LCERPA

Laurier Centre for Economic Research & Policy Analysis

LCERPA Working paper No. 2025-6

June 2025

Exact Mixed-Frequency Data Sampling (eMIDAS)

Quinlan Lee
University of Toronto

Stephen Snudden
Wilfrid Laurier University

Exact Mixed-Frequency Data Sampling (eMIDAS)

Quinlan Lee
University of Toronto

Stephen Snudden*†
Wilfrid Laurier University

June 25, 2025

Abstract

We derive the exact bottom-up functional form of mixed-frequency data sampling (eMIDAS) models for temporally aggregated VARMA processes. The eMIDAS specification depends solely on the structure of the disaggregated data-generating process and is invariant to the degree of time aggregation. This reduces the number of estimated parameters by a factor equal to the aggregation frequency. Compared to existing unrestricted and restricted MIDAS approaches, eMIDAS simplifies implementation and yields more efficient and robust estimation. We show that high-frequency information on both explanatory and response variables is required for efficiency and quantify the value of measuring data at high frequency. Simulation results demonstrate that eMIDAS is the only MIDAS-type method that rivals recursive bottom-up methods and often outperforms them in real-time empirical applications.

JEL classification: C32, C43, C52.
Keywords: Temporal Aggregation, Mixed Frequency, MIDAS, Time Series, Inflation, Interest Rates.

*Corresponding Author: Department of Economics, 64 University Ave W, Waterloo, Ontario, N2L 3C7, Canada; ORCID iD 0000-0002-4033-3235; Email: ssnudden@wlu.ca

†Many thanks to Dalibor Stevanovic, Heiner Mikosch, Martin Burda, Reinhard Ellwanger, and Christian Gourieroux, as well as participants at the XIX New York Camp Econometrics, the 59th Annual Meetings of the Canadian Economics Association, and seminars at Wilfrid Laurier University and the University of Toronto for their discussions, comments, and suggestions. This research was supported by the Social Sciences and Humanities Research Council (SSHRC) Insight Grant 435-2025-1423.

1 Introduction

This study advances our understanding of how to construct direct mixed-frequency forecasts of temporally aggregated data, such as monthly averages of exchange rates or quarterly sums of output. Such aggregations are a primary focus in economics, as they are understood to best reflect economic conditions over a period of time, including total costs and revenue. However, forecasts constructed with temporally aggregated data can theoretically lead to a substantial reduction in forecast accuracy (Tiao, 1972; Amemiya and Wu, 1972; Kohn, 1982; Lütkepohl, 1986). One class of models to rectify this information loss is the MIxed DAta Sampling (MIDAS) framework. For instance, Restricted MIDAS (RMIDAS) techniques (Ghysels et al., 2007; Andreou et al., 2010, 2013) synthesize data with varying frequencies into a single econometric model. The more recent literature on Unrestricted MIDAS (UMIDAS) (Foroni and Marcellino, 2016; Foroni et al., 2015, 2019) assumes the existence of an underlying high-frequency data generating process, which is translated into a low-frequency model. While these methods are an effective means of mitigating some of the bias arising from temporal aggregation, they feature an immense parameterization structure which depends on the degree of aggregation in the data. When the discrepancy in frequencies is large, such as aggregation from daily to monthly or quarterly, traditional MIDAS forecasts can suffer from a substantial decline in efficiency.

We revisit the UMIDAS setting under the assumption that information on the high-frequency data generating process is available to the practitioner. We then derive, under this context, the optimal direct bottom-up forecast. The bottom-up approach is a novel way to approach the MIDAS structure, but is aligned with the literature on the recursive bottom-up forecasts of (V)ARIMA processes (Zellner and Montmarquette, 1971; Tiao, 1972; Amemiya and Wu, 1972; Kohn, 1982; Lütkepohl, 1986). We show that eMIDAS depends solely on the Markov structure of the underlying high-frequency process, as opposed to the immense parameterization structure induced by aggregation polynomials. This facilitates a sizable reduction in dimensionality by a factor equal to the degree of aggregation. With rapid growth in access to high-frequency data sets, eMIDAS is becoming increasingly feasible in many applied contexts and is easily implemented in practice.[1] Moreover, the gains in efficiency are significant, and showcase the importance and value of high-frequency information.

Simulations are used to evaluate the efficiency of alternative methods of integrating high-

[1]For example, daily CPI (Cavallo and Rigobon, 2016) and effective exchange rates (McCarthy and Snudden, 2024), as well as macroeconomic data from financial markets such as interest rates and commodity prices.

frequency information in both univariate and multivariate settings: (i) recursive "bottom-up" forecasts, (ii) MIDAS with restricted parameter profiles (RMIDAS), (iii) MIDAS with unrestricted parameter profiles (UMIDAS), and (iv) the proposed eMIDAS. We find that eMIDAS is more efficient than RMIDAS and UMIDAS and is the only method with gains commensurate with efficient disaggregated recursive bottom-up approaches. The gains in forecast accuracy are substantive and can reduce forecast error by over 40 percent compared to the use of monthly or quarterly data. In-sample testing of parameterization corroborates the simpler eMIDAS structure.

To demonstrate the practical relevance of our approach, we evaluate the real-time forecast accuracy of alternative methods using monthly U.S. interest rates and headline inflation. The eMIDAS forecasts consistently outperform both UMIDAS and RMIDAS specifications, and importantly also improve upon recursive bottom-up forecasts. These results indicate that efficiently constructed MIDAS models are not only simpler to implement, but also yield more accurate forecasts than previously recognized. For inflation in particular, the forecast improvements over UMIDAS and RMIDAS are concentrated in the tails of the distribution, where accuracy is especially critical for policy and risk management. Our results corroborate evidence that gains from high-frequency information enhance tail- and downside-risk forecasts (e.g., Ferrara et al., 2022; Chan et al., 2025).

The key contribution of this study is to extend the mixed-frequency literature by reinterpreting MIDAS as the direct forecast analog of a recursive bottom-up approach. We are the first to derive a closed-form, finite-dimensional exact MIDAS representation in which the high-frequency realization of the dependent variable is observed. This stands in contrast to the use of weighting functions (Brewer, 1973) in the existing MIDAS literature (Marcellino, 1999; Ghysels et al., 2004; Foroni et al., 2015). There are two immediate advantages. First, the bottom-up approach guarantees efficiency (Lütkepohl, 1986). Second, while weighting functions offer a convenient way to express the information structure, they do not always exhibit finite representations (see, Brewer, 1973; Silvestrini and Veredas, 2008).[2] The eMIDAS form therefore supplies a structural benchmark against which any finite-dimensional approximation can be evaluated.

Our second contribution is a systematic comparison of bottom-up forecasts—both recursive and eMIDAS—with conventional UMIDAS and RMIDAS estimators. Martins and Teles (2025) show analytically that temporal aggregation can bias RMIDAS/UMIDAS coefficients and inflate MSFE when the high-frequency lag weights are treated as free parameters. We demonstrate that this bias

[2]For example, an AR(2) aggregated to two periods has UMIDAS structure with infinite high-frequency moving average terms.

vanishes when the high-frequency path of the target is available (or credibly proxied) because the exact bottom-up structure fully determines the required lags. Accordingly, when the high-frequency target is observed, eMIDAS and recursive bottom-up forecasts attain the same efficiency, whereas models that do not use such information are sub-optimal.

Despite the uniqueness of the eMIDAS functional form, our theoretical conclusions are consistent with recent empirical findings, and help explain why these patterns occur. For example, Kronenberg et al. (2023) show that parsimoniously parameterized UMIDAS models without low-frequency moving average terms consistently outperform alternative mixed-frequency approaches in real-time GDP forecasting. Similarly, Benmoussa et al. (2025) find that restricted MIDAS models for daily oil prices assign all weight to the last daily observation. Bayesian, frequentist, and machine learning shrinkage methods have been shown to improve forecast accuracy by recovering sparse lag structures (Mogliani and Simoni, 2021; Babii et al., 2022; Kohns and Potjagailo, 2025; Chan et al., 2025). For instance, Daniele et al. (2025) demonstrate that applying shrinkage in observation-driven mixed-frequency VARs reduces overparameterization and enhances performance. Our analysis provides a theoretical rationale for the smaller parameterizations often found in practice.

While eMIDAS offers a new approach to mixed-frequency modelling, it complements rather than replaces existing MIDAS methods. A key insight is that forecast efficiency hinges on access to disaggregated high-frequency information. Conversely, our VAR simulations show that when the high-frequency data of the dependent (target) variable is unobserved—so neither the bottom-up approach nor eMIDAS can be applied—efficient estimation is unattainable and UMIDAS/RMIDAS provide the best performance. Section 4 further demonstrates that comparable gains can still be realized—and the eMIDAS structure retained—when a strong high-frequency proxy is available; in its absence, UMIDAS and RMIDAS remain the more reliable choices. More broadly, eMIDAS underscores the value of measuring disaggregated data and of constructing robust proxies when direct observation is not feasible.

The analysis also contributes to the ongoing debate over direct versus recursive forecasting approaches (see, e.g., Ing, 2003; Marcellino et al., 2006; Chevillon, 2007). Existing studies have compared direct and recursive forecast approaches within a single frequency, suggesting that recursive forecasts often outperform direct forecasts (see, e.g., Ing, 2003; Marcellino et al., 2006; Chevillon, 2007). Simulation evidence reveals that direct bottom-up forecasts often outperform recursive bottom-up forecasts. In practice, the efficient eMIDAS forecasts offer several practical

advantages, particularly at longer horizons.

Finally, the findings contribute to the understanding of information loss in forecasts arising from temporal aggregation. The forecast gains, for both the eMIDAS and bottom-up approaches, are commensurate with those suggested in theory (Tiao, 1972; Kohn, 1982; Lütkepohl, 1986), and substantially larger than those documented in prior empirical investigations, which focused on monthly-to-quarterly or quarterly-to-annual aggregation (see, e.g., Zellner and Montmarquette, 1971; Lütkepohl, 1986; Athanasopoulos et al., 2011). Utilizing disaggregated data in mixed-frequency forecasts can be efficient under the eMIDAS framework, albeit with the stronger data requirement that the disaggregated components of the low-frequency observation are observed.

Overall, these insights reveal that direct forecasts from mixed-frequency techniques can be simpler to implement and more efficient than previously understood. The eMIDAS approach offers key guidance for achieving the substantial information gains available through temporal disaggregation.

2 Exact MIDAS

We start by deriving the optimal MIDAS forecast under the assumption that the high-frequency data-generating process is observable to the practitioner. In doing so, we demonstrate that the methods in the existing literature can deviate from this structure quite substantially, especially when the degree of aggregation is large. Our approach is consistent with the assumptions in (Tiao, 1972; Amemiya and Wu, 1972; Kohn, 1982; Lütkepohl, 1986; Marcellino, 1999), where the recursive bottom-up forecast is mean-squared forecast error (MSFE) optimal (Kohn, 1982; Lütkepohl, 1986). We begin by presenting two motivating examples which showcase the key differences between the exact MIDAS structure and the structure implied by traditional MIDAS approaches. Then, we generalize our results to the (V)ARMA setting.

2.1 Motivating Examples

2.1.1 AR(1) Process

Our first exposition will be through the lens of a high-frequency daily AR(1) model, given by:

$$(1-\rho L)y_{t,i} \quad = \quad \varepsilon_{t,i} \quad \text{for } i=1,2,...,n;\ t=1,2,...,T, \tag{1}$$

where $\varepsilon_{t,i}$ denotes the daily structural innovation for the i-th day within month t (with a total of n days) and transition $y_{t,n+1} = y_{t+1,1}$. We define its temporal aggregate in month t as $\bar{y}_t = w(L)y_{t,i}$, where $w(L) = 1 + L + ... + L^{n-1}$ denotes the aggregation operator for an arbitrary n. Here, L is the one–day lag operator, i.e. $L\, y_{t,i} = y_{t,i-1}$; hence $L^n y_{t,i} = y_{t-1,i}$, so L^n corresponds to a one-month lag when the month contains n daily observations. This is the univariate analog of the original UMIDAS setting (Foroni et al., 2015), where it was assumed that the practitioner had information on $\bar{y}_t$ (or some other temporal aggregate), but not $y_{t,i}$.

Following (Brewer, 1973), let us introduce a polynomial $\beta(L)$ such that $h(L) = \beta(L)(1 - \rho L)$ contains only the powers of L^n. The role of such polynomial is to transform the high-frequency model in 1 into low frequency. It can be shown that the polynomial $\beta(L)$ is equal to $(1 + \rho L + ... + \rho^{n-1}L^{n-1})$ in this context. By multiplying both sides of equation 1 by $\omega(L)$ and $\beta(L)$ we obtain the UMIDAS structure:

$$h(L)\bar{y}_t = \left[1 + (1+\rho)L + ... + (1+\rho^{n-1})L^{2(n-1)}\right]\varepsilon_{t,i}. \tag{2}$$

Thus, $\bar{y}_t$ admits an ARMA type representation at low frequency, with autoregressive component characterized by $h(L)$, and a high-frequency moving average component characterized by $\beta(L)$ such that the highest lag at high frequency is equal to $2(n-1)$. For illustration, we plot in Figure 1 the values of these moving average coefficients for a month with $n = 21$ business days. From this figure we deduce that the weights on the high frequency moving averages are nonlinear when the underlying process is stationary. Moreover, we see that weights tend to be much higher for more recent values.

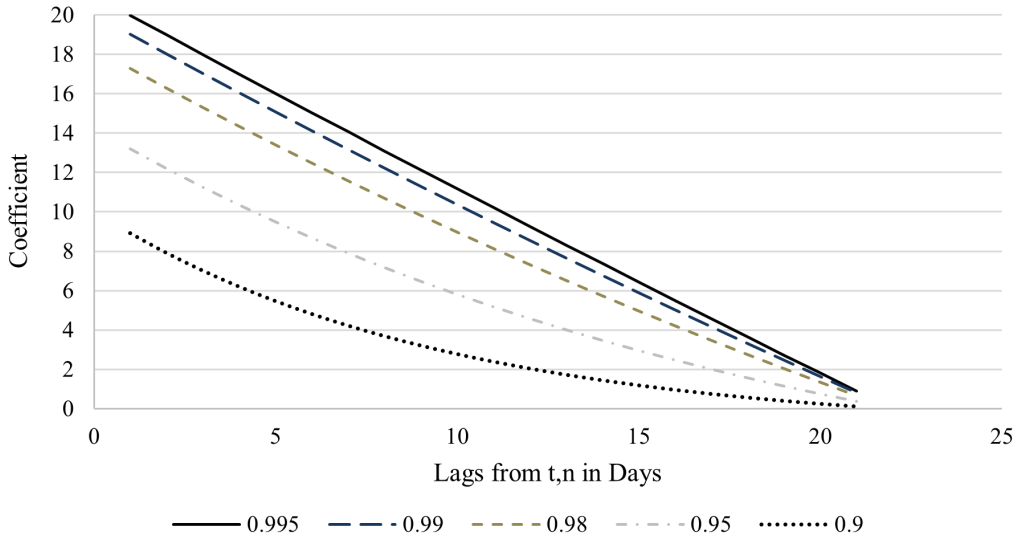
At monthly horizon 1, the forecast implied by the UMIDAS model is given by:

$$\mathbb{E}_T\left[\bar{y}_{T+1}\right] = c_1 \bar{y}_T, \tag{3}$$

where the coefficient c_1 arises from the autoregressive structure implied by polynomial $h(L)$. This equation demonstrates that the UMIDAS forecast is a projection on the average of the most recent $n-1$ historical daily values, and can be implemented as a direct forecast in this setting.

To understand the uncertainty implied by the UMIDAS forecast, let us now derive its variance at horizon 1 conditional on the information set at month T.:

Figure 1. Coefficients on Lagged Daily Values are Not Constant



Notes: Coefficients on lagged daily values for alternative values of ρ of an AR(1) model with temporal aggregation to monthly data, $n = 21$.

Proposition 1. *For UMIDAS model in 2, the variance of the forecast at horizon 1 is given by:*

$$\mathbb{V}_T[\bar{y}_{T+1}] = \frac{1}{n^2} \sum_{i=0}^{2(n-1)} \gamma_i^2, \tag{4}$$

where $\gamma_i = \sum_{j=\max(0,i-n)}^{\min(n,i)} \rho^j$ *and variance is taken with respect to the information set at month* T*.*

Proof. The low-frequency process at horizon 1 is of the form:

$$\bar{y}_{T+1} = c_1 \bar{y}_T + \beta(L)\varepsilon_{T+1}, \tag{5}$$

where c_1 is the low-frequency autoregressive coefficient (corresponding to the polynomial $h(L)$) and $\varepsilon_{T+1} = \frac{1}{n}[\varepsilon_{T+1,1} + ... + \varepsilon_{T+1,n}]$. Therefore, the only source of uncertainty conditioning at month T is the future LF innovation:

$$\begin{aligned} \beta(L)\varepsilon_{T+1} &= \left[(1 + \rho L + ... + \rho^{n-1}L^{n-1})\frac{1}{n}(1 + L + ... + L^{n-1})\right]\varepsilon_{T+1,n} \\ &= \sum_{i=0}^{2(n-1)} \gamma_i \varepsilon_{T+1,n-i}, \end{aligned} \tag{6}$$

where $\gamma_i = \sum_{j=\max(0,i-n)}^{\min(n,i)} \rho^j$. Its variance is given by:

$$\begin{aligned}
\mathbb{V}_T[\beta(L)\varepsilon_{T+1}] &= \frac{1}{n^2}\mathbb{V}\left[\sum_{i=0}^{2(n-1)} \gamma_i \varepsilon_{T+1,n-i}\right], \\
&= \frac{1}{n^2}\sum_{i=0}^{2(n-1)} \mathbb{V}\left[\gamma_i \varepsilon_{T+1,n-i}\right], \\
&= \frac{1}{n^2}\sum_{i=0}^{2(n-1)} \gamma_i^2,
\end{aligned} \tag{7}$$

which is the desired result. □

Intuitively, the uncertainty of the forecast is completely explained by the low frequency moving average and comprises two sources. The first source is the uncertainty due to the future, that is, the variation arising from daily innovations beyond $\varepsilon_{T,n}$. The second source is the uncertainty due to aggregation; although we condition on the information set at month T, the daily process is unobserved and therefore add to the uncertainty of the forecast—even though some of the daily moving averages are "in the past".

In contrast, we assume perfect observability of the daily process. Our aim is to derive what is the "best" forecast in this setting, and to compare its structure to that of UMIDAS. In more technical language, we derive the MSFE-optimal forecast under these assumptions, to reveal the exact MIDAS (eMIDAS) forecast. This is given in the following proposition:

Proposition 2. *The exact mixed-frequency direct forecast of a temporally aggregated AR(1) with* $n \geq 2$ *and forecast horizon* k *is given by:*

$$\mathbb{E}_{T,n}\left[\bar{y}_{T+k}\right] = \beta_k y_{T,n},$$

where $\mathbb{E}_{T,n}[\cdot] = \mathbb{E}[\cdot|y_{T,n}]$ *denotes an expectation taken over the information set* $y_{T,n}$ *and* β_k *is a function of the daily autoregressive parameter* ρ*.*

Proof. In this setting, the daily forecast for day i in period $T+k$ is given by

$$\mathbb{E}_{T,n}\left[y_{T+k,i}\right] = \rho^{(k-1)n+i} y_{T,n} \tag{8}$$

The optimal recursive forecast is given by the bottom-up approach (Lütkepohl, 1986). As such, the period-average forecast is the ex-post average of the daily forecasts. For an arbitrary monthly

horizon k we get:

$$\mathbb{E}_{T,n}\left[\bar{y}_{T+k}\right] = \frac{1}{n}\sum_{i=1}^{n} \rho^{(k-1)n+i} y_{T,n}.$$

This implies that for arbitrary k, the coefficient β_k is just the sum of terms:

$$\beta_k = \frac{1}{n}\sum_{i=1}^{n} \rho^{(k-1)n+i},$$

which gives us the desired result. □

This yields a powerful insight: the optimal forecast can be obtained by projecting only on the most recent daily value $y_{T,n}$. Hence, it aligns with the Markov structure of the underlying daily AR(1) process, rather than an ARMA structure as seen in equation 2 for UMIDAS. Moreover, our result demonstrates that the optimal structure of MIDAS forecasts does not depend on observations other than the most recent daily value, as opposed to UMIDAS that employs $n-1$ historical daily observations. This allows for a reduction in dimensionality in the number of parameters to estimate, for example, from $n-1$ to one parameter, as well as eliminating the need for approximations using restrictive profiles. For a practitioner, this offers an efficient and simple way of forecasting temporal aggregates, and provides compelling evidence on the importance of high-frequency information.

In the context of efficiency, the variance of the eMIDAS forecast is given in the following proposition:

Proposition 3. *For eMIDAS in 2, the variance of the forecast at horizon 1 is given by:*

$$\mathbb{V}_{T,n}\left[\bar{y}_{t+1}\right] = \frac{1}{n^2}\sum_{i=0}^{n-1} \gamma_i^2 \tag{9}$$

where $\gamma_i = \sum_{j=\max(0,i-n)}^{\min(n,i)} \rho^j$ *and variance is taken with respect to the information set at day* T, n.

Proof. At each daily horizon, we have:

$$y_{t+1,i} = \rho^i y_{T,n} + \sum_{j=1}^{i} \rho^{i-j} \varepsilon_{t+1,i}. \tag{10}$$

When the values are summed together and taken at average we get:

$$\bar{y}_{t+k} = \sum_{i=1}^{n} \rho^i y_{T,n} + \sum_{i=1}^{n}\sum_{j=1}^{i} \frac{1}{n} \rho^{(i-j)} \varepsilon_{t+1,i}. \tag{11}$$

The only source of uncertainty conditioning at daily T, n is the sequence of future HF innovations. It follows directly that:

$$\sum_{i=1}^{n}\sum_{j=1}^{i}\frac{1}{n}\rho^{(i-j)}\varepsilon_{T+1,i} = \sum_{i=0}^{n-1}\frac{1}{n}\gamma_i\varepsilon_{T+1,n-i}, \tag{12}$$

where $\gamma_i = \sum_{j=\max(0,i-n)}^{\min(n,i)}\rho^j$. Thus, the variance becomes:

$$\mathbb{V}_{T,n}\left[\sum_{i=0}^{n-1}\frac{1}{n}\gamma_i\varepsilon_{T+1,n-i}\right] = \frac{1}{n^2}\sum_{i=0}^{n-1}\gamma_i^2, \tag{13}$$

which is the result as desired. □

The expression in equation 9 is very similar to the variance of UMIDAS in 4. The key difference is that the summation is taken up to $n-1$ lags, instead of $2(n-1)$ in the UMIDAS framework. Intuitively, this means that eMIDAS does not take into account uncertainty from past daily observations, since the daily process is observable to the practitioner. This reveals precisely the efficiency gains from high-frequency information and can be summarized by the following corollary:

Corollary 1. *Let $\mathbb{V}_T^{UMIDAS}[\bar{y}_{T+1}]$ denote forecast variance implied by UMIDAS at horizon 1 and let $\mathbb{V}_{T,n}^{UMIDAS}[\bar{y}_{T+1}]$ be the forecast variance implied by eMIDAS at horizon 1. Then:*

$$\mathbb{V}_T^{UMIDAS}[\bar{y}_{T+1}] \geq \mathbb{V}_{T,n}^{UMIDAS}[\bar{y}_{T+1}], \tag{14}$$

with equality as $n \to 1$.

Therefore, only when we do not aggregate (i.e. $n = 1$) is eMIDAS and UMIDAS equivalent. Conversely, as n increases, the efficiency gap widens.

2.1.2 External Predictor

Next, let us consider an example where a process $\bar{y}_t$ is observed at monthly frequency with underlying daily data-generating process given by:

$$y_{t,i} \quad = \quad \gamma x_{t,i-1} + \nu_{t,i} \quad \text{for } i = 1, 2, ..., n;\ t = 1, 2, ..., T. \tag{15}$$

In the context of RMIDAS (Ghysels et al., 2007), the aggregate $\bar{y}_t$ can synthesized into a single model with daily $x_{t,i-1}$ using a MIDAS regression of the form:

$$\bar{y}_t = B(L^{n-1})x_{t,i} + \varepsilon_{t,i}, \tag{16}$$

where $B(L^{n-1}) = \sum_{i=1}^{i_{max}} B(i)L^{n-i}$ is a polynomial with length at most i_{max}. The polynomial $B(L^{n-1})$ can be parameterized by either exponential Almon or Beta lags. This again induces a parameterization structure which depends on the degree of aggregation, $n-1$.

However, if we are willing to assume that x_t follows a daily AR(1) process of the form:

$$(1-\rho L)x_{t,i} \quad = \quad \varepsilon_{t,i} \quad \text{ for } i = 1, 2, ..., n;\ t = 1, 2, ..., T, \tag{17}$$

then we can apply the principles of eMIDAS in this setting as well.

Proposition 4. *The exact mixed-frequency direct forecast of equation 17 with $n \geq 2$ and forecast horizon k is given by*

$$\mathbb{E}_{T,n}\left[\bar{y}_{T+k}\right] = \beta_k x_{T,n}$$

Proof. In this setting, the daily forecast for $x_{T+k,i}$ is given by

$$\mathbb{E}_{T,n}\left[x_{T+k,i}\right] = \rho^{(k-1)n+i} x_{T,n} \tag{18}$$

Given $\mathbb{E}_{T,n}\left[x_{T+k,i}\right]$ for arbitrary k, i the daily forecast is given by

$$\mathbb{E}_{T,n}\left[y_{T+k,i+1}\right] = \gamma\rho^{(k-1)n+i} x_{T,n}$$

Then the forecast for $\bar{y}_{T+k}$ is given by

$$\mathbb{E}_{T,n}\left[\bar{y}_{T+k}\right] = \frac{1}{n}\sum_{i=1}^{n} \gamma\rho^{(k-1)n+i} x_{T,n}.$$

This implies that for arbitrary k, the coefficient β_k is just the sum of terms:

$$\beta_k = \frac{1}{n}\sum_{i=1}^{n} \gamma\rho^{(k-1)n+i}.$$

□

In this context, the exact MIDAS structure depends only on the Markov structure of the underlying external predictor. It does not require the additional lags induced by the Almon or Beta lag polynomials, as in the RMIDAS specification. Moreover, there is no need for the practitioner to have any high-frequency information on y_t; by imposing the additional assumption that x_t follows an AR structure, the implementation of eMIDAS is simple and requires only a projection of the aggregate $\bar{y}_{t+k}$ on the most recent daily value $x_{T,n}$.

2.2 A General Result for Exact MIDAS

We now present the general result for Exact MIDAS under a general high-frequency VARMA framework. Consider a daily data-generating process given by:

$$a(L)Y_{t,i} = b(L)\varepsilon_{t,i}, \quad \text{for } i = 1, ..., n;\ t = 1, ..., T, \tag{19}$$

where L is a lag operator, such that $Ly_{t,i} = y_{t,i-1}$, $b(L) = (1 + \theta_1 L + \cdots + \theta_q L^q)$, and $a(L) = (1 - \rho_1 L - \cdots - \rho_p L^p)$. Here, n is the number of daily price observations within a month, and we define $y_{t,0} = y_{t-1,n}$ to transition between months.

Proposition 5. *For UMIDAS, the optimal direct forecast of a temporally aggregated VARMA(p,q) with $n \geq 2$ and forecast horizon k is given by*

$$\mathbb{E}_{T,n}\left[\bar{Y}_{T+k}\right] = \sum_{j=1}^{p} \beta_{k,j} Y_{T,n-j} + \sum_{\ell=1}^{q} \alpha_{k,\ell} \varepsilon_{T,n-\ell+1},$$

where $\beta_{k,j}$ are functions of the autoregressive matrices $A_1, ..., A_p$ and $\alpha_{k,\ell}$ are functions of the autoregressive matrices and the moving average coefficients $\Theta_1, ..., \Theta_q$.

Proof. The daily VARMA process is given by:

$$Y_{t,i} = \sum_{j=1}^{p} A_j Y_{t,i-j} + \varepsilon_{t,i} + \sum_{\ell=1}^{q} \Theta_\ell \varepsilon_{t,i-\ell} \tag{20}$$

At horizon $T + 1, 1$, the optimal forecast is:

$$\mathbb{E}_{T,n}[Y_{T+1,1}] = \sum_{j=1}^{p} A_j Y_{T,n-j+1} + \varepsilon_{T,1} + \sum_{\ell=1}^{q} \Theta_\ell \varepsilon_{T,n-\ell+1} \tag{21}$$

By forward substitution, we get at horizon $T+K, i$:

$$\mathbb{E}_{T,n}[Y_{T+k,i}] = \sum_{j=1}^{p} \Phi_j(k,i) Y_{T,n-j+1} + \sum_{\ell=1}^{q} \Gamma_\ell(k,i) \varepsilon_{T,n-\ell+1}, \tag{22}$$

where $\Phi_j(k,i)$ are the coefficients for the VAR terms and $\Gamma_\ell(k,i)$ are the coefficients for the VMA terms. In the first p forecasts, the autoregressive coefficients are generated by:

$$\Phi_j(1,i) = \begin{cases} A_{j+i-1} + \sum_{1 \le s \le i-1} A_s \Phi_j(k, i-s) & \text{for } \; j+i-1 < p \\ \sum_{1 \le s \le i-1} A_s \Phi_j(k, i-s) & \text{otherwise} \end{cases} \tag{23}$$

with initial conditions $\Phi_j(1,1) = A_j$ for $j = 1, ..., p$. In the first q forecasts, the moving average coefficients are generated by:

$$\Gamma_\ell(1,i) = \begin{cases} \Theta_{\ell+i-1} + \sum_{1 \le s \le i-1} A_s \Gamma_\ell(k, i-s) & \text{for } \; \ell+i-1 < q \\ \sum_{1 \le s \le i-1} A_s \Gamma_\ell(k, i-s) & \text{otherwise} \end{cases} \tag{24}$$

with initial conditions $\Gamma_\ell(1,1) = \Theta_\ell$ for $\ell = 1, ..., q$. Thereafter, the coefficients evolve according to the recurrence relations:

$$\begin{aligned} \Phi_j(k,i) &= A_1 \Phi_j(k, i-1) + ... + A_p \Phi_j(k, i-p), \\ \Gamma_\ell(k,i) &= A_1 \Gamma_\ell(k, i-1) + ... + A_p \Gamma_\ell(k, i-p), \end{aligned} \tag{25}$$

Thus, the autoregressive terms $\Phi_j(k,i)$ are functions of only the autoregressive parameters $A_1, ..., A_p$, but the moving average terms $\Gamma_\ell(k,i)$ are functions of both autoregressive parameters $A_1, ..., A_p$ and the moving average parameters $\Theta_1, ..., \Theta_q$. The expected trajectory of the process $Y_{T+k,i}$ conditional on T, n is then given by:

$$\mathbb{E}\left[Y_{T+k,i} | Y_{T,n}\right] = \sum_{j=1}^{p} \Phi_j(k,i) Y_{T,n-j+1} + \sum_{\ell=1}^{q} \Gamma_\ell(k,i) \varepsilon_{T,n-\ell+1}. \tag{26}$$

Then, we can get the forecast of $\bar{Y}_{T+k}$ by aggregation:

$$\mathbb{E}\left[\bar{Y}_{T+k} | Y_{T,n}\right] = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p} \Phi_j(k,i) Y_{T,n-j+1} + \frac{1}{n} \sum_{i=1}^{n} \sum_{\ell=1}^{q} \Gamma_\ell(k,i) \varepsilon_{T,n-\ell+1}. \tag{27}$$

Hence, we have $\beta_{k,j} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\Phi_j(k,i)$ and $\alpha_{k,\ell} = \frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{q}\Gamma_\ell(k,i)$ □

Therefore, Proposition 26 extends the exact MIDAS structure to the VARMA setting and shows that the compelling insight from our motivating AR(1) example is retained in this context. In particular, what matters is only the underlying Markov structure of the daily VARMA process; the most recently observed p autoregressive and q moving average values are needed for efficiency in this setting. Moreover, this result implies the exact MIDAS structure of several special cases of interest to the practitioner.

2.2.1 Moving Averages

Corollary 2. *The optimal direct forecast of a temporally aggregated (V)MA(q) with $n \geq 2$ and forecast horizon k is given by:*

$$\mathbb{E}_{T,n}\left[\bar{Y}_{T+k}\right] = \begin{cases} \sum_{\ell=1}^{q}\alpha_{k,\ell}\varepsilon_{T,n-\ell+1}, & \text{if } (k-1)n < q \\ 0, & \text{otherwise.} \end{cases}$$

This result for (V)MA processes is a direct application of Proposition 5. In particular, for the (V)ARMA(0,q) setting, since $\alpha_{k,\ell}$ is a function of the autoregressive parameters, when they are set to 0, the optimal forecast will be 0 when you go beyond q days of forecast.

The intuition of this result is straightforward: a finite order MA process carries no serial dependence beyond q daily shocks, so once the forecast horizon exceeds those q steps, the conditional mean of every remaining shock is zero. More importantly, this memoryless property is inherited from the underlying daily innovations rather than the lower-frequency moving average. This suggests that aggregation alone does not reintroduce memory.

2.2.2 Point-in-Time Sampling

Aside from temporally aggregated data, one may also be interested in point-in-time sampling, including the use of end-of-period data. MIDAS applications commonly model lower-frequency end-of-period returns (e.g., monthly) using high-frequency predictors (e.g., daily or hourly observations). We show that the principles of exact MIDAS can be readily applied in this setting.

Corollary 3. *The exact mixed-frequency direct forecast of a point-in-time sampled VARMA(p,q)*

at point v and forecast horizon k is given by

$$\mathbb{E}_{T,n}\left[Y_{T+k,v}\right] = \sum_{j=1}^{p} \beta_{k,j} Y_{T,n-j+1} + \sum_{\ell=1}^{q} \alpha_{k,\ell} \varepsilon_{T,n-\ell+1}$$

Proof. As in equation 26, the expected trajectory of the process $Y_{T+k,i}$ conditional on T, n is given by:

$$\mathbb{E}\left[Y_{T+k,i}|Y_{T,n}\right] = \sum_{j=1}^{p} \Phi_j(k,i) Y_{T,n-j+1} + \sum_{\ell=1}^{q} \Gamma_\ell(k,i) \varepsilon_{T,n-\ell+1}. \tag{28}$$

Then, we can get the forecast of $Y_{T+k,v}$ by sampling the forecast at only point v:

$$\mathbb{E}\left[Y_{T+k,v}|Y_{T,n}\right] = \sum_{j=1}^{p} \Phi_j(k,v) Y_{T,n-j+1} + \sum_{\ell=1}^{q} \Gamma_\ell(k,v) \varepsilon_{T,n-\ell+1}. \tag{29}$$

Hence, we have $\beta_{k,j} = \Phi_j(k,v)$ and $\alpha_{k,\ell} = \Gamma_\ell(k,v)$ as required. □

The implication is that regardless if modeling temporally aggregated or point-in-time sampled VARMA(p,q) processes, the structure depends entirely on the degree of p and q, but not the time sampling n.

3 Quantifying Predictability

We now compare the efficiency of eMIDAS to alternative mixed-frequency techniques, including RMIDAS, UMIDAS, and a VARMA model estimated on aggregated data. We also examine the bottom-up approach (Zellner and Montmarquette, 1971; Lütkepohl, 1986), which estimates the model at the daily frequency, produces daily forecasts, and then averages these forecasts to obtain the monthly forecast (following Benmoussa et al., 2025). Additionally, we include the random-walk forecast based on daily data, the end-of-month observation, which serves as the no-change forecast for all future values—whether averaged or not—under the random walk null, where it is MSFE-optimal (Ellwanger and Snudden, 2023) and directionally accurate (McCarthy and Snudden, 2024).

We report two common forecast evaluation criteria: the MSFE ratio and the success ratio for directional accuracy. Both measures are expressed relative to the period-average no-change forecast.

Let $\bar{y}_{m+h}$ denote the realized period average for target y over horizon h, and let $\hat{y}_{m+h|m}$ denote the forecast made at time m for $\bar{y}_{m+h}$. The MSFE ratio compares the candidate model's forecast

accuracy to that of the benchmark forecast $\hat{y}^{bench}_{m+h|m}$:

$$MSFE^{ratio}_{h} = \frac{\frac{1}{T}\sum_{m=1}^{T}\left(\bar{y}_{m+h} - \hat{y}_{m+h|m}\right)^2}{\frac{1}{T}\sum_{m=1}^{T}\left(\bar{y}_{m+h} - \bar{y}_{m}\right)^2}. \tag{30}$$

Where appropriate, we also conduct Diebold-Mariano tests (Diebold and Mariano, 1995) of equal forecast accuracy. Loss differentials are defined as the difference in squared errors, and statistical significance is assessed via Newey-West HAC standard errors (Newey and West, 1987).

Mean directional accuracy is denoted as the success ratio, SR_h, and measures the proportion of times the model correctly predicts the direction of change in the target variable:

$$SR_h = \frac{1}{T}\sum_{m=1}^{T} \mathbb{1}\left[\text{sgn}(\bar{y}_{m+h} - \bar{y}_m) = \text{sgn}(\hat{y}_{m+h|m} - \bar{y}_m)\right], \tag{31}$$

where $\mathbb{1}[\cdot]$ is the indicator function. We test for directional predictability using the independence test developed by Pesaran and Timmermann (2009).

3.1 Forecast Performance

We now simulate the forecast performance of monthly aggregates from a daily data-generating process estimated using alternative methods. The models considered include AR(p), MA(q), ARDL, and VAR(p) specifications. Unless otherwise noted, structural innovations are drawn i.i.d. from a standard normal distribution.[3]

We run 5,000 Monte Carlo replications, using 40 years of daily data and 21 trading days per month, to mimic the span of macro-financial series available since the mid-1980s. The sample is split 75–25 percent into estimation and forecast windows. Results are robust to shorter 10- or 20-year samples: the method ranking is unchanged, although the gap between eMIDAS/bottom-up and UMIDAS/RMIDAS widens as sample length falls because the latter suffer a heavier parameter penalty.

3.1.1 AR(1)

We begin with a high-frequency AR(1) data-generating process and consider a range of values for the persistence parameter ρ. We vary the persistence parameter over $0.90 \leq \rho \leq 1.00$, a range

[3]All findings are robust to Student-t innovations (5 degrees of freedom) and to conditional heteroskedasticity generated by a GARCH(1,1) process; the ranking of methods is unchanged.

typical for daily macro-financial series that are nearly indistinguishable from random walks. Unlike Tiao (1972), who stopped at $\rho \leq 0.90$, we therefore probe the high-persistence boundary where aggregation losses approach those of a unit-root process.

Table 1 shows that bottom-up forecasts consistently outperform the end-of-period no-change forecast for all $\rho < 1$. The bottom-up approach also delivers substantial improvements over direct forecasts based on monthly averages, with more than a 44 percent reduction in MSFE at $\rho = 0.995$. These findings confirm the forecast efficiency losses associated with temporal aggregation, echoing results for recursive forecasting (Lütkepohl, 1986; Benmoussa et al., 2025).

Table 1. AR(1) One-Month-Ahead Forecast Performance

Method	Average	Bottom-Up	eMIDAS	UMIDAS	RMIDAS	No-Change
Data	Monthly	Daily	Mixed	Mixed	Mixed	Mixed
ρ			MSFE Ratio			
1.00	5.815 (6.972)	0.544 (0.067)	0.545 (0.069)	0.578 (0.077)	0.617 (0.074)	0.540 (0.064)
0.995	0.972 (0.097)	0.540 (0.065)	0.540 (0.065)	0.573 (0.072)	0.609 (0.068)	0.561 (0.069)
0.99	0.919 (0.062)	0.539 (0.064)	0.539 (0.064)	0.572 (0.072)	0.606 (0.067)	0.582 (0.073)
0.98	0.863 (0.049)	0.537 (0.064)	0.537 (0.064)	0.570 (0.071)	0.599 (0.066)	0.626 (0.083)
0.95	0.756 (0.044)	0.531 (0.061)	0.532 (0.061)	0.564 (0.069)	0.583 (0.062)	0.768 (0.113)
0.90	0.654 (0.045)	0.525 (0.056)	0.526 (0.057)	0.558 (0.065)	0.563 (0.056)	1.035 (0.170)
ρ			Success Ratio			
1.00	0.577 (0.046)	0.737 (0.039)	0.737 (0.040)	0.727 (0.041)	0.714 (0.041)	0.738 (0.039)
0.995	0.596 (0.042)	0.740 (0.039)	0.740 (0.039)	0.729 (0.040)	0.718 (0.040)	0.733 (0.039)
0.99	0.608 (0.041)	0.740 (0.039)	0.739 (0.039)	0.730 (0.040)	0.719 (0.040)	0.727 (0.039)
0.98	0.627 (0.038)	0.740 (0.039)	0.740 (0.039)	0.731 (0.040)	0.720 (0.040)	0.715 (0.040)
0.95	0.666 (0.036)	0.742 (0.039)	0.741 (0.039)	0.733 (0.040)	0.725 (0.039)	0.688 (0.043)
0.90	0.702 (0.036)	0.743 (0.038)	0.743 (0.038)	0.735 (0.038)	0.731 (0.037)	0.654 (0.044)

Note: Comparison of monthly forecasts for simulated AR(1) model at the daily frequency when estimated with alternative methods. 5000 Monte Carlo simulations. All MSFE and success ratios relative to the monthly average no-change. Assumes 21 days in a month and 40 years of observations in addition to burning the first 500 days, with a 75–25 percent estimation-forecast sample. Last column is the no-change of the daily data using the end-of-month observation.

The eMIDAS forecasts match the accuracy of the bottom-up approach, illustrating the equivalence between efficient direct (eMIDAS) and recursive (bottom-up) forecasts. The UMIDAS specification, which uses $n - 1$ unrestricted parameters, performs worse, suggesting inefficiency due to overparameterization. The RMIDAS model, which imposes a restricted Almon lag structure, does not improve upon UMIDAS and underperforms relative to both eMIDAS and the bottom-up method.[4] These results reinforce the theoretical equivalence between bottom-up and eMIDAS

[4]We tested RMIDAS with Beta weights; in all cases these produced higher MSFEs than the corresponding Almon weights. For both weight schemes the default window was the $n - 1$ daily lags used in UMIDAS. Trimming that window to 15, 10, or 5 lags narrowed the gap between RMIDAS and UMIDAS, but all variants were outperformed

forecasts.

3.1.2 MA(q)

We now consider a daily MA(q) data-generating process aggregated to the monthly frequency. In addition to the forecasting methods from the previous subsection, we test whether incorporating a low-frequency moving average term (i.e., adding an MA(1)) improves accuracy, as suggested by Foroni et al. (2019).

Table 2. MA(q) One-Month-Ahead Forecast Performance

Model	Daily Lags	MA(5)	MA(10)	MA(15)
			MSFE Ratio	
No-Change	-	2.113 (0.389)	1.249 (0.219)	0.914 (0.156)
Monthly MA(1)	-	0.523 (0.051)	0.558 (0.047)	0.589 (0.046)
UMIDAS	$n-1+q$	0.529 (0.060)	0.512 (0.064)	0.484 (0.066)
RMIDAS	$n-1$	0.526 (0.049)	0.520 (0.051)	0.485 (0.055)
Bottom-up	q	0.521 (0.050)	0.554 (0.053)	0.595 (0.059)
eMIDAS+MA(1)	q	0.513 (0.056)	0.496 (0.056)	0.477 (0.065)
eMIDAS	q	0.506 (0.052)	0.493 (0.056)	0.473 (0.060)
			Success Ratio	
No-Change	-	0.596 (0.046)	0.638 (0.046)	0.668 (0.045)
Monthly MA(1)	-	0.747 (0.036)	0.734 (0.036)	0.722 (0.035)
UMIDAS	$n-1+q$	0.744 (0.037)	0.749 (0.037)	0.759 (0.038)
RMIDAS	$n-1$	0.743 (0.035)	0.746 (0.036)	0.757 (0.038)
Bottom-up	q	0.745 (0.034)	0.735 (0.036)	0.723 (0.036)
eMIDAS+MA(1)	q	0.752 (0.038)	0.754 (0.035)	0.760 (0.038)
eMIDAS	q	0.750 (0.036)	0.754 (0.037)	0.761 (0.039)

Note: Comparison of monthly forecasts for simulated MA(q) model at the daily frequency when estimated with alternative methods. 5000 Monte Carlo simulations. All MSFE and success ratios relative to the monthly average no-change. Assumes 21 days in a month and 40 years of observations in addition to burning the first 500 days, with a 75-25 percent estimation-forecast sample.

The forecasts depend on high-frequency moving average components that are not directly observable in practice. For the bottom-up approach, these are estimated on the daily data using maximum likelihood. For the "Monthly MA(1)", a single moving average term is estimated on the monthly data. In contrast, for UMIDAS and eMIDAS, we estimate on the last $n-1+q$ and q daily observations, respectively. For RMIDAS we use Almon weights, estimated using the $n-1$ daily observations.

As shown in Table 2, eMIDAS consistently delivers the lowest MSFE across all lag lengths.

by eMIDAS.

Adding a monthly MA(1) term does not improve forecast accuracy; instead, it increases the standard deviation of the estimates, indicating that such terms are redundant in the presence of mixed-frequency predictors. This finding suggests that low-frequency moving average terms are not needed.

Interestingly, both UMIDAS, RMIDAS and eMIDAS outperform the bottom-up approach, even though the true data-generating process is a daily MA process. The gap reflects estimation efficiency: recursive bottom-up must fit a daily MA(q) and infer q latent shocks, whereas MIDAS regressions work directly with the observed daily series, introducing less estimation noise into the monthly forecast.

3.1.3 Exogenous Predictor

We now consider a multivariate data-generating process in which the dependent variable is observed only at a low frequency, while the predictor is observed at a higher frequency. This setting is common in macroeconomic applications, where indicators such as oil prices or interest rates are available daily, while target variables such as inflation or GDP are reported monthly. The model follows the external predictor framework in section 2.1.2, equations 15 and 17. We simulate outcomes for different degrees of persistence in the predictor process, using alternative parameterizations of the persistent component ($\rho = 0.995$ and $\rho = 0.98$), and varying signal strength in the distributed lag coefficients ($\gamma = 20$ and $\gamma = 0.5$).

We compare forecast performance across four specifications: eMIDAS, UMIDAS, RMIDAS, and a baseline end-of-period no-change (EoP NC) forecast. Unlike in the AR(1) case, forecasts must be constructed using mixed-frequency methods by design, as the low-frequency dependent variable cannot be reconstructed from daily data alone.

Table 3 presents the results. Across all parameterizations, eMIDAS delivers the largest reductions in MSFE—up to 53 percent relative to the monthly average no-change benchmark. These gains are robust to both the persistence of the predictor and the strength of the distributed lag signal. Notably, eMIDAS consistently outperforms both UMIDAS and RMIDAS specifications. The performance of UMIDAS and RMIDAS is more sensitive to the lag structure. In particular, RMIDAS—despite using a parsimonious Almon lag profile—tends to underperform the unrestricted UMIDAS. This finding aligns with the theoretical results (Proposition 4), which establish the optimality of eMIDAS when the high-frequency predictor is observed and the aggregation structure of the dependent variable is known.

Table 3. Exogenous Predictor One-Month-Ahead Forecast Performance

Parameters		MSFE Ratio			
ρ	γ	EoP NC	eMIDAS	UMIDAS	RMIDAS
0.995	20	0.562 (0.069)	0.473 (0.061)	0.501 (0.067)	0.541 (0.064)
0.995	0.50	0.864 (0.128)	0.473 (0.061)	0.502 (0.068)	0.540 (0.065)
0.980	20	0.629 (0.083)	0.475 (0.061)	0.503 (0.068)	0.538 (0.063)
0.980	0.50	0.997 (0.153)	0.476 (0.061)	0.504 (0.068)	0.537 (0.064)
		Success Ratio			
ρ	γ	EoP NC	eMIDAS	UMIDAS	RMIDAS
0.995	20	0.732 (0.040)	0.760 (0.038)	0.752 (0.039)	0.739 (0.040)
0.995	0.50	0.674 (0.043)	0.760 (0.038)	0.752 (0.039)	0.740 (0.040)
0.980	20	0.715 (0.041)	0.759 (0.038)	0.751 (0.038)	0.739 (0.039)
0.980	0.50	0.658 (0.044)	0.758 (0.038)	0.750 (0.039)	0.739 (0.040)

Note: Comparison of monthly forecasts for simulated Exogenous Predictor model at the daily frequency when estimated with alternative methods. 5000 Monte Carlo simulations. All MSFE and success ratios relative to the monthly average no-change. Assumes 21 days in a month and 40 years of observations in addition to burning the first 500 days, with a 75-25 percent estimation-forecast sample.

Finally, we note that extending the model to include additional low-frequency moving average terms did not improve forecast performance. These findings reinforce the core insight that high-frequency predictors can be used efficiently for mixed-frequency forecasting without expanding the low-frequency state space. Overall, the simulation results confirm that eMIDAS yields robust gains in forecast accuracy and directional precision when the target variable is observed only at a low frequency.

3.1.4 VAR

Consider now mixed-frequency direct forecasts from a daily bivariate VAR(p) data-generating process, where the series are aggregated to the monthly frequency. Let the daily observations $y_{1,t,i}$ and $y_{2,t,i}$ be governed by a structural VAR with a recursive identification scheme.

$$\begin{bmatrix} y_{1t,i} \\ y_{2t,i} \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} \begin{bmatrix} y_{1t,i-1} \\ y_{2t,i-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \eta_2 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t,i} \\ \epsilon_{2t,i} \end{bmatrix}$$

We consider four alternative parameterizations of the SVAR. In all cases, the first variable exhibits a high degree of autoregressive persistence with $\rho_{11} = 0.99$, while the second variable has either high or low persistence, with $\rho_{22} = 0.99$ or $\rho_{22} = 0.45$. We assume a recursive ordering such that $y_{1,t,i}$ affects $y_{2,t,i}$ contemporaneously through the parameter η_2. For the cross-variable

dynamics, lags of $y_{1,t,i}$ affect $y_{2,t,i}$ either positively or negatively, with $\rho_{21} = 0.5$, which induces greater persistence in the second variable, or $\rho_{21} = -0.5$, which reduces it.

The results, shown in Table 4, support the efficiency of eMIDAS forecasts. When both variables are sampled and forecasted using monthly averages, forecast performance is poor relative to the bottom-up approach. The UMIDAS forecast, which uses $n - 1$ parameters, is inefficient and less precise, suggesting overparameterization. The RMIDAS model, estimated with an Almon lag profile, performs better than UMIDAS but not as well as the eMIDAS or bottom-up approaches. When the second variable has low persistence but responds positively to $y_{1,t-1,i}$, there is little effect on the forecast performance of the first variable. In contrast, when $y_{2,t,i}$ responds negatively to $y_{1,t-1,i}$, UMIDAS performs notably worse, though forecast accuracy remains unaffected for both the bottom-up and eMIDAS methods.

Table 4. VAR One-Month-Ahead Forecast Performance

Parameters		MSFE Ratio				
ρ_{22}	ρ_{21}	Monthly	Bottom Up	eMIDAS	UMIDAS	RMIDAS
0.99	0.50	0.928 (0.039)	0.540 (0.064)	0.536 (0.063)	0.612 (0.069)	0.603 (0.070)
0.99	-0.50	0.928 (0.039)	0.540 (0.064)	0.536 (0.063)	0.761 (0.075)	0.603 (0.069)
0.45	0.50	0.905 (0.041)	0.539 (0.063)	0.536 (0.063)	0.612 (0.069)	0.607 (0.070)
0.45	-0.50	0.778 (0.064)	0.539 (0.063)	0.536 (0.063)	0.761 (0.075)	0.591 (0.068)
		Success Ratio				
ρ_{22}	ρ_{21}	Monthly	Bottom Up	eMIDAS	UMIDAS	RMIDAS
0.99	0.50	0.589 (0.038)	0.739 (0.039)	0.741 (0.040)	0.715 (0.041)	0.719 (0.041)
0.99	-0.50	0.588 (0.038)	0.739 (0.039)	0.741 (0.040)	0.665 (0.043)	0.719 (0.040)
0.45	0.50	0.601 (0.040)	0.739 (0.038)	0.741 (0.040)	0.715 (0.041)	0.717 (0.041)
0.45	-0.50	0.659 (0.042)	0.739 (0.039)	0.741 (0.039)	0.666 (0.043)	0.723 (0.040)

Note: Comparison of monthly forecasts for simulated VAR(1) model at the daily frequency when estimated with alternative methods. 5000 Monte Carlo simulations. All MSFE and success ratios relative to the monthly average no-change. Assumes 21 days in a month and 40 years of observations in addition to burning the first 500 days, with a 75-25 percent estimation-forecast sample.

In the VAR(1) framework, we also capture differences between the direct (eMIDAS) and indirect (bottom-up) approaches. In this context, the inefficiency of the latter arises from downward biases in the autoregressive parameter due to finite-sample effects. In the direct forecast, these biases are diluted by averaging across the autoregressive parameters in each daily forecast, whereas they are compounded in the recursive approach. As such, we observe a slight improvement in eMIDAS relative to the bottom-up method.

Next, we consider a VAR(1) scenario in which high-frequency information on the first variable,

Table 5. Daily VAR(1) - Missing High-Frequency Information for y_{1t}

Parameters		MSFE Ratio				
ρ_{22}	ρ_{21}	Monthly	~~Bottom Up~~	~~eMIDAS~~	UMIDAS	RMIDAS
0.99	0.50	0.918 (0.029)	0.540 (0.064)	0.921 (0.028)	0.601 (0.066)	0.725 (0.069)
0.99	-0.50	0.918 (0.029)	0.540 (0.064)	0.919 (0.029)	0.680 (0.067)	0.745 (0.070)
0.45	0.50	0.897 (0.038)	0.539 (0.063)	0.640 (0.066)	0.599 (0.066)	0.674 (0.069)
0.45	-0.50	0.770 (0.061)	0.539 (0.063)	0.782 (0.061)	0.676 (0.067)	0.728 (0.069)
		Success Ratio				
ρ_{22}	ρ_{21}	Monthly	~~Bottom Up~~	~~eMIDAS~~	UMIDAS	RMIDAS
0.99	0.50	0.591 (0.038)	0.739 (0.039)	0.589 (0.037)	0.719 (0.040)	0.677 (0.042)
0.99	-0.50	0.590 (0.038)	0.739 (0.039)	0.590 (0.037)	0.693 (0.042)	0.671 (0.042)
0.45	0.50	0.603 (0.040)	0.739 (0.038)	0.706 (0.041)	0.719 (0.040)	0.695 (0.041)
0.45	-0.50	0.661 (0.042)	0.739 (0.039)	0.655 (0.042)	0.694 (0.042)	0.677 (0.042)

Note: Bottom-up and eMIDAS are struck out since they cannot be implemented in this setting. "Bottom-up" assumes high-frequency information for $y_{t,i}$ is observed, and "eMIDAS" is approximated by replacing $y_{1,t,i}$ with its monthly aggregates.

$y_{1,t,i}$, is not observed by the practitioner. This setting corresponds precisely to that proposed in the UMIDAS literature. The results are shown below in Table 5.

In this setting, we note that neither the bottom-up approach nor eMIDAS can be implemented in practice, as both require high-frequency information for all variables. Instead, for eMIDAS, we attempt to approximate its structure by replacing $y_{1,t,i}$ with its monthly aggregates. As expected, this approximation performs poorly. When the autoregressive parameter is high (i.e., $\rho_{22} = 0.99$), the problem of missing data is amplified, and the performance of the approximated eMIDAS is even worse than that of forecasts based on monthly aggregates. However, when persistence in the variable with missing data is lower, eMIDAS performs comparably to RMIDAS—for example, when $\rho_{22} = 0.45$ and $\rho_{21} = 0.50$. None of the approaches, however, match the efficiency of the bottom-up method, which recall, is not actually feasible in this setting as it assumes the availability of high-frequency data for $y_{1,t,i}$.

This result offers a revealing insight about existing MIDAS methods: while they yield gains relative to forecasts based on temporally aggregated data, they differ from best practices when the practitioner observes all high-frequency information. As such, it should be a priority for practitioners to obtain high-frequency information whenever possible. In the absence of such high-frequency information, the UMIDAS and RMIDAS structures offer a complementary way to recover the high-frequency information.

3.2 In-Sample Testing

Simulations are now used to analyze in-sample tests of parameterization in mixed-frequency models. We begin by testing the parameterization of the MIDAS structure when the daily data-generating process follows the AR(1) process described in equation 1, using a UMIDAS specification that includes a low-frequency moving average term:

$$\bar{y}_{t+k} = \sum_{j=0}^{n-1} \beta_j y_{t-1,n+1-j} + \varepsilon_{t+k} + \alpha\varepsilon_{t+k-1},$$

where the β_j's are the UMIDAS coefficients on the daily observations, covering up to $p(n-1)$ high-frequency lags (Foroni et al., 2015), and α is the coefficient on the low-frequency moving average term (Foroni et al., 2019).

Table 6 reports the share of simulations rejecting the null hypothesis of zero coefficients at the 5 percent level, based on 5,000 simulations of daily data aggregated to the monthly frequency ($n = 21$). Rejection rates are shown for the null hypotheses $\beta_1 = 0$, $\beta_2 = \beta_3 = \cdots = \beta_n = 0$, and $\alpha = 0$, using t-, F-, and Wald tests, respectively.

Table 6. Rejection Frequency, High-Frequency Parameterization

	Sample Size = 20			Sample Size = 40		
ρ	$\beta_1 = 0$	$\beta_j = 0, \forall j > 1$	$\alpha = 0$	$\beta_1 = 0$	$\beta_j = 0, \forall j > 1$	$\alpha = 0$
1.00	1.000	0.050	0.052	1.000	0.051	0.046
0.995	1.000	0.048	0.056	1.000	0.050	0.046
0.990	1.000	0.049	0.058	1.000	0.050	0.048
0.980	1.000	0.050	0.062	1.000	0.051	0.050
0.950	1.000	0.052	0.063	1.000	0.052	0.057
0.900	1.000	0.051	0.057	1.000	0.050	0.052
0.500	0.635	0.046	0.056	0.888	0.053	0.047
0.000	0.052	0.048	0.054	0.049	0.056	0.050

Notes: Tested at the 0.05 level using 5000 simulations of a daily AR(1) using UMIDAS estimated with OLS. DM-test refers to Diebold and Mariano (1995) with Newey and West (1987) standard errors and compared against the standard normal.

The results show near-perfect power to reject the null hypothesis that $\beta_1 = 0$ for monthly data across the sample sizes considered. In contrast, we reject the null that $\beta_2 = \beta_3 = \cdots = \beta_n = 0$ only at the nominal significance level, providing no evidence that additional daily terms are needed. Moreover, there is little evidence to reject $\alpha = 0$ for monthly data, regardless of sample size.[5] These

[5]Rejection above the nominal level occurs only when the model is already overparameterized with $j_{\max} = n - 1$.

findings numerically confirm the exact UMIDAS structure described in Section 2.

Additional analysis using alternative underlying data-generating processes confirms that statistical significance arises only for the coefficients implied by the eMIDAS structure. This includes high-frequency data-generating processes with additional autoregressive lags, as well as MA and VARMA models. The results support the validity of classical inference on high-frequency parameters within the UMIDAS framework for uncovering the structure of the data.

Given its use in existing studies, we also examine the Diebold-Mariano (DM) test statistic of Diebold and Mariano (1995) to assess equality in mean-squared out-of-sample one-step forecast error over the second half of the sample. The DM test likewise fails to detect any significant improvement in forecast accuracy from the inclusion of additional low-frequency moving average terms.[6] This is consistent with the forecast results in section 3.2, which showed no additional forecast gains from incorporating extra terms.

4 Exact MIDAS when High-Frequency Data is Not Available

We show that high-frequency information on the explanatory variable is sufficient for forecast efficiency when the response does not depend on its own lags; otherwise, observing the high-frequency response is also required. However, in practice, it is often the case that a noisy proxy is observed that closely relates to the unobserved underlying high-frequency data. For example, daily measures of inflation may exist that are closely related to official monthly inflation measures. In such a situation, we want to understand the value of observing such a proxy, as well as the appropriate MIDAS structure to employ. We begin by considering the conditions under which the use of proxy variables will still emulate the structure of eMIDAS.

4.1 Proxy Structure

Suppose we have high-frequency information on $x_{t,i}$, which is a candidate proxy for unobserved $y_{t,i}$. In this setting, can we infer the daily structure from $y_{t,i}$ using $x_{t,i}$, and forecast the aggregate of $\bar{y}_t$ with comparable efficiency to eMIDAS? To formalize the setting, we consider when the unobserved

[6] Size distortion of the DM test is found to be minor in our setting; adjustments that account for nested models yield identical conclusions. In the empirical applications, the benchmark period-average no-change forecast is non-nested, so these adjustments are unnecessary.

daily data follows an AR(1), and the proxy is described by:

$$\begin{aligned} y_{t,i} &= Ay_{t,i-1} + \varepsilon_{t,i}, \\ x_{t,i} &= By_{t,i} + \nu_{t,i}, \end{aligned} \tag{32}$$

where B is the "strength" of the proxy and $\nu_{t,i}$ is an i.i.d. innovation which measures its noise.

4.2 Proxy Variable: Direct Replacement

Let us first consider a "naive" method of using the proxy variable, where we directly replace $y_{t,i}$ with $x_{t,i}$. In a first step, we regress $x_{t,i}$ on $x_{t,i-1}$ at the daily level, that is, we fit the model to the observed proxy:

$$x_{t,i} = A'x_{t,i-1} + \upsilon_{t,i}, \tag{33}$$

which yields an estimate for the parameter A'. This implies the daily forecast:

$$x_{T+k,i} = (A')^{(k-1)n+i} x_{T,n}, \tag{34}$$

at all horizons $T + k, i$. Next, using the bottom approach, we consider the ex-post average and obtain a similar expression to the distributed lag example in Section 2.2.1:

$$\mathbb{E}_{T,n}\left[\bar{y}_{T+k}\right] = \beta'_k x_{T,n}, \tag{35}$$

where $\beta'_k = \frac{1}{n}\sum_{i=1}^{n} BA'^{(k-1)n+i}$. In other words, the bottom-up approach using the proxy variable as a replacement will yield a forecast which projects directly on the most recent observed value for the proxy, $x_{T,n}$. Immediately, this tells us that the use of a proxy variable in this context will yield a structure consistent to that of eMIDAS. In particular, we are still projecting only on information from the most recent value, which is consistent with the underlying daily structure of y_t.

However, to understand exactly how it relates to eMIDAS, note that $x_{T,n} = By_{t,n} + \upsilon_{t,i}$ and the forecast can be written in terms of $y_{t,n}$ as follows:

$$\mathbb{E}_{T,n}\left[\bar{y}_{T+k}\right] = \beta'_k B y_{T,n} + \beta'_k \upsilon_{T,n}. \tag{36}$$

Hence, the projection on the most recent proxy will yield a result that depends on the most recent daily value $y_{T,n}$ in addition to a noise term $\upsilon_{T,n}$. Indeed, whether eMIDAS can be recovered in this

context depends solely on two parameters: i) the strength of the proxy B (which implicitly governs A'), and ii) the latest measurement noise $\upsilon_{t.n}$. When $B = 1$, it can be shown that $A \approx A'$ and therefore $\beta_k \approx \beta'_k$. Further, if the noise approaches 0, then we recover the exact MIDAS structure. In other words, this proxy strategy is a highly effective method of forecast in this context if a strong and noiseless proxy for y_t is available to the practitioner.

4.3 Proxy Variable: State Space Representation and Kalman Filter

It is possible that a strong proxy may not be available, or the practitioner is unable to gauge the strength of such a proxy. For instance, if $B \to 0$ or if the measurement error $\upsilon_{T,n}$ is large, then we have a biased forecast that will greatly differ from exact MIDAS. In this case, one may prefer a more conservative approach to reduce the potential bias in the forecast. Indeed, the model in equation 32 admits a state space representation where $y_{t,i}$ is the state equation (containing the unobservables) and $x_{t,i}$ is the observation equation. By using the Kalman filter, we obtain at horizon $T + k, i$:

$$\mathbb{E}\left[y_{t+k,i}\right] = A^{(k-1)n+i}\hat{y}_{T,n}, \tag{37}$$

where $\hat{y}_{T,n} = \hat{y}_{T,n-1} + K_{T,n}(x_{T,n} - B\hat{y}_{T,n-1})$ and $K_{T,n}$ is the Kalman gain at time T, n. Thus, using the bottom up approach we get:

$$\mathbb{E}_{T,n}\left[\bar{y}_{t+k}\right] = \beta_k \hat{y}_{T,n}. \tag{38}$$

Thus, we again get a similar structure to eMIDAS, where instead of using a projection on the true value of $y_{T,n}$, it is replaced with the filtered version through a proxy $x_{t,i}$. Of course, the accuracy and similarity to the eMIDAS forecast will depend on the result of the filtering; like in the naive proxy case, if the proxy is strong with little to no noise, then we expect to see a close resemblance between this state space approach and eMIDAS.

4.4 Proxy Variable Simulations

Simulation results in Table 7 report the relative MSPE ratios comparing the performance of alternative MIDAS forecasts of $\bar{y}_t$ using the noisy proxy $x_{t,i}$. The two proxy-based eMIDAS estimation methods–direct replacement and Kalman Filter—are compared, along with RMIDAS and UMIDAS, for varying levels of proxy strength and noise-to-signal ratios.[7] Each row corresponds to a

[7]State–space parameters are estimated by maximum prediction-error likelihood (see Harvey, 1989; Durbin and Koopman, 2012, Chs. 3–4). We initialise the Kalman filter with a diffuse prior following De Jong (1991), and estimate

distinct data-generating process based on the structure in equation 32. Higher values of B indicate a stronger relationship between the proxy $x_{t,i}$ and the high-frequency component of the target variable $y_{t,i}$. Higher noise-to-signal ratios, defined as $\sigma_\nu/\sigma_\varepsilon$, reflect increased measurement noise.

Table 7. Relative MSPE Ratios: Alternative Noise-to-Signal Ratios and Proxy Strengths ($\rho = 0.995$)

Noise to Signal	Proxy Strength	eMIDAS (state space)	eMIDAS (direct)	UMIDAS	RMIDAS
0	1	0.609 (0.068)	**0.540** (0.065)	0.573 (0.072)	0.609 (0.068)
	0.75	0.609 (0.068)	**0.540** (0.065)	0.573 (0.072)	0.609 (0.068)
	0.50	0.609 (0.068)	**0.540** (0.065)	0.573 (0.072)	0.609 (0.068)
	0.25	0.609 (0.068)	**0.540** (0.065)	0.573 (0.072)	0.609 (0.068)
0.25	1	0.614 (0.067)	**0.545** (0.064)	0.577 (0.071)	0.611 (0.068)
	0.75	0.617 (0.067)	**0.548** (0.065)	0.580 (0.071)	0.612 (0.068)
	0.50	0.624 (0.068)	**0.558** (0.067)	0.588 (0.072)	0.615 (0.068)
	0.25	0.653 (0.070)	**0.611** (0.078)	0.617 (0.074)	0.628 (0.069)
0.50	1	0.624 (0.068)	**0.558** (0.067)	0.588 (0.072)	0.615 (0.068)
	0.75	0.633 (0.069)	**0.572** (0.070)	0.597 (0.072)	0.618 (0.068)
	0.50	0.655 (0.072)	**0.611** (0.078)	0.617 (0.074)	0.628 (0.069)
	0.25	0.716 (0.073)	0.813 (0.118)	0.686 (0.079)	**0.673** (0.072)
0.75	1	0.637 (0.068)	**0.580** (0.072)	0.602 (0.073)	0.620 (0.068)
	0.75	0.652 (0.069)	**0.611** (0.078)	0.617 (0.074)	0.628 (0.069)
	0.50	0.687 (0.074)	0.696 (0.095)	0.651 (0.077)	**0.648** (0.070)
	0.25	0.786 (0.091)	1.125 (0.181)	0.756 (0.085)	**0.730** (0.076)
1	1	0.655 (0.071)	**0.611** (0.078)	0.617 (0.074)	0.628 (0.069)
	0.75	0.673 (0.070)	0.664 (0.089)	**0.640** (0.076)	0.641 (0.070)
	0.50	0.716 (0.073)	0.813 (0.118)	0.686 (0.079)	**0.673** (0.072)
	0.25	0.848 (0.087)	1.517 (0.265)	0.826 (0.091)	**0.792** (0.081)

Note: Simulation results based on 5000 Monte Carlo replications of a daily AR(1) process with autoregressive coefficient $\rho = 0.995$. Forecasts are evaluated relative to the monthly average no-change benchmark. Each simulation assumes 21 trading days per month, 40 years of data, and a 75%–25% split between estimation and forecast samples, with the first 500 daily observations discarded as burn-in.

First, consider the case when the noise-to-signal ratio is 0. With $B = 1$, the high-frequency data is effectively observed since $y_{t,i} = x_{t,i}$, and when $B < 1$, we observe a perfectly correlated proxy. In such a setting, direct substitution for eMIDAS provides the same efficiency gains as eMIDAS and the bottom-up approach, showing the same results as Table 1.

Similarly, the results show that eMIDAS via direct-substitution consistently achieves the lowest MSPE when the proxy is strong and the noise level is low. For example, when the proxy strength is high $B \geq 0.5$ and the noise-to-signal ratio is 0.5, eMIDAS outperforms both UMIDAS and RMIDAS. However, when the information in the proxy becomes unreliable, i.e. as the signal weakens or

transition/measurement covariances jointly with the system matrices. Simulation results are insensitive to alternative (informative) initialisations.

noise increases, this performance gap narrows. In such a setting, the state space representation outperforms direct substitution of eMIDAS, but both are outperformed by UMIDAS and RMIDAS.

Overall, the simulation results confirm the theoretical findings of Sections 4.2 and 4.3. When a strong and clean proxy is available, using it directly in an eMIDAS structure is preferred. However, when the proxy is noisy and only weakly correlated with the target, alternative forms of MIDAS may be preferred. Thus, eMIDAS can be considered complementary to existing MIDAS methods, where the latter are expected to outperform when neither the underlying high-frequency data nor a strong proxy exists.

5 Application

5.1 Monthly Average Interest Rates

We now analyze real-time monthly average level forecasts of the nominal short-term interest rate on the U.S. 3-Month Treasury Bill. Daily closing data are obtained from the FRED database of the Federal Reserve Bank of St. Louis (ticker: DGS3MO). End-of-month observations correspond to the closing price on the last trading day of each month, and the monthly average is calculated as the simple average of daily closing prices. These data are available in real time and are not subject to historical revisions. The time series begins in January 1990, with the forecast evaluation sample running from January 2000 to June 2023.

We evaluate the forecast performance of five methods: recursive forecasts from the bottom-up approach and direct forecasts using monthly averages, eMIDAS, UMIDAS, and restricted MIDAS with an Almon lag profile.[8] MIDAS specifications include 21 daily lags, corresponding to the average number of trading days in a month. The bottom-up forecast is estimated on the first difference of the rate, with model parameterization selected using Akaike Information Criterion (AIC; Akaike, 1974).[9] Forecasts for the nominal level on day i of month $t+h$, conditional on information available at month t, are based on the model-implied cumulative sum between day t, n and day $t+h, i$. These daily forecasts are then averaged to obtain the monthly forecast. The eMIDAS specification uses the number of lags suggested by the parameterization of the bottom-up model.

[8] Beta restrictions failed to converge—regardless of starting values or estimation methods—during at least one period of the expanding window estimation and are therefore not reported.

[9] Using other standard information criteria yields the same forecast ranking. Likewise, results are insensitive to ancillary modelling choices such as drift terms, seasonal adjustment, or error-distribution assumptions—the first-order driver of the results is the information loss from temporal aggregation.

We report both the MSFE ratio and the success ratio, each expressed relative to the conventional monthly average no-change forecast to facilitate comparison with previous studies. Tests for forecast gains relative to the end-of-month no-change forecast are reported in the appendix. P-values relative to the monthly average benchmark are reported in brackets for the null hypothesis of equal MSFE between the model-based forecast and the no-change forecast, and for the null of no directional accuracy.

Table 8. Real-Time Model-based Forecasts of the Nominal 3-Month Treasury Bills

Method	Direct	RMIDAS	eMIDAS	UMIDAS	Bottom Up
Data	Average	Mixed	Mixed	Mixed	Daily
horizon			MSFE Ratios		
1	1.01 (0.579)	0.66 (0.001)	0.56 (0.000)	0.70 (0.008)	0.57 (0.000)
3	0.99 (0.412)	0.89 (0.014)	0.82 (0.003)	0.97 (0.399)	0.86 (0.027)
6	0.96 (0.239)	1.02 (0.588)	0.88 (0.017)	1.01 (0.524)	0.96 (0.305)
12	0.89 (0.042)	1.09 (0.826)	0.85 (0.013)	0.91 (0.154)	1.03 (0.614)
24	0.81 (0.005)	1.08 (0.760)	0.80 (0.003)	0.83 (0.012)	1.13 (0.820)
			Success Ratios		
1	0.46 (1.000)	0.63 (0.000)	0.74 (0.000)	0.68 (0.000)	0.70 (0.000)
3	0.49 (0.975)	0.57 (0.005)	0.61 (0.000)	0.60 (0.001)	0.54 (0.000)
6	0.50 (0.566)	0.56 (0.020)	0.58 (0.002)	0.57 (0.011)	0.51 (0.147)
12	0.54 (0.197)	0.55 (0.000)	0.60 (0.000)	0.65 (0.000)	0.54 (1.000)
24	0.55 (0.260)	0.56 (0.000)	0.58 (0.003)	0.60 (0.010)	0.55 (1.000)

Note: Real-time, out-of-sample forecasts of the real monthly average 3-month Treasury bills in levels, 2000M1–2023M6. Forecast criteria relative to the monthly average no-change forecast, with p-values reported in brackets. RMIDAS uses Almon restrictions.

The results for the monthly average forecasts of the yields on the U.S. 3-Month Treasury Bill are reported in Table 8. eMIDAS outperforms the other approaches. The bottom-up method performs well at short horizons but fails to converge to the long-run mean. eMIDAS improves forecast accuracy at short horizons relative to both the RMIDAS and UMIDAS models. At longer horizons, eMIDAS and RMIDAS perform similarly and both outperform the UMIDAS forecasts. These results confirm the practical gains from using eMIDAS.

5.2 Multivariate Headline Inflation Forecasts

We next assess the performance of eMIDAS in forecasting U.S. headline Consumer Price Index (CPI) inflation. The focus is on one-month-ahead forecasts of month-over-month changes in the seasonally adjusted headline CPI, expressed in percentage terms. Forecasts are constructed in real time over the period January 2000 to January 2023 using an expanding window estimation. We

compare the forecast performance of eMIDAS with a standard monthly AR(2) model, a random walk with drift (RWWD), UMIDAS, and RMIDAS. The benchmark models are estimated using monthly data, while the eMIDAS and UMIDAS specifications incorporate high-frequency predictors.

All mixed-frequency models use a common set of explanatory variables, including daily lags of West Texas Intermediate (WTI) crude oil prices and the 3-Month Treasury Bill rate, as well as weekly lags of U.S. retail gasoline prices. The eMIDAS model employs a sequential specification search, beginning with the UMIDAS specification and removing higher-order lags based on recursive backward F-tests. RMIDAS applies the restricted Almon lag polynomial to the same set of high-frequency predictors. The number of daily and weekly lags corresponds to one calendar month.

Table 9. Forecast Performance One-Month-Ahead Headline Inflation

Method	RWWD	AR(2)	eMIDAS	UMIDAS	RMIDAS
Data	Average	Average	Mixed	Mixed	Average
			MSFE		
	0.896 (0.227)	0.723 (0.002)	0.468 (0.000)	0.589 (0.000)	0.489 (0.000)
			Success Ratio		
	0.717 (0.000)	0.713 (0.000)	0.781 (0.000)	0.773 (0.000)	0.753 (0.000)
NDA	0.709	0.709	0.764	0.748	0.724
PDA	0.726	0.718	0.798	0.798	0.782
EDA	0.833	0.800	0.933	0.900	0.900
NEDA	0.701	0.701	0.760	0.756	0.733

Note: Real-time, out-of-sample forecasts of the monthly headline CPI inflation rate, 2000M1–2023M1. Forecast criteria relative to the no-change forecast, with p-values reported in brackets. RMIDAS uses Almon restrictions. DA means Directional Accuracy. NDA and PDA measure accuracy conditional on negative and positive inflation, respectively. EDA (Extreme DA) measures accuracy during extreme movements—defined as one standard deviation beyond the forecast sample mean — while NEDA (Non-Extreme DA) evaluates performance inside those events.

Table 9 presents the forecast results. The MSFE results reveal that eMIDAS substantially outperforms all other methods, achieving a 35% reduction in MSFE relative to the AR(2) model and even larger gains compared to the RWWD and the no-change forecast. RMIDAS also performs well in terms of MSFE, though it consistently underperforms eMIDAS. The success ratio results tell a complementary story. Interestingly, although RMIDAS outperforms UMIDAS in terms of MSFE, UMIDAS achieves higher directional accuracy. These results highlight the value of using quantitatively reduced unrestricted functional forms to capture the contribution of high-frequency predictors in modeling monthly inflation dynamics.

While headline directional accuracy provides useful information, performance in cases of large

inflation movements is particularly important for monetary policy. eMIDAS exhibits strong performance in both positive and negative inflation realizations—PDA and NDA—and its accuracy is especially pronounced in extreme inflation episodes (EDA), defined as outcomes exceeding one standard deviation above or below the forecast sample mean. In these extreme cases, the success ratio of eMIDAS reaches 0.933, compared to 0.900 for both RMIDAS and UMIDAS, and 0.800 for the AR(2) model.

Taken together, these results demonstrate that eMIDAS offers substantial improvements in forecast accuracy over competing methods when high-frequency information is available. Notably, the gains over RMIDAS and UMIDAS are concentrated in the tails of the inflation distribution, where precision is especially valuable for risk management (see the GDP-at-Risk framework of Adrian et al., 2019). Similar tail-risk benefits from exploiting high-frequency data have been documented for GDP Growth-at-Risk (Ferrara et al., 2022) and downside GDP risk (Chan et al., 2025). That eMIDAS delivers additional gains over existing MIDAS approaches therefore points to a promising direction for future research on tail-focused forecasting.

6 Conclusion

This paper uses the bottom-up approach to derive the exact structure of direct mixed-frequency sampling models. We demonstrate that the structure of eMIDAS is independent of the degree of time aggregation, relying solely on the disaggregated data-generating process. These insights substantially reduce parameterization relative to existing MIDAS specifications, resulting in more efficient estimation and improved forecast accuracy. Moreover, high-frequency information on both the predictors and the predicted variable is necessary to ensure efficiency.

These findings have important implications for researchers and practitioners using mixed-frequency models in economic forecasting, enhancing both theoretical understanding and practical implementation. Future research may extend these insights to nonlinear and high-dimensional settings, further exploring the benefits of efficient parameterization in mixed-frequency forecasting. By adopting eMIDAS, forecasters can achieve greater predictive accuracy with a more parsimonious model structure.

References

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–1289.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Amemiya, T. and Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, 67(339):628–632.

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2):246–261.

Andreou, E., Ghysels, E., and Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240–251.

Athanasopoulos, G., Hyndman, R. J., Song, H., and Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3):822–844.

Babii, A., Ghysels, E., and Striaukas, J. (2022). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106.

Benmoussa, A. A., Ellwanger, R., and Snudden, S. (2025). Carpe diem: Can daily oil prices improve model-based forecasts of the real price of crude oil? *International Journal of Forecasting*, https://doi.org/10.1016/j.ijforecast.2025.02.00.

Brewer, K. R. (1973). Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics*, 1(2):133–154.

Cavallo, A. and Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, 30(2):151–178.

Chan, J. C., Poon, A., and Zhu, D. (2025). Time-varying parameter MIDAS models: Application to nowcasting US real GDP. Available at SSRN: https://ssrn.com/abstract=4802887.

Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4):746–785.

Daniele, M., Mikosch, H., and Neuwirth, S. (2025). An observation-driven mixed-frequency VAR model with closed-form solution. Available at SSRN: https://ssrn.com/abstract=4776351.

De Jong, P. (1991). The diffuse kalman filter. *The Annals of Statistics*, 19(2):1073–1083.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press, 2nd edition.

Ellwanger, R. and Snudden, S. (2023). Forecasts of the real price of oil revisited: Do they beat the random walk? *Journal of Banking & Finance*, 154(106962):1–8.

Ferrara, L., Mogliani, M., and Sahuc, J.-G. (2022). High-frequency monitoring of growth at risk. *International Journal of Forecasting*, 38(2):582–595.

Foroni, C. and Marcellino, M. (2016). Mixed frequency structural vector auto-regressive models. *Journal of the Royal Statistical Society: Series A*, 179(2):403–425.

Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society. Series A*, 178(1):57–82.

Foroni, C., Marcellino, M., and Stevanovic, D. (2019). Mixed-frequency models with moving-average components. *Journal of Applied Econometrics*, 34(5):688–706.

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

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Harvey, A. C. (1989). *Forecasting, Structural Time Series and the Kalman Filter*. Cambridge University Press.

Ing, C.-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory*, 19(2):254–279.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics*, 18(3):337–349.

Kohns, D. and Potjagailo, G. (2025). Flexible bayesian MIDAS: Time variation, group-shrinkage and sparsity. *Journal of Business & Economic Statistics*, https://doi.org/10.1080/07350015.2025.2467898.

Kronenberg, P., Mikosch, H., Neuwirth, S., Bannert, M., and Thöni, S. (2023). The nowcasting lab: Live out-of-sample forecasting and model testing. Available at SSRN: https://ssrn.com/abstract=4353052.

Lütkepohl, H. (1986). Forecasting temporally aggregated vector ARMA processes. *Journal of Forecasting*, 5(2):85–95.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics*, 17(1):129–136.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135(1-2):499–526.

Martins, L. and Teles, P. (2025). The effects of temporal aggregation on MIDAS regressions. *Journal of Business & Economic Statistics*. https://doi.org/10.1080/07350015.2025.2507370.

McCarthy, M. and Snudden, S. (2024). Predictable by construction: Assessing forecast directional accuracy of temporal aggregates. *LCERPA Working Paper, 2024-7*.

Mogliani, M. and Simoni, A. (2021). Bayesian MIDAS penalized regressions: estimation, selection, and prediction. *Journal of Econometrics*, 222(1):833–860.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Pesaran, M. H. and Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485):325–337.

Silvestrini, A. and Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys*, 22(3):458–497.

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, 59(3):525–531.

Zellner, A. and Montmarquette, C. (1971). A study of some aspects of temporal aggregation problems in econometric analyses. *The Review of Economics and Statistics*, 53(4):335–342.

A Online Appendix

Data Appendix

All empirical results in the paper are based on publicly available series. No proprietary or restricted–access data are employed. The precise identifiers and download locations are listed below; the retrieval date is 12 March 2025.

U.S. 3-Month Treasury Constant-Maturity Rate

Daily secondary-market yields (percent p.a., not seasonally adjusted) are taken from the Federal Reserve Bank of St. Louis *FRED* database, series DGS3MO ("Market Yield on U.S. Treasury Securities at 3-Month Constant Maturity, Quoted on an Investment Basis"). Let $S_{t,i}$ denote the yield on trading day $i = 1, \ldots, n$ of month t. The period-average level is the simple average of daily values,

$$\bar{S}_t = \frac{1}{n} \sum_{i=1}^{n} S_{t,i},$$

while the end-of-month value is $S_{t,n}$. *FRED* releases are real-time and never revised.

Consumer Price Index (CPI)

Real-time headline CPI (seasonally adjusted, index 2015 = 100) is obtained from the Federal Reserve Bank of Philadelphia's Real-Time Data Research Center, vintage series PCPI. (Vintages "PCPI98M11–PCPI23M6")

The Bureau of Labor Statistics publishes the CPI with a one-month delay that is available around the middle of the subsequent month. We nowcast the missing observation with the historical average monthly growth rate; alternative nowcasting rules (last observed inflation rate; zero-change) leave all conclusions unchanged.

West Texas Intermediate Crude Oil

Daily WTI spot prices (US$ barrel) come from the U.S. Energy Information Administration's (EIA) *Petroleum and Other Liquids* portal, series RWTC ("Cushing, OK WTI Spot Price FOB (Dollars per Barrel)"). The *EIA* posts the series each business day with no subsequent revision.

U.S. Retail Gasoline Prices

Weekly retail gasoline prices are taken from Table 14 of the *EIA Weekly Petroleum Status Report* Sourcekey: EMM_EPMRU_PTE_NUS_DPG. We used the series "U.S. Regular Weekly U.S. Regular Conventional Retail Gasoline Prices (Dollars per Gallon)" for which the longest history is available. The EIA's weekly price corresponds to the price on the Monday of the week and is published on the same day, except on government holidays, for which the data is released on Tuesday, see Definitions, Sources and Explanatory Notes. This price is observable in real time. A monthly average of the series corresponds to the series for "Regular Motor Gasoline, All Areas, Retail Price" available in the pen-ultimate column in the EIA's Monthly Energy Review.