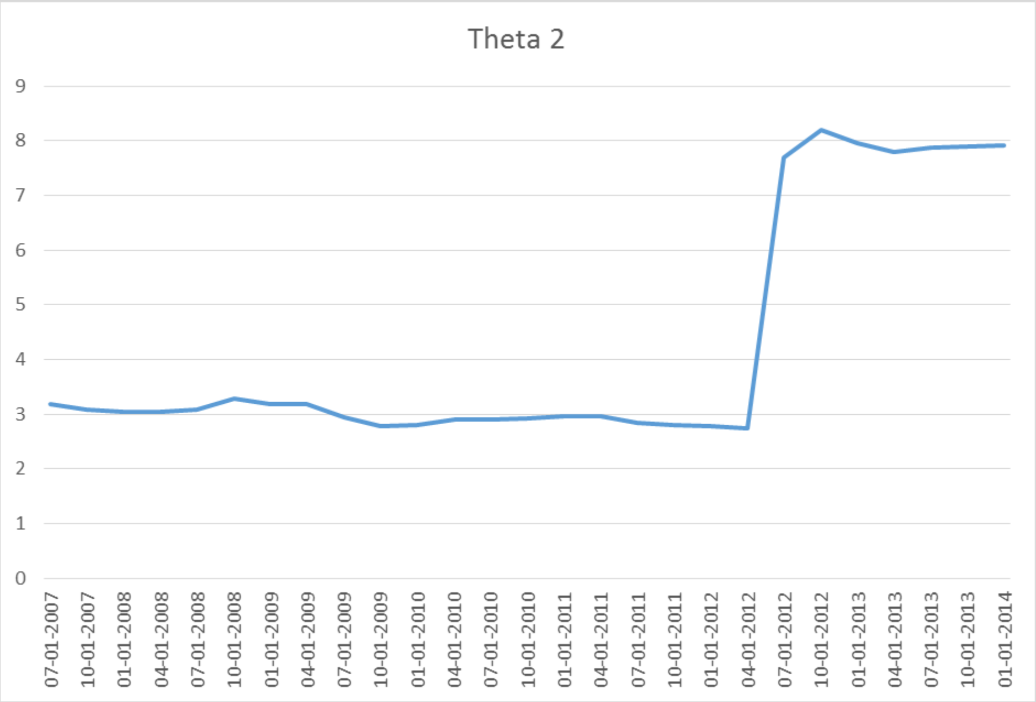
Table A5: RMSE for 3 year GDP forecasts			
Horizon / **Predictor**	3 years		
	Almon	Beta	Exp Almon
Weekly initial claims adj.	0.5440	0.5542	0.5533
Weekly initial claims	0.6197	0.5918	0.5707
Monthly CPI	0.5501	0.5561	0.5689
Monthly initial claims	0.6293	0.6090	0.6117
Monthly initial claims adj.	0.6346	0.6144	0.6096
Monthly ISM % change	0.5869	0.5940	0.6023
Monthly ISM level	0.6094	0.6036	0.6203
Monthly payroll	0.6425	0.6505	0.6504
Monthly unemployment	0.6718	0.6384	0.6115
AR(1)	0.8425		

The table shows RMSEs for AR-MIDAS and AR(1) for the different horizons. All model specification use 25 lags. The results for initial claims are presented both for the labor force adjusted and unadjusted predictor. Monthly initial claims is the monthly average of weekly initial claims.

7.3 Lag polynomial estimates

Figure A1: Beta lag polynomial estimates for weekly initial claims, adjusted for labor force





The figures above plot the beta lag polynomial estimates $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ when forecasting GDP over one horizon using weekly initial claims, adjusted for labor force. The figures show how changes in the relationship between the dependent and independent variables are picked up and incorporated in the later forecasts when using the rolling window method.