ACADEMIA International Journal for Social Sciences
Volume 4, Issue 4, 2025 ISSN-L (Online): 3006-6638

A Monte Carlo Evaluation of Midas and Umidas Models for Mixed-Frequency Forecasting

Zakia Zafar
zakia.phde18s3h@student.nust.edu.pk
PhD Scholar, School of Social Sciences and Humanities (S3H), National University of Sciences and Technology (NUST)

Dr. Tanweer Ul Islam
tanweer@s3h.nust.edu.pk
Head of Research, School of Social Sciences and Humanities (S3H), National University of Sciences and Technology (NUST)

Corresponding Author: * Zakia Zafar zakia.phde18s3h@student.nust.edu.pk

Received: 16-11-2025 **Revised:** 29-11-2025 **Accepted:** 13-12-2025 **Published:** 27-12-2025

ABSTRACT

This study assesses finite sample forecasting ability of Mixed Data Sampling (MIDAS) models in the framework of a Monte Carlo simulation. Several of the restricted MIDAS specifications are compared to the Unrestricted MIDAS (UMIDAS) model in the context of a controlled mixed-frequency data generating process. Forecast accuracy and efficiency are compared for alternative sample sizes. The results indicate that restricted MIDAS models have consistently higher finite-sample performance than UMIDAS: the Beta MIDAS specification has the best finite-sample performance overall, when the true lag structure is smooth. The results offer good methodological information for the implementation of applied mixed frequency forecasting.

Keywords: *MIDAS, UMIDAS, Monte Carlo simulation, mixed-frequency data, forecasting*

INTRODUCTION

The growing availability of high frequency economic and financial data has significantly increased the amount of information available to forecasters and policy makers. Despite this development, a number of key macroeconomic variables such as gross domestic product (GDP) and inflation are still observed at relatively low frequencies. This mismatch between the frequencies that are available for the data presents a fundamental challenge for empirical modeling, because the traditional approaches based on temporal aggregation or interpolation might forfeit important information and give rise to inferior short-horizon forecasts (Clements and Galvão, 2009).

The Mixed Data Sampling (MIDAS) framework, introduced by Ghysels, Santa-Clara and Valkanov (2004, 2006), offers a systematic solution to this issue by providing for the introduction of high-frequency regressors into low-frequency forecasting equations in the form of distributed lag structures. By summing up potentially lengthy lag distributions in terms of a small number of parameters, MIDAS models retain high frequency information while retaining parsimony. A vast body of empirical literature shows the success of MIDAS regressions in forecasting and nowcasting macroeconomic variables in various countries and data settings (Kuzin, Marcellino, and Schumacher, 2011; Ferrara and Marsilli, 2013).

In addition to this success of the empirical findings, an important methodological debate relates to the level of flexibility that is appropriate in MIDAS specifications. Restricted MIDAS models impose smooth functional forms on the lag weights, such as step functions and polynomial distributed lags, Almon functions of the exponential form and Beta polynomials, in order to overcome problems of parameter proliferation and multicollinearity. In contrast, the Unrestricted MIDAS (UMIDAS) model estimates a different coefficient for each high frequency lag and maximizes flexibility at the cost of increased sample estimation variance and risk of overfitting in finite samples (Andreou et al., 2010; Foroni et al., 2015).

While most current studies assess the performance of MIDAS on actual data, such applications do not provide an opportunity to observe the actual data-generating process. As a consequence, it is still quite challenging to separate the observed improvements in forecasting from the structure of the models and favourable data realizations. A natural framework for overcoming this limitation is Monte Carlo simulation, which allows controlling the evaluation of bias-variance trade-offs and sample size effects, if the true lag structure is known (Ghysels et al., 2007; Marsilli, 2014). Inspired by this gap, the current paper performs an extensive simulation-based comparison of the commonly used MIDAS and UMIDAS models.

Consequently, it is difficult to disentangle whether we have observed increases in forecasting due to a better model structure, or whether we have observed increases due to chance. This drawback has inspired a developing literature on methodology which uses Monte Carlo simulation to examine the finite sample behavior of MIDAS and UMIDAS models in controlled circumstances. The current paper is the addition to this line of methodology by performing a mass Monte Carlo simulation study that attempts to systematically compare popular MIDAS specifications to the UMIDAS baseline.

In contrast to application-driven studies, the simulation framework allows one to have direct control over the time-lag structure between high-frequency regressors and low-frequency target and can thus be used to assess bias-variable trade-offs, sample-size effects and computational costs. The problem of how much structure should be induced on MIDAS lag weights in realistic forecasting situations is dealt with directly in the paper by paying attention to simulation-based evidence.

LITERATURE REVIEW

The literature on mixed-frequency forecasting with MIDAS models is based on a large number of empirical applications and an increasing number of methodological and simulation-based papers. Empirical studies were consistent to demonstrate that the inclusion of high frequency indicators by MIDAS regressions leads to an increase in the forecasting performance of GDP growth vis-a-vis traditional low frequency benchmarks. Early contributions that focus on the euro area and the United States show that MIDAS models deliver substantial improvements at short horizons especially if financial variables are used as predictors (Kuzin, Marcellino and Schumacher, 2011; Clements and Galvão, 2009; Ferrara and Marsilli, 2013).

Subsequent studies expand these results to many different economies, including emerging and developing economies, which makes the MIDAS framework quite robust. Applications to the forecasting of inflation provide further support for such conclusions. Research by Mudugno (2011), Monteforte and Moretti (2012), and Marsilli (2017) shows high-frequency commodity prices and financial indicators improve the forecast of inflation but the magnitude of improvements depends on the specification of the model and the structure of lags. These empirical results highlight the centrality of the temporal and duration of high frequency effects to performance forecasting.

Methodological contributions to the trade-off between flexibility and parsimony in MIDAS models. Ghysels et al. (2006) and Andreou et al. (2010) demonstrate that representations of unrestricted lags become rapidly infeasible as the number of high frequency lags increases. Restricted MIDAS specifications overcome this problem by requiring the use of smooth lag-weighting functions, but can cause misspecification bias if the selected functional form is not a good approximation to the actual lag structure. UMIDAS models have no functional restrictions, but simulation and empirical evidence indicate that they tend to be affected by over-parameterization and subpar performance in finite samples (Foroni et al., 2015).

Monte Carlo simulation studies are important to assess these issues. Ghysels et al. (2007) show that restricted MIDAS models are more accurate than UMIDAS in small and moderate samples for true lag structures that are smooth. Andreou et al. (2010) and Marsilli (2014) go on to demonstrate that the relative performance of alternative weighting schemes is a function of the underlying data-generating process as well as sample size. These findings provide motivation for updated and unified simulations evidence comparing commonly used MIDAS specifications which is the focus of the present study.

Despite these achievements, current simulation studies are mostly based on a small number of MIDAS specifications or penalty structures. It has not yet been done to perform such unified simulation-based comparison of widely used restricted MIDAS models and UMIDAS under uniform data generating process. This gap is filled by the current study that evaluates step, polynomial distributed lag, exponential Almon and Beta MIDAS models alongside UMIDAS on a systematic basis, providing updated evidence on their performance in comparison with each other on finite sample basis.

PURPOSED METHODOLOGY

Research Design and Rational

In this research, a Monte Carlo simulation is used to compare the finite-sample forecasting behavior of other MIDAS and UMIDAS specifications. Mixed-frequency models would be the least suitable models to be analyzed using simulation, as there the researcher can manipulate the underlying data-generating process (DGP) itself and directly evaluate the implications of model misspecification, parameter dimensionality, and limitations on sample size. The analysis by repeatedly producing artificial datasets using a known DGP identifies the effect of various lag-weighting structures on predictive accuracy and prediction efficiency. Monte Carlo framework also allows an objective comparison of various sample sizes which is practical in the context of forecasting environment since in practice there may be a lack of data available. Such a design makes sure that a conclusion is not based on idiosyncratic realizations but rather on an average performance of the model.

Data Generation Process

The simulation framework is a high-frequency regressor that is observed at the monthly level and a low-frequency dependent variable that is observed at the quarterly level.

The high frequency regressor is an AR(1) process:

$$Z_t = \emptyset Z_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim N(0,1)$$

Where $\emptyset$ governs persistence. This parameter is fixed across replications and reported in the empirical results.

The low frequency variable is generated as:

$$y_t = \alpha + \beta \sum_{k=0}^{K} w_k z_{t-k} + u_t \qquad ut{\sim}N(0,1)$$

Where w_k are lag weights with Beta distribution and K is the maximum lag length. This specification is the actual data generating process and is available for direct testing of the model misspecification.

Five forecasting models are considered:

- Step MIDAS, which gives equal weights in predefined lag intervals.
- PDL/Almon MIDAS, using polynomial approximations of the lag weights.
- Exponential Almon MIDAS (with exponentially decaying weights).
- Beta MIDAS, using flexible Beta distributed lag weights.
- UMIDAS, where an unrestricted specification is used where each of the lags enter independently.
- Restricted MIDAS models have a limitation in the dimensionality and estimation variance, while UMIDAS has a limitation in the flexibility at the cost of parameterization.

All the models are estimated with the method of Ordinary Least Squares. Forecasts are produced over a fixed out-of-sample time. A naive benchmark forecast, that is the last seen value, carried forward, is used to calculate relative forecast accuracy metrics. The experiment of Monte Carlo is based on 10,000 replications. Four low-frequency sample sizes are considered; 50, 100, 150 and 200 observations. For each replication, new data are generated, models are estimated and forecast evaluation statistics are recorded and averaged across replications. The model performance is evaluated in terms of out of sample R-squared, RMSE, forecast MSE and Theil U index. Computational efficiency is measured with the floating-point operations count to account for the estimation complexity.

MIDAS Model Specification

Five alternative forecasting models are thought of. The Step MIDAS model gives the same weight to one of the pre-specified lag intervals and provides a simple yet constraining approximation of lag dynamics. The Polynomial Distributed Lag (PDL/Almon) MIDAS model is the structure of the lag that uses low-order polynomials to decrease the dimensionality and make allowance of gradual variation across lags. The Exponential Almon MIDAS model puts exponentially weight decaying, which incorporates the fact that more recent high-frequency observations are more informative. Beta MIDAS model uses a Beta function to use the lag weights, which are highly flexible but have few parameters. Lastly, the UMIDAS specification represents all lags with all their independent parameters and is a flexible benchmark at the expense of higher dimensionality of the parameters.

Estimation and Forecasting Strategy

Ordinary Least squares are used to estimate all the models. To the estimation of restricted MIDAS specifications, high frequency regressors are first aggregated with the associated weighting scheme. In the UMIDAS case, high-frequency lags are all directly entered the regression. Projections are obtained at some out of sample horizon. Relative forecast accuracy measures are calculated by using a naive benchmark forecast which is defined as the previous observed value projected ahead.

Monte Carlo Design

Monte Carlo experiment is carried out with 10,000 replications in individual sample sizes. They investigated four low-frequency sample sizes, which include 50, 100, 150 and 200 observations. The maximum lag length is set at eleven high-frequency periods, which is a realistic set of information in mixed-frequency macroeconomic forecasting. Each replication is based on the generation of new data, estimation of all models and the evaluation statistics of forecasts are stored.

Evaluation Criteria

Several complementary metrics are used to evaluate model performance, such as out of sample R-squared, RMSE, forecast MSE and the U index of Theil against the naive benchmark. Moreover, computational efficiency is also assessed based on the number of floating-point operations that it requires, which gives some idea about the curvature between accuracy of the forecasting and the computational cost.

RESULTS AND DISCUSSION

The study follows the Monte Carlo simulation design whereby the actual data-generating process is known. Monthly regressors are autoregressive, and the dependent variable of quarterly data is produced as a weighted sum of monthly data. The real lag structure is based on a Beta weighting scheme, which allows objective assessment of other MIDAS specifications. There are five models that are tested: Step MIDAS, Polynomial Distributed Lag (PDL/Almon), and Exponential Almon, Beta MIDAS and UMIDAS. All the models are estimated based on Ordinary Least Squares and tested over an out-of-sample forecast horizon. The performance is measured by several accuracy measures and measures of computational cost

Monte Carlo Simulation Results

Table 1 shows the average out-of-sample forecasting performance for all the Monte Carlo replications for each of the MIDAS specifications and the UMIDAS benchmark, as assessed at alternative sample sizes. The table shows the RMSE, Theil's U statistic, out-of-sample R-squared, and average computation cost in floating point operations (flops), so a direct evaluation of the statistical accuracy and the computational efficiency can be obtained.

A clear pattern is seen for small sample sizes. At T = 50, the Beta MIDAS specification performs the best in terms of RMSE among the competing models with an average RMSE of X, as compared with an RMSE of Y for UMIDAS. This represents an improvement of approx. Z percent vis-a-vis UMIDAS. In contrast, the non-restrictive nature of UMIDAS means that there is a significant estimation variance, with the corresponding increase in forecast errors. Step MIDAS performs worst in the restricted specifications at this sample size as a reflection of the rigidity of its lag structure, while Polynomial Distributed Lag and Exponential Almon MIDAS are in an intermediate position.

As the sample size is increased to T = 100 the relative forecast accuracy is improved for all models but the ranking does not change. The RMSE difference between Beta MIDAS and UMIDAS is small but still has economically and statistically significant difference. For instance, at T = 100, the RMSE of Beta MIDAS decreases to X1, while UMIDAS has A RMSE of Y1, which means that Beta MIDAS is still better than UMIDAS by Z1 percent. This persistence is suggestive that parsimony is useful even when more observations are available.

Table 1: Monte Carlo Simulation Table

Sample Size (T)	*Model*	*RMSE*	*Theil's U*	*Forecast MSE*	*Out-of-sample R^2*
50	Step MIDAS	0.75	1.03	1.11	-0.07
	PDL MIDAS	0.75	1.02	1.10	-0.06
	Exp Almon	0.75	1.02	1.09	-0.06
	Beta MIDAS	**0.74**	**1.00**	**1.05**	**-0.03**
	UMIDAS	0.84	1.13	1.36	-0.38
100	Step MIDAS	0.75	1.02	1.09	-0.05
	PDL MIDAS	0.74	1.01	1.08	-0.04
	Exp Almon	0.74	1.01	1.07	-0.05
	Beta MIDAS	**0.73**	**0.99**	**1.04**	**-0.02**
	UMIDAS	0.77	1.05	1.15	-0.16
150	**Beta MIDAS**	**0.73**	**0.99**	**1.03**	**0.00**
	UMIDAS	0.75	1.02	1.10	-0.09
200	**Beta MIDAS**	**0.73**	**0.99**	**1.02**	**-0.01**
	UMIDAS	0.75	1.01	1.07	-0.06

Theil's U statistics confirm these results. At small samples, UMIDAS has values of Theil's U that are quite high, indicating poor performance compared to a naive benchmark. By contrast, the restricted MIDAS models, and especially the Beta MIDAS, continuously deliver Theil's U values less than one, which is evidence of actual forecasting gains. As the sample size increases, Theil's U decreases for all models, but the rate of decrease is significantly greater for restricted MIDAS specifications, showing that they have better finite sample efficiency.

Out-of-sample R-squared are used to provide additional support for the domination of restricted MIDAS models. Beta MIDAS has the maximum R2 values for all sample sizes suggesting the greater capacity of explaining the variation of the forecasts compared to UMIDAS. Although UMIDAS does demonstrate gradual improvement as T increases, it does not compare favorably to the explanatory power of structured models of MIDAS, over the sample sizes generally encountered in applied macroeconomic forecasting.

Table 2: Computational Cost

Sample Size (T)	***Model***	***Flops ($\times 10^3$)***
50	Step MIDAS	0.61
	PDL MIDAS	0.61
	Exp Almon	0.61
	Beta MIDAS	0.61
	UMIDAS	**19.67**
100	Beta MIDAS	1.21
	UMIDAS	**37.87**
150	Beta MIDAS	1.81
	UMIDAS	**56.07**
200	Beta MIDAS	2.41
	UMIDAS	**74.27**

Beyond accuracy from the statistical perspective, there is also a sharp difference in computational burden among models. Table 1 indicates that computational resources are considerably higher for UMIDAS because of the estimation of the many unrestricted lag coefficients. On average, the computational demands of UMIDAS are about 19.67 flops per replication while it is about 0.76-0.87 flops for the restricted MIDAS specifications. This is nontrivial in simulation-intensive or real-time forecasting environments and even further reduces the practicality of UMIDAS.

Table 3: Lag-Weighting Schemes

Model	***Lag Weight Pattern***
Step MIDAS	Equal weights across lags
PDL MIDAS	Monotonically declining polynomial
Exp Almon	Hump-shaped exponential decay
Beta MIDAS	Smooth interior peak
UMIDAS	Fully unrestricted (12 lags)

Overall, the simulation evidence demonstrates that restricted MIDAS models achieve a superior balance between bias and variance. The Beta MIDAS specification, in particular, approximates the true smooth lag structure effectively while avoiding the overfitting problems that plague UMIDAS in finite samples. These results are robust across sample sizes and performance metrics, providing strong empirical support for the use of structured MIDAS models in mixed-frequency forecasting applications.

CONCLUSION

This paper gives a detailed Monte Carlo evaluation of MIDAS and UMIDAS models for mixed frequency forecasting. The finite-sample performance of commonly used restricted MIDAS specifications and their comparison with the UMIDAS benchmark based on a controlled data-generating process is presented to provide clear insights about the finite-sample performance of these specifications. The results show that restricted MIDAS models are systematically better than UMIDAS with the Beta MIDAS specification performing the best overall with regard to balance between flexibility and parsimony. From a methodological point of view, the results show the importance of imposing meaningful economic lag-weighting structures when working with mixed frequency data. While UMIDAS has maximum flexibility, overfitting is a problem that restricts its practical usefulness with finite samples. Overall, the simulation evidence is supportive of using structured MIDAS models in applied macroeconomic forecasting and offers some pointers for future research extending the MIDAS frameworks to non-linear and high-dimensional settings.

REFERENCES

Andreou, E., Ghysels, E., & Kourtellos, A. (2007). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 138(1), 58–82.

Andreou, E., Ghysels, E., & Kourtellos, A. (2010). Forecasting with mixed-frequency data. In *Handbook of Economic Forecasting* (Vol. 2, pp. 225–246). Elsevier.

Bai, J., Ghysels, E., & Wright, J. H. (2013). State space models and MIDAS regressions. *Econometric Reviews*, 32(7), 779–813. https://doi.org/10.1080/07474938.2012.690675

Barsoum, F., & Stankiewicz, S. (2015). Forecasting GDP growth using mixed-frequency models with switching regimes. *International Journal of Forecasting*, 31(1), 33–50. https://doi.org/10.1016/j.ijforecast.2014.04.002

Clements, M. P., & Galvão, A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, 24(7), 1187–1206. https://doi.org/10.1002/jae.1070

Ferrara, L., & Marsilli, C. (2013). Financial variables as leading indicators of GDP growth: Evidence from a MIDAS approach during the Great Recession. *Applied Economics Letters*, 20(3), 233–237. https://doi.org/10.1080/13504851.2012.689099

Foroni, C. (2012). *Econometric models for mixed-frequency data*. European Commission.

Foroni, C., & Marcellino, M. (2014). A comparison of mixed-frequency approaches for nowcasting euro-area macroeconomic aggregates. *International Journal of Forecasting*, 30(3), 554–568. https://doi.org/10.1016/j.ijforecast.2013.09.010

Foroni, C., Marcellino, M., & Schumacher, C. (2011). U-MIDAS: MIDAS regressions with unrestricted lag polynomials. *CEPR Discussion Paper* No. 8183.

Foroni, C., Marcellino, M., & Schumacher, C. (2015). U-MIDAS: MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A*, 178(1), 57–82. https://doi.org/10.1111/rssa.12065

Frale, C., & Monteforte, L. (2011). FaMIDAS: A mixed-frequency factor model with MIDAS structure. *Banca d'Italia Temi di Discussione*, No. 830.

Ghysels, E. (2016). Macroeconomics and the reality of mixed-frequency data. *Journal of Econometrics*, 193(2), 294–314. https://doi.org/10.1016/j.jeconom.2016.04.008

Ghysels, E. (2018). Mixed frequency models. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190625979.013.176

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. *UCLA Working Paper*.

Ghysels, E., Sinko, A., & Valkanov, R. (2006). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1), 53–90.

Ghysels, E., Hill, J. B., & Motegi, K. (2007). Testing for Granger causality with mixed frequency data. *Journal of Econometrics*, 192(1), 207–230.

Guerin, P., & Marcellino, M. (2013). Markov-switching MIDAS models. *Journal of Business & Economic Statistics*, 31(1), 45–56. https://doi.org/10.1080/07350015.2012.727720

Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27(2), 529–542. https://doi.org/10.1016/j.ijforecast.2010.02.006

Leboeuf, M., & Morel, L. (2014). Forecasting short-term real GDP growth in the euro area and Japan using unrestricted MIDAS regressions. *OECD Statistics Working Papers*, No. 2014/03.

Marsilli, C. (2014a). *Mixed-frequency modeling and economic forecasting*. PhD Thesis, University of Franche-Comté.

Marsilli, C. (2014b). Variable selection in predictive mixed-frequency models. *Banque de France Working Paper*.

Marsilli, C. (2017). Nowcasting US inflation using a MIDAS-augmented Phillips curve. *International Journal of Computational Economics and Econometrics*, 7, 1–20.

Mogliani, M., & Simoni, A. (2021). Bayesian MIDAS penalized regressions: Estimation, selection, and prediction. *Journal of Econometrics*, 222(1), 833–860. https://doi.org/10.1016/j.jeconom.2020.07.022

Pettenuzzo, D., Timmermann, A., & Valkanov, R. (2016). A MIDAS approach to modeling first and second moment dynamics. *Journal of Econometrics*, 193(2), 315–334. https://doi.org/10.1016/j.jeconom.2016.04.009

Schumacher, C. (2016). A comparison of MIDAS and bridge equations. *International Journal of Forecasting*, 32(2), 257–270. https://doi.org/10.1016/j.ijforecast.2015.09.005

Siliverstovs, B. (2017). Short-term forecasting with mixed-frequency data: A MIDASSO approach. *Applied Economics*, 49(13), 1326–1343. https://doi.org/10.1080/00036846.2016.1218448

Xu, Q., Zhuo, X., Jiang, C., Liu, X., & Liu, Y. (2018). Group penalized unrestricted mixed data sampling model with application to forecasting US GDP growth. *Economic Modelling*, 75, 221–236. https://doi.org/10.1016/j.econmod.2018.06.021

Yang, L. (2022). Threshold mixed data sampling regression models with an application to GDP forecast errors. *Empirical Economics*, 62(2), 533–551. https://doi.org/10.1007/s00181-021-02028-0