ISCTE IUL
REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2026-01-27

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Martins, L. & Teles, P. (2026). The effects of temporal aggregation on MIDAS regressions. Journal of Business and Economic Statistics. 44 (1), 120-133

Further information on publisher's website:
10.1080/07350015.2025.2507370

Publisher's copyright statement:
This is the peer reviewed version of the following article: Martins, L. & Teles, P. (2026). The effects of temporal aggregation on MIDAS regressions. Journal of Business and Economic Statistics. 44 (1), 120-133, which has been published in final form at https://dx.doi.org/10.1080/07350015.2025.2507370. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

The Effects of Temporal Aggregation on MIDAS Regressions

Luís Martins*[a] and Paulo Teles[b]

[a]ISCTE, Instituto Universitário de Lisboa, Business Research Unit (BRU-IUL), and CIMS-Univ. of Surrey

[b]School of Economics, University of Porto, and LIAAD-INESC Porto LA, Portugal

ABSTRACT

Time series used in practice are often temporal aggregates. A MIDAS regression model may be fitted wrongly assuming the independent variables are not aggregated when in fact they are. We derive the correct model's specification under temporal aggregation of the independent variables, thus identifying the correct number of dynamic terms and the order of the MA component. We also show that these depend on the frequency of the variables and the aggregation order used. As a result, the three alternative estimators considered are asymptotically biased and it is possible to rank them according to their bias in some cases. Our results are also confirmed by Monte Carlo experiments and are illustrated with an empirical application.

KEYWORDS

Aggregate model; Asymptotic bias; MIDAS regression; Overlapping sums; Temporal aggregation

*Corresponding author. E-mail address: luis.martins@iscte-iul.pt.

1 Introduction

A growing amount of economic data recorded at different frequencies is currently available. In a regression framework, the model equation may be set up at the lowest frequency of all the variables involved, with those observed at higher frequencies transformed to the lower frequency by temporal aggregation (flow variables) or systematic sampling (stock variables). Alternatively, an econometric model can be defined with all variables at their original, observed frequencies, even if some of them are already aggregates.

One of the most popular approaches to mitigate the frequency-mismatch problem without losing potentially relevant high-frequency information is the mixed data sampling (henceforth MIDAS) regression as proposed by Ghysels, Santa-Clara, and Valkanov (2004, 2005, 2006), Ghysels, Sinko, and Valkanov (2007), Ghysels, and Wright (2009) and Andreou, Ghysels, and Kourtellos (2010a, 2010b, 2013). In this model, the dependent variable is recorded at a lower frequency and, to ensure parsimony, this method aggregates the high-frequency data with particular polynomial functions, namely exponential Almon distributed lags or the Beta function. This transformation results in a regression problem whose coefficients are estimated by nonlinear least squares (NLS). MIDAS regressions have been applied in several fields such as macroeconomics, finance or forecasting. For example, see Clements and Galvão (2008, 2009), Armesto *et al.* (2010), Andreou *et al.* (2013), Ghysels (2016) or Foroni *et al.* (2013, 2015, 2018).

A MIDAS regression model with an autoregressive component of the dependent variable (or ARDL-MIDAS) has also been proposed, which introduces lagged values of that variable on the right-hand side of the equation (Ghysels *et al.* 2007, Clements and Galvão 2008, Armesto *et al.* 2010, Andreou *et al.* 2013, Foroni *et al.* 2018, Babii *et al.* 2022). More recently, a moving average component has been added to the ARDL-MIDAS model as well (Foroni *et al.* 2018). These extensions yielded gains in forecasting accuracy over the original MIDAS regression in some simulation studies and empirical applications but were not based on theoretical grounds, i.e., it has

not been shown they are superior or what the most appropriate model for mixed-frequency data is.

The motivation for this paper follows from the fact that so far we do not know the correct number of lags in the dynamic component of the ARDL-MIDAS model nor the order of the moving average specification. It is well known in the time series literature that misspecification of the dynamic part of a model induces bias in the estimation and erroneous inference. Consequently, the forecasts found for the variable of interest are not optimal. In this paper, we derive the correct specification of the ARDL-MIDAS model under aggregation. We show that the dynamic component includes rolling overlapping aggregates of the dependent variable and we find the exact number of such terms as well as the order of the moving average polynomial, both depending on the frequency and the order of aggregation introduced.

This contribution to the literature follows from applying to the MIDAS specification the definition and existing results on temporal aggregation in the usual time-series models. As previously mentioned, the values of a flow variable such as the Gross Domestic Product (GDP), sales, investment, stock log returns or rainfall amount are often available in practice through temporal aggregation of higher-frequency data and are therefore time series aggregates. For example, the annual GDP is the aggregate of monthly data, weekly log stock returns may be found by aggregating daily values. Literature on temporal aggregation of (univariate and multivariate) ARIMA models is vast, comprising model structure, parameter estimation, forecasting and causal relationships. Among others, this issue is discussed in Stram and Wei (1986), Lütkepohl (1987), Hotta and Neto (1993), Marcellino (1999), Breitung and Swanson (2002), Wei (2006, Chapter 20, a comprehensive review), Silvestrini and Veredas (2008), Lütkepohl (2009) or Mamingi (2017).

In fact, a MIDAS regression is defined with an aggregate dependent variable and assuming the independent variables are nonaggregated (or basic) time series. But, as previously mentioned, the only available data of many variables are time series aggregates, despite being the existing highest-frequency information. For instance, even though a monthly time series may be the highest-frequency data available, it may be the aggregate of weekly values, or weekly data may also be

the aggregates of daily values. Moreover, even when higher-frequency data are available, their (lower-frequency) aggregates are sometimes used in empirical analysis for a number of reasons. For instance, quarterly GDP may be the highest frequency available but it is obtained by the national statistical institutes through aggregation of the values within each quarter, some at a daily frequency and subject to constant revisions. Similarly, although daily stock returns are available, weekly or monthly values are often used by practitioners (as a matter of a fact, stock prices are nowadays observed every second). Consequently, a MIDAS model may be fitted wrongly assuming the data of the independent variables are not aggregated, i.e., it may be fitted ignoring that such data are in fact aggregates rather than the basic (nonaggregated) variables. The paper studies the effects of fitting a MIDAS regression with aggregated independent variables as if they were basic time series, i.e., wrongly assuming they are basic series when in fact they are aggregates.

After deriving the model with the aggregated variables and showing that the MIDAS model is not closed under aggregation, we study analytically and by means of Monte Carlo simulations the properties of three distinct estimators each associated to its own misspecified regression model: The FLAT-LS estimator where the independent variable shares the same frequency as the left-hand-side variable, the MIDAS-LS and the MIDAS-NLS estimators based on the original MIDAS model but where the latter assumes the usual exponential weights. Finally, an empirical application is provided as an illustration, where the quarterly US economic growth is explained by a component of the monthly index of coincident indicators or by different financial determinants.

The paper is organized as follows. In Section 2, we begin with the main result of the paper by deriving the correct MIDAS specification of a regression model where temporal aggregation was imposed on the independent variables. Section 3 provides the first and second-order moments of the aggregate variables and the expectations of their cross products implied by the MIDAS model. In Section 4, we define the FLAT-LS, MIDAS-LS, and MIDAS-NLS estimators, some general expressions of their asymptotic biases and variances, and closed-form expressions of the biases of the parameter of interest with low-order aggregation. Monte Carlo experiments are in

Section 5 and an empirical application in Section 6. We conclude in Section 7. Proofs are provided in the appendix in the Supplemental material.

2 Aggregate model

In this section we introduce one of the main results of the paper which is the derivation of the correct specification of a regression model whose independent variables have been subject to temporal aggregation. Such model includes lags of the dependent and independent variables with the number of terms depending on the frequency of the data and the order of aggregation undertaken.

2.1 The MIDAS model

Consider the time series Y_t^A observed at $t = 1, \ldots, T$ and $X_t^{(r)}$ observed r times between t and $t-1$, i.e., $X_{t-i/r}^{(r)}$ with $i = 0, \ldots, r-1$. They are respectively an aggregate (low frequency) and a basic (high frequency) variable. The MIDAS regression model proposed by Ghysels *et al.* (2004, 2006, 2007) and Andreou *et al.* (2010a, 2010b) is

$$Y_t^A = \alpha + \beta^* \sum_{i=0}^{r-1} \omega_i X_{t-i/r}^{(r)} + \varepsilon_t \tag{1}$$

where α, β^* are the regression parameters and ε_t is a martingale difference process with respect to the σ-field generated by $\left\{X_{t-i/r}^{(r)}, \varepsilon_{t-1-i} : i \geq 0\right\}$ and $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2 < \infty$. The lag coefficients $\omega_i = \lambda_i / \sum_{i=0}^{r-1} \lambda_i$ are the standardized weights, where λ_i are the unstandardized ones. Following Ghysels *et al.* (2004, 2005, 2006) and Andreou *et al.* (2010a, 2010b) among others, we assumed exponential weights $\lambda_i = \exp\left\{\theta_1 i + \theta_2 i^2\right\}$ $(i = 0, \ldots, r-1)$, where θ_1 and θ_2 are parameters. Sometimes, they have also been defined by using the Beta function (for instance, Ghysels *et al.* 2006, 2007 or Andreou *et al.* 2010b) or Legendre polynomials (Babii *et al.* 2022), but we considered exponential weights to further understand the properties of the well-known original MIDAS regression model of Ghysels *et al.* (2004, 2005) and because they are the most common. Considering unstandardized weights

first, model (1) becomes

$$Y_t^A = \alpha + \beta \sum_{i=0}^{r-1} \lambda_i X_{t-i/r}^{(r)} + \varepsilon_t, = \alpha + \beta \sum_{i=0}^{r-1} \lambda_i L^{i/r} X_t^{(r)} + \varepsilon_t = \alpha + \beta \prod_{i=1}^{r-1} \left(1 - \delta_i L^{1/r}\right) X_t^{(r)} + \varepsilon_t \quad (2)$$

where $\lambda_0 = 1$, which allows the identification of β, $L^{i/r}$ is the high-frequency lag operator such that $L^{i/r} X_t^{(r)} = X_{t-i/r}^{(r)}$ and δ_i^{-1} $(i = 1, \ldots, r-1)$ are the roots of the polynomial $\sum_{i=0}^{r-1} \lambda_i L^{i/r}$. Note that the pattern of the ω_i is given by the λ_i, i.e., they exhibit the same pattern, and that their respective polynomials in (1) and (2) share the same roots. Moreover, the slope parameters in (1) and (2) are such that $\beta = \beta^* / \sum_{i=0}^{r-1} \lambda_i$. For the sake of exposition, unstandardized weights were used in the analysis.

2.2 Model with aggregate independent variables

The first step of the analysis was to find the true regression model with an aggregate independent variable, because the process of data collection or the practice of researchers often aggregates the basic independent time series. Starting from (2), the next proposition derives the MIDAS model with an aggregate independent variable showing this model is not closed under temporal aggregation (the proof is provided in the Appendix).

Proposition 1 *Let Y_t^A and $X_t^{(r)}$ be an aggregate and a basic time series, respectively, as defined above, and let $X_t^a = \sum_{i=0}^{m-1} L^{i/r} X_t^{(r)}$ denote the m-period nonoverlapping aggregates of $X_t^{(r)}$, where the positive integer m is the order of aggregation ($m = 1$ means the basic data). For example, if $X_t^{(r)}$ is a monthly time series and $m = 3$, X_t^a is a quarterly series. Assume the data-generating process of Y_t^A is the MIDAS regression model defined in (2) such that $\delta_i^m = \delta_j^m$ if and only if $\delta_i = \delta_j$. Then, the regression model with X_t^a as independent variable is*

$$Y_t^A = m\alpha \prod_{i=1}^{r-1} \frac{1 - \delta_i^m}{1 - \delta_i} - \sum_{i=1}^{r(m-1)} \Lambda_i Y_{t-i/r}^A + \beta \prod_{i=1}^{r-1} \left(1 - \delta_i^m L^{m/r}\right) X_t^a + \sum_{i=0}^{r(m-1)} \Lambda_i \varepsilon_{t-i/r} \quad (3)$$

where Y_t^A is assumed to be the r-aggregate of a basic time series $Y_t^{(r)}$, i.e., $Y_t^A = \sum_{i=0}^{r-1} L^{i/r} Y_t^{(r)}$,

$$\prod_{i=1}^{r-1} \left(1 - \delta_i^m L^{m/r}\right) = 1 + \sum_{i=1}^{r-1} (-1)^i \sum_{s_1=1}^{r-i} \sum_{s_2=s_1+1}^{r-(i-1)} \sum_{s_3=s_2+1}^{r-(i-2)} \cdots \sum_{s_i=s_{i-1}+1}^{r-1} \prod_{h=1}^{i} \delta_{s_h}^m L^{m/r} = \sum_{i=0}^{r-1} \lambda_i^a, \quad (4)$$

where λ_i^a *are the weights of* $X_{t-mi/r}^a$ $(\lambda_0^a = 1)$*, and*

$$\Lambda_i = \begin{cases} 1 & \text{if} \quad i = 0 \\ 1 + \sum_{j=1}^{i} \sum_{s_1=1}^{r-1} \sum_{s_2=s_1}^{r-1} \cdots \sum_{s_j=s_{j-1}}^{r-1} \prod_{h=1}^{j} \delta_{s_h} & \text{if} \quad i = 1, \ldots, m-1 \\ \sum_{j=i-(m-1)}^{i} \sum_{s_1=1}^{r-1} \sum_{s_2=s_1}^{r-1} \cdots \sum_{s_j=s_{j-1}}^{r-1} \prod_{h=1}^{j} \delta_{s_h} & \text{if} \quad i = m, \ldots, (r-1)(m-1) \\ \sum_{j=i-(m-1)}^{(r-1)(m-1)} \sum_{s_1=1}^{r-1} \sum_{s_2=s_1}^{r-1} \cdots \sum_{s_j=s_{j-1}}^{r-1} \prod_{h=1}^{j} \delta_{s_h} & \text{if} \quad i = (r-1)(m-1)+1, \ldots, r(m-1), \end{cases} \tag{5}$$

where the resulting powers $\delta_{s_h}^u$ *are such that* $u = min(j, m-1)$ *for* $i = m, \ldots, r(m-1)$*, i.e., the largest possible power is* $\delta_{s_h}^{m-1}$*. Concerning the error term and assuming that* $\varepsilon_t = \sum_{j=0}^{r-1} \epsilon_{t-j/r}$*, where* ϵ_t *denotes a martingale difference sequence with* $E(\epsilon_t^2) = \sigma_\epsilon^2 < \infty$*, we have*

$$\sum_{i=0}^{r(m-1)} \Lambda_i \varepsilon_{t-i/r} = \sum_{i=0}^{r-1} \sum_{j=0}^{i} \Lambda_j \epsilon_{t-i/r} + \sum_{i=r}^{r(m-1)} \sum_{j=i-(r-1)}^{i} \Lambda_j \epsilon_{t-i/r} + \sum_{i=r(m-1)+1}^{rm-1} \sum_{j=i-(r-1)}^{r(m-1)} \Lambda_j \epsilon_{t-i/r} \quad (\Lambda_0 = 1) \tag{6}$$

which, in terms of the aggregate time unit t*, is a moving average of order* $m-1$ *and can be written as* $\xi_t - \sum_{i=1}^{m-1} \Theta_i \xi_{t-i}$ *where* ξ_t *is a white noise with mean zero and variance* σ_ξ^2*. The parameters* Θ_i $(i = 1, \ldots, m-1)$ *and* σ_ξ^2 *are functions of the* δ_j $(j = 1, \ldots, r-1)$ *and* σ_ε^2*.*

Examples of the aggregate model for some r and m are provided in the Supplemental material. Equation (3) shows the aggregate model is not a MIDAS model. Thus, the class of MIDAS regression models is not closed under temporal aggregation of the independent variable. Such conclusion is also true for (strong) GARCH models, for example (Drost and Nijman 1993), unlike ARIMA models which are closed (Wei 2006, Chapter 20). Furthermore, (3) also shows that aggregation of the independent variable introduces lagged overlapping values of the dependent variable, which are not usually available, in the right-hand side of the equation. Recall that the right-hand side of the ARDL-MIDAS model includes lagged nonoverlapping values of the latter variable and, since these are the best proxies for its lagged overlapping values, this model is closer to the true one when the independent variable is aggregated. Similarly, a moving average component in the equation is also

required, as shown by (3). This helps to explain why the ARDL-MIDAS model with or without an MA component outperformed the standard MIDAS regression in forecasting, as previously mentioned (Ghysels *et al.* 2007, Clements and Galvão 2008, Armesto *et al.* 2010, Andreou *et al.* 2013, Foroni *et al.* 2018, Babii *et al.* 2022).

We note this is not the framework discussed in Foroni *et al.* (2015) since their basic regression is derived by assuming a VAR model with the same frequency both for the dependent and the independent variables. Then, unlike our approach, only the former variable is aggregated in order to find a MIDAS regression model they denoted by U-MIDAS (or unrestricted MIDAS) with different parameters and number of lags. In this paper we take the more general aggregated specification which was never before considered in the literature.

The prior data-generating process (2) can be generalized by including several independent variables,

$$\begin{aligned} Y_t^A &= \alpha + \beta_1 \sum_{i=0}^{r-1} \lambda_{1,i} X_{1,t-i/r}^{(r)} + \beta_2 \sum_{i=0}^{r-1} \lambda_{2,i} X_{2,t-i/r}^{(r)} + \ldots + \beta_p \sum_{i=0}^{r-1} \lambda_{p,i} X_{p,t-i/r}^{(r)} + \varepsilon_t \\ &= \alpha + \sum_{u=1}^{p} \beta_u \sum_{i=0}^{r-1} \lambda_{u,i} L^{i/r} X_{u,t}^{(r)} + \varepsilon_t = \alpha + \sum_{u=1}^{p} \beta_u \prod_{i=1}^{r-1} \left(1 - \delta_{u,i} L^{1/r}\right) X_{u,t}^{(r)} + \varepsilon_t, \end{aligned} \tag{7}$$

where $\lambda_{u,i}$ are the unstandardized weights of $X_{u,t}^{(r)}$ $(u = 1, \ldots, p;\ i = 0, \ldots, r-1)$, with $\lambda_{u,0} = 1$, and $\delta_{u,i}^{-1}$ are the roots of $\sum_{i=0}^{r-1} \lambda_{u,i} L^{i/r}$. Then, the aggregate model is (details provided in the Appendix)

$$\begin{aligned} Y_t^A &= m\alpha \prod_{j=1}^{p} \prod_{i=1}^{r-1} \frac{1 - \delta_{j,i}^m}{1 - \delta_{j,i}} - \sum_{i=1}^{[p(r-1)+1](m-1)} \Lambda_{p_i} Y_{t-i/r}^A + \sum_{u=1}^{p} \beta_u \frac{\prod_{j=1}^{p} \prod_{i=1}^{r-1} \left(1 - \delta_{j,i}^m L^{m/r}\right)}{\prod_{\substack{j=1 \\ j \neq u}}^{p} \prod_{i=1}^{r-1} \left(1 - \delta_{j,i} L^{1/r}\right)} X_{u,t}^a \\ &+ \sum_{i=0}^{[p(r-1)+1](m-1)} \Lambda_{p_i} \varepsilon_{t-i/r} \end{aligned}$$

which again proves that the class of MIDAS regression models is not closed under temporal aggregation of the independent variables. The above equation shows that the aggregate model with multiple independent variables leads to cumbersome expressions, unlike the case of a single variable. Consequently, estimation of the latter case was studied in detail in order to find tractable, closed-form expressions.

3 Moments with aggregate variables

The following results concerning estimation of a MIDAS regression model assume that the (higher-frequency) independent variable is a stationary ARMA$(1,1)$ process which is more general than the cases considered in the literature (such as Andreou *et al.* 2010a, 2010b). For completeness, we derived the required raw higher-order moments of the basic $X_t^{(r)}$ in the Supplemental material since, to the best of our knowledge, such moments of a stationary ARMA$(1,1)$ process have not been derived yet. Using them, we also added to the existing literature the expectations of the cross products of the aggregate X_t^a with Y_t^A implied by the MIDAS model described in the previous Section. To this purpose, the following assumption is required first.

Assumption 1 *Let $X_t^{(r)}$ be the ARMA(1,1) process $X_t^{(r)} = \phi_0 + \phi_1 X_{t-1/r}^{(r)} + \nu_t - \omega_1 \nu_{t-1/r}$, where $\phi_1 \neq \omega_1$, $|\phi_1| < 1$, $|\omega_1| < 1$ for stationarity and invertibility respectively and ν_t is a white noise process with mean zero, $E(\nu_t^2) = \sigma_\nu^2$ and $E(\nu_t^k) = \mu_{k_\nu} < \infty \, (k = 3, 4)$.*

Let the m-period nonoverlapping aggregates of $X_t^{(r)}$ be $X_t^a = \sum_{i=0}^{m-1} L^{i/r} X_t^{(r)} = \sum_{i=0}^{m-1} X_{t-i/r}^{(r)}$, with $m \leq r$. It is well known that X_t^a is also an ARMA(1,1) model (Wei 2006, Chapter 20) and it can be shown that it has the moving average representation

$$X_t^a = \frac{m\phi_0}{1-\phi_1} + \sum_{i=0}^{m-1} \left[1 + (\phi_1 - \omega_1) \frac{1-\phi_1^i}{1-\phi_1} \right] \nu_{t-i/r} + (\phi_1 - \omega_1) \frac{1-\phi_1^m}{1-\phi_1} \sum_{i=0}^{\infty} \phi_1^i \nu_{t-(i+m)/r} \tag{8}$$

so that

$$E(X_t^a) = \frac{m\phi_0}{1-\phi_1}. \tag{9}$$

Some moments of X_t^a that will be required later are derived next as functions of the parameters of interest ϕ_0, ϕ_1, ω_1 and σ_ν^2. Furthermore, if $m = r$ (full aggregation of $X_t^{(r)}$), $X_t^A = \sum_{i=0}^{r-1} X_{t-i/r}^{(r)}$. Then, X_t^A and its expected value can be written as (8) and (9) respectively replacing m by r. All the other expectations, namely the raw moments between the aggregate independent variable and the (aggregate) dependent variable, are provided in the next proposition and corollaries (the proof can be found in the appendix).

Proposition 2 *Let Y_t^A and $X_t^{(r)}$ be an aggregate and a basic time series respectively such that the latter is observed r times between t and $t-1$, i.e., $X_{t-i/r}^{(r)}$ with $i=0,\ldots,r-1$. Suppose also that Assumption 1 and the definitions thereafter hold. Furthermore, consider the MIDAS regression model given in (2). Thus,*

$$\begin{aligned} E\left[\left(X_t^a\right)^2\right] &= \frac{1+\omega_1^2-2\phi_1\omega_1}{1-\phi_1^2}m\sigma_\nu^2+\left(\frac{m\phi_0}{1-\phi_1}\right)^2 \\ &+ \frac{\left(1-\phi_1\omega_1\right)\left(\phi_1-\omega_1\right)}{1-\phi_1^2}\frac{2\sigma_\nu^2}{1-\phi_1}\left(m-1-\phi_1\frac{1-\phi_1^{m-1}}{1-\phi_1}\right) \qquad (10) \\ E\left(Y_t^A\right) &= \alpha+\frac{\beta\phi_0}{1-\phi_1}\sum_{j=0}^{r-1}(-1)^j\sum_{s_1=1}^{r-j}\sum_{s_2=s_1+1}^{r-(j-1)}\cdots\sum_{s_j=s_{j-1}+1}^{r-1}\prod_{h=1}^{j}\delta_{s_h} \qquad (11) \\ E\left(X_t^aX_{t-im/r}^a\right) &= \left(\frac{m\phi_0}{1-\phi_1}\right)^2+\frac{\left(1-\phi_1\omega_1\right)\left(\phi_1-\omega_1\right)}{1-\phi_1^2}\sigma_\nu^2\left(\frac{1-\phi_1^m}{1-\phi_1}\right)^2\phi_1^{(i-1)m} \qquad (12) \end{aligned}$$

with $i\geq 1$, and

$$E\left(X_t^aY_{t-i/r}^A\right)=m\frac{\alpha\phi_0}{1-\phi_1}+\beta\sum_{j=0}^{r-1}(-1)^j\sum_{s_1=1}^{r-j}\sum_{s_2=s_1+1}^{r-(j-1)}\cdots\sum_{s_j=s_{j-1}+1}^{r-1}\prod_{h=1}^{j}\delta_{s_h}E\left(X_{t-(i+j)/r}^{(r)}\sum_{u=0}^{m-1}X_{t-u/r}^{(r)}\right). \qquad (13)$$

Furthermore, with $u=1,2,\ldots$ and $r\geq(u+1)m$,

$$\begin{aligned} E\left(X_{t-um/r}^aY_{t-i/r}^A\right) &= E\left[\sum_{v=0}^{m-1}X_{t-(um+v)/r}^{(r)}\left(\alpha+\beta\prod_{j=1}^{r-1}\left(1-\delta_jL^{1/r}\right)X_{t-i/r}^{(r)}+\varepsilon_{t-i/r}\right)\right] \\ &= m\frac{\alpha\phi_0}{1-\phi_1}+\beta\sum_{j=0}^{r-1}(-1)^j\sum_{s_1=1}^{r-j}\sum_{s_2=s_1+1}^{r-(j-1)}\cdots\sum_{s_j=s_{j-1}+1}^{r-1}\prod_{h=1}^{j}\delta_{s_h} \\ &\times E\left(X_{t-(i+j)/r}^{(r)}\sum_{v=0}^{m-1}X_{t-(um+v)/r}^{(r)}\right). \qquad (14) \end{aligned}$$

Corollary 1 shows (13) and (14) according to the values of $i=1,\ldots,r\left(m-1\right)$ and Corollary 2 shows the expectations in Proposition 2 when $m=r$, i.e., for X_t^A (both corollaries and their proofs are in the appendix).

4 FLAT-LS, MIDAS-LS, and MIDAS-NLS estimation

In the previous sections we derived the true model's specification for aggregated variables and found the aggregate moments when the basic regressor follows an ARMA(1,1) process. The properties

of the estimators associated to the three different misspecified models of interest are analytically compared next.

4.1 The estimators and their asymptotic bias and variance

Assuming the true model with X_t^a is given by (3) and that $X_t^{(r)}$ follows a stationary ARMA(1,1) process, we compared three distinct estimation methods in terms of their asymptotic bias (ABIAS) and variance (AVAR), each based on a misspecified regression model: the flat least squares estimator (FLAT-LS), the MIDAS least squares estimator (MIDAS-LS), and the MIDAS nonlinear least squares estimator with exponential weights (MIDAS-NLS).

Regarding the FLAT-LS estimator, the regression model is

$$Y_t^A = \alpha_1^* + \beta X_t^A + U_{FLAT,t} = \pi' \widetilde{X}_t^A + U_{FLAT,t},$$

where $\pi = (\alpha_1^*, \beta)'$ and $\widetilde{X}_t^A \equiv \widetilde{X}_{t,r}^A = \left(1, X_{t,r}^A\right)'$, which is the full-aggregated $X_t^A \equiv X_{t,r}^A = \sum_{i=0}^{r-1} X_{t-i/r}^{(r)}$, such that $R = r/m = 1$. In fact, the FLAT-LS regression model, also known as bridge model (Foroni and Marcellino 2013, 2014, Schumacher 2016), is nothing but the original MIDAS model with weights equal to $1/r$ in (1) or equal to 1 and $\beta = \beta^*/r$ in (2). The corresponding estimator of π is the sample minimizer of the LS criterion whose probability limit is

$$\pi_{FLAT-LS} = \left(E\left(\widetilde{X}_t^A \widetilde{X}_t^{A\prime}\right)\right)^{-1} E\left(\widetilde{X}_t^A Y_t^A\right), \tag{15}$$

under the usual assumptions (see, for example, Andreou *et al.*, 2010a, 2010b). For the MIDAS-LS estimator,

$$Y_t^A = \alpha_2^* + \beta X_t^a + \psi_{\beta,1} X_{t-m/r}^a + ... + \psi_{\beta,R-1} X_{t-(R-1)m/r}^a + U_{MLS,t} = \pi' \widetilde{X}_t^a + U_{MLS,t},$$

where $\pi = (\alpha_2^*, \beta, \psi_{\beta,1}, \ldots, \psi_{\beta,R-1})'$, such that the integer $R = r/m > 1$ and $\psi_{\beta,i} = \beta\psi_i$ for some constants ψ_i $(i = 1, ..., R-1)$ and $\widetilde{X}_t^a \equiv \widetilde{X}_{t,m,r}^a = \left(1, X_t^a, X_{t-m/r}^a, ..., X_{t-(R-1)m/r}^a\right)'$. The probability limit of the estimator takes the form

$$\pi_{MIDAS-LS} = \left(E\left(\widetilde{X}_t^a \widetilde{X}_t^{a\prime}\right)\right)^{-1} E\left(\widetilde{X}_t^a Y_t^A\right). \tag{16}$$

Note that the relationship between X_t^A and X_t^a is given by

$$X_t^A = \sum_{i=0}^{r-1} X_{t-i/r}^{(r)} = \sum_{i=0}^{R-1} \sum_{j=mi}^{m(i+1)-1} X_{t-j/r}^{(r)} = \sum_{i=0}^{R-1} X_{t-im/r}^a.$$

Finally, the MIDAS-NLS estimator assuming exponential weights follows from the model

$$Y_t^A = h\left(X_t^a, X_{t-m/r}^a, ..., X_{t-(R-1)m/r}^a; \pi\right) + U_{MNLS,t} = \alpha_3^* + \beta X_t^a\left(\theta^*\right) + U_{MNLS,t},$$

where $\pi = (\alpha_3^*, \beta, \theta^{*\prime})' = (\alpha_3^*, \beta, \theta_1^*, \theta_2^*)'$ and, for $R > 1$,

$$X_t^a\left(\theta^*\right) \equiv X_{t,m,r}^a\left(\theta^*\right) = X_t^a + \lambda_1^{a^*} X_{t-m/r}^a + ... + \lambda_{R-1}^{a^*} X_{t-(R-1)m/r}^a$$

such that $\lambda_i^{a^*} = \exp\left(\theta_1^* i + \theta_2^* i^2\right), i = 1, ..., R-1$. The nonlinear LS estimator is defined as

$$\widehat{\pi}_{MIDAS-NLS} = \arg\min_{\pi \in \Pi} \sum_{t=1}^{T} \left[Y_t^A - h\left(X_t^a, X_{t-m/r}^a, ..., X_{t-(R-1)m/r}^a; \pi\right)\right]^2.$$

Unlike the previous two estimators, there is no closed-form expression for $\widehat{\pi}_{MIDAS-NLS}$ because the solution for the minimization problem follows from non-linear first order conditions associated to θ^*. If we take the optimal θ^* as given,

$$\widehat{\beta}_{MIDAS-NLS}\left(\theta^*\right) = \frac{\frac{1}{T}\sum_{t=1}^{T} Y_t^A X_t^a\left(\theta^*\right) - \left(\frac{1}{T}\sum_{t=1}^{T} Y_t^A\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t^a\left(\theta^*\right)\right)}{\frac{1}{T}\sum_{t=1}^{T} X_t^a\left(\theta^*\right)^2 - \left[\frac{1}{T}\sum_{t=1}^{T} X_t^a\left(\theta^*\right)\right]^2}$$

which, as $T \to \infty$, converges in probability to

$$\beta_{MIDAS-NLS}\left(\theta^*\right) = \frac{Cov\left(Y_t^A, X_t^a\left(\theta^*\right)\right)}{Var\left(X_t^a\left(\theta^*\right)\right)} = \frac{Cov\left(Y_t^A, X_t^a + \sum_{i=1}^{R-1} \lambda_i^{a^*} X_{t-im/r}^a\right)}{Var\left(X_t^a + \sum_{i=1}^{R-1} \lambda_i^{a^*} X_{t-im/r}^a\right)}.$$

All three estimators are biased because of the omission of relevant regressors. Notice that the true model given by (3) can be written as

$$Y_t^A = \pi_{0,a}' \widetilde{X}_t^a + \pi_{0,z}' Z_t^a + \xi_t - \Theta_1 \xi_{t-1} - \ldots - \Theta_{m-1} \xi_{t-(m-1)},$$

where, for $R > 1$,

$$\pi_{0,a} \equiv \pi_{0,a,m,r} = \left(\alpha^*, \beta, \beta\lambda_1^a, \ldots, \beta\lambda_{(R-1)}^a\right)',$$

$$\pi_{0,z} \equiv \pi_{0,z,m,r} = \left(-\Lambda_1, -\Lambda_2, \ldots, -\Lambda_{r(m-1)}, \beta\lambda_R^a, \ldots, \beta\lambda_{r-1}^a\right)'$$

and

$$Z_t^a \equiv Z_{t,m,r}^a = \left(Y_{t-1/r}^A, Y_{t-2/r}^A, ..., Y_{t-r(m-1)/r}^A, X_{t-Rm/r}^a, ..., X_{t-(r-1)m/r}^a\right)',$$

with $\alpha^* \equiv \alpha_{m,r}^* = m\alpha \prod_{i=1}^{r-1} \frac{1-\delta_i^m}{1-\delta_i}$ (a function of α and δ_i), $\Lambda_j \equiv \Lambda_{j,m,r}$ and $\lambda_h^a \equiv \lambda_{h,m,r}^a$ (both functions of the δ_i), or, equivalently,

$$Y_t^A = \pi_{0,A}' \widetilde{X}_t^A + \pi_{0,\tilde{z}}' \widetilde{Z}_t^a + \xi_t - \Theta_1 \xi_{t-1} - \ldots - \Theta_{m-1} \xi_{t-(m-1)},$$

where, instead,

$$\pi_{0,A} = (\alpha^*, \beta)',$$

$$\pi_{0,\tilde{z}} \equiv \pi_{0,\tilde{z},m,r} = \left(-\Lambda_1, -\Lambda_2, \ldots, -\Lambda_{r(m-1)}, \beta(\lambda_1^a - 1), \ldots, \beta(\lambda_{(R-1)}^a - 1), \beta\lambda_R^a, \ldots, \beta\lambda_{r-1}^a\right)'$$

and

$$\widetilde{Z}_t^a \equiv \widetilde{Z}_{t,m,r}^a = \left(Y_{t-1/r}^A, Y_{t-2/r}^A, ..., Y_{t-r(m-1)/r}^A, X_{t-m/r}^a, ..., X_{t-(R-1)m/r}^a, X_{t-Rm/r}^a, ..., X_{t-(r-1)m/r}^a\right)'.$$

An additional source of BIAS is parameter misspecification. Although β is not misspecified in the regression models, the intercept is because in general it is not equal to α^*. The same happens to ψ_i $(i = 1, ..., R-1)$:

$$Y_t^A = \pi' \widetilde{X}_t^a + (\pi_{0,a} - \pi)' \widetilde{X}_t^a + \pi_{0,z}' Z_t^a + \xi_t - \Theta_1 \xi_{t-1} - ... - \Theta_{m-1} \xi_{t-(m-1)}. \tag{17}$$

The MIDAS-NLS suffers from the additional misspecification of assuming the exponential parametric form of their weights. In fact, the true MIDAS-NLS coefficient is

$$\pi_{0,MNLS} \neq \arg\min_{\pi \in \Pi} E\left[Y_t^A - h\left(X_t^a, X_{t-m/r}^a, ..., X_{t-(R-1)m/r}^a; \pi\right)\right]^2$$

because when exponential weights are assumed for the independent variable in the basic regression model, such weighting scheme may not be valid in the aggregate model. As an example, consider model (2) with exponential weights and $r = 4, m = 2$, i.e., $\lambda_i = \exp\{\theta_1 i + \theta_2 i^2\}$ $(i = 0, 1, 2, 3)$.

Then, from (2),

$$\lambda_0=1;\lambda_1=-\sum_{i=1}^{3}\delta_i=\exp(\theta_1+\theta_2);\lambda_2=\sum_{i=1}^{2}\sum_{j=i+1}^{3}\delta_i\delta_j=\exp[2\,(\theta_1+2\theta_2)];\lambda_3=-\prod_{i=1}^{3}\delta_i=\exp[3\,(\theta_1+3\theta_2)].$$

Furthermore, the weights of X_t^a in the aggregate model (3), given in (4) and in the Supplemental material, are such that

$$\begin{aligned}
\lambda_0^a &= 1;\\
\lambda_1^a &= -\sum_{i=1}^{3}\delta_i^2=-\left[\exp\left[2\left(\theta_1+\theta_2\right)\right]-2\exp\left[2\left(\theta_1+2\theta_2\right)\right]\right]=\exp\left[2\left(\theta_1+\theta_2\right)\right]\left[2\exp\left(2\theta_2\right)-1\right]\\
&\neq \exp\left(\theta_1^*+\theta_2^*\right)\\
\lambda_2^a &= \sum_{i=1}^{2}\sum_{j=i+1}^{3}\delta_i^2\delta_j^2=\left(\sum_{i=1}^{2}\sum_{j=i+1}^{3}\delta_i\delta_j\right)^2-2\left(\delta_1^2\delta_2\delta_3+\delta_1\delta_2^2\delta_3+\delta_1\delta_2\delta_3^2\right)=\exp\left[2\left(2\theta_1+4\theta_2\right)\right]\\
&\quad -2\exp\left(\theta_1+\theta_2\right)\exp\left[3\left(\theta_1+3\theta_2\right)\right]=\exp\left[2\left(2\theta_1+4\theta_2\right)\right]\left[1-2\exp\left(2\theta_2\right)\right]\neq\exp\left[2\left(\theta_1^*+2\theta_2^*\right)\right]\\
\lambda_3^a &= -\prod_{i=1}^{3}\delta_i^2=-\exp\left[2\left(3\theta_1+9\theta_2\right)\right]\neq\exp\left[3\left(\theta_1^*+3\theta_2^*\right)\right].
\end{aligned}$$

These equations show it is not possible to write all the weights λ_i^a of the aggregate model in exponential form, although the basic weights λ_i have such a form. In other words, the choice of exponential-type weights in the aggregate model is questionable and induces additional misspecification in the model. This is not specific of the simplest case we considered of $r=4, m=2$. Another example is provided in the Supplemental material.

By plugging Y_t^A in (15), we find the asymptotic bias for the FLAT-LS estimator,

$$ABIAS\left(\widehat{\pi}_{FLAT-LS};\pi_{0,A},\pi,\pi_{0,\widetilde{z}}\right)=\left(\pi_{0,A}-\pi\right)+\left(E\left(\widetilde{X}_t^A\widetilde{X}_t^{A\prime}\right)\right)^{-1}E\left(\widetilde{X}_t^A\widetilde{Z}_t^{a\prime}\right)\pi_{0,\widetilde{z}}$$

and, if $\alpha_1^*=\alpha^*$,

$$ABIAS\left(\widehat{\pi}_{FLAT-LS};\pi_{0,\widetilde{z}}\right)=\left(E\left(\widetilde{X}_t^A\widetilde{X}_t^{A\prime}\right)\right)^{-1}E\left(\widetilde{X}_t^A\widetilde{Z}_t^{a\prime}\right)\pi_{0,\widetilde{z}}. \tag{18}$$

Similarly,

$$ABIAS\left(\widehat{\pi}_{MIDAS-LS};\pi_{0,a},\pi,\pi_{0,z}\right)=\left(\pi_{0,a}-\pi\right)+\left(E\left(\widetilde{X}_t^a\widetilde{X}_t^{a\prime}\right)\right)^{-1}E\left(\widetilde{X}_t^aZ_t^{a\prime}\right)\pi_{0,z}$$

and, if $\pi_{0,a} = \pi$,

$$ABIAS\left(\widehat{\pi}_{MIDAS-LS};\pi_{0,z}\right) = \left(E\left(\widetilde{X}_t^a \widetilde{X}_t^{a\prime}\right)\right)^{-1} E\left(\widetilde{X}_t^a Z_t^{a\prime}\right)\pi_{0,z}. \tag{19}$$

The asymptotic bias of these two estimators is easily found from standard matrix calculus using the moments provided in Section 3. The matrices involved are shown in the Supplemental material.

The general ABIAS for the MIDAS-NLS follows from the fact that the criterion function of this extremum estimator does converge in probability, under the usual conditions, to the true expected value that defines $\pi_{0,MNLS}$. For the particular case of β,

$$\begin{aligned}\beta_{MIDAS-NLS}\left(\theta^*\right) &= \beta\frac{Cov\left(\left(X_t^a + \sum_{i=1}^{R-1}\lambda_i^a X_{t-im/r}^a\right), \left(X_t^a + \sum_{i=1}^{R-1}\lambda_i^{a^*} X_{t-im/r}^a\right)\right)}{Var\left(X_t^a\left(\theta^*\right)\right)} \\ &+ \frac{Cov\left(\pi_{0,z}' Z_t^a, X_t^a\left(\theta^*\right)\right)}{Var\left(X_t^a\left(\theta^*\right)\right)} \neq \beta + \frac{Cov\left(\pi_{0,z}' Z_t^a, X_t^a\left(\theta^*\right)\right)}{Var\left(X_t^a\left(\theta^*\right)\right)},\end{aligned}$$

because $\lambda_i^a \neq \lambda_i^{a^*}$, and we have

$$\begin{aligned}&ABIAS\left(\widehat{\beta}_{MIDAS-NLS};\pi_{0,z},\theta^*\right) = \beta_{MIDAS-NLS}\left(\theta^*\right) - \beta \\ &= \beta\left[\frac{Cov\left(\left(X_t^a + \sum_{i=1}^{R-1}\lambda_i^a X_{t-im/r}^a\right), \left(X_t^a + \sum_{i=1}^{R-1}\lambda_i^{a^*} X_{t-im/r}^a\right)\right)}{Var\left(X_t^a\left(\theta^*\right)\right)} - 1\right] + \frac{Cov\left(\pi_{0,z}' Z_t^a, X_t^a\left(\theta^*\right)\right)}{Var\left(X_t^a\left(\theta^*\right)\right)}.\end{aligned}$$

These results show that, besides the bias due to the omitted regressors, the MIDAS-NLS estimator suffers from the additional bias caused by the fact that the exponential weights are not valid in the aggregate model, as shown earlier.

The bias of the estimators can be derived by using standard calculus, namely after inverting a matrix of moments. Then, the exact closed-form expressions can be found for simple cases, given next and in the section on Monte Carlo simulations. On the contrary, the variances are extremely complicated, making cumbersome any effort to derive their detailed formulas. We provide some results for the variances based on Monte Carlo experiments and show in the Supplemental material some intermediate calculations required in the derivation of the final analytical result.

4.2 The simplest cases of $r = m = 2$ and $r = 4, m = 2$

The amount of bias and variance of each estimator can be obtained by Monte Carlo simulations for any r, m and the bias may also be derived analytically from the previous general expressions using the simplest cases of $r = m = 2$ and $r = 4, m = 2$.

Consider the case of $r = m = 2$ as a toy example because the FLAT-LS and MIDAS-LS estimators coincide and the MIDAS-NLS is not defined. After some tedious but straightforward calculations, shown in the Supplemental material, we conclude that

$$ABIAS\left(\widehat{\pi}_{FLAT-LS};\alpha,\beta,\delta_1\right) = \begin{pmatrix} -\left(1+2\delta_1\right)\alpha \\ -\frac{1}{2}\left(1+\delta_1\right)\beta \end{pmatrix}.$$

In particular, $ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right) = \frac{1}{2}\beta\left(-\delta_1 - 1\right)$, where $-\delta_1^2 = \lambda_1^a$. Notice that the ABIAS for $r = m = 2$ does not depend on the persistence of the basic $X_t^{(r)}$.

Next, we study the properties of the estimators for $r = 4, m = 2$. The following Proposition presents the closed-form expressions of the biases.

Proposition 3 *Assume that the conditions in Proposition 2 hold for $r = 4, m = 2$ and define the ABIAS of the FLAT-LS and MIDAS-LS estimators as in (18) and (19), respectively. Denoting the autocorrelation function of X_t^a by ρ,*

$$\begin{aligned} ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right) = & \frac{1}{2Var\left(X_t^a\right)\left[1+\rho\left(X_t^a, X_{t-2/4}^a\right)\right]} \\ \times & \left\{ -2\left[-\sum_{i=1}^{4}\Lambda_i E\left(Y_t^A\right)E\left(X_t^a\right) + \beta\sum_{i=2}^{3}\lambda_i^a\left[E\left(X_t^a\right)\right]^2\right]\right. \\ - & \left[\sum_{i=1}^{4}\Lambda_i E\left(X_t^a Y_{t-i/4}^A\right) - \beta\sum_{i=2}^{3}\lambda_i^a E\left(X_t^a X_{t-(2i)/4}^a\right)\right]\frac{2}{1-\rho\left(X_t^a, X_{t-2/4}^a\right)} \\ - & \left.\left[\sum_{i=1}^{4}\Lambda_i E\left(X_{t-2/4}^a Y_{t-i/4}^A\right) - \beta\sum_{i=2}^{3}\lambda_i^a E\left(X_{t-2/4}^a X_{t-(2i)/4}^a\right)\right]\frac{2}{1-\frac{1}{\rho\left(X_t^a, X_{t-2/4}^a\right)}}\right\} \end{aligned}$$

and

$$\begin{aligned} ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right) &= \frac{1}{2Var\left(X_t^a\right)\left[1+\rho\left(X_t^a,X_{t-2/4}^a\right)\right]} \\ &\times \left\{-2\left[-\sum_{i=1}^{4}\Lambda_i E\left(Y_t^A\right)E\left(X_t^a\right)+\beta\sum_{i=2}^{3}\lambda_i^a\left[E\left(X_t^a\right)\right]^2\right]\right. \\ &- \left[\sum_{i=1}^{4}\Lambda_i E\left(X_t^a Y_{t-i/4}^A\right)-\beta\sum_{i=2}^{3}\lambda_i^a E\left(X_t^a X_{t-(2i)/4}^a\right)\right] \\ &- \left[\sum_{i=1}^{4}\Lambda_i E\left(X_{t-2/4}^a Y_{t-i/4}^A\right)-\beta\sum_{i=2}^{3}\lambda_i^a E\left(X_{t-2/4}^a X_{t-(2i)/4}^a\right)\right] \\ &+ \left.\beta\left(\lambda_1^a-1\right)\left[Var\left(X_t^a\right)+Cov\left(X_t^a,X_{t-2/4}^a\right)\right]\right\}, \end{aligned}$$

so that,

$$ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right)-ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right)=\left[\sum_{i=1}^{4}\Lambda_i\left[E\left(X_t^a Y_{t-i/4}^A\right)-E\left(X_{t-2/4}^a Y_{t-i/4}^A\right)\right]\right.$$
$$\left.-\beta\sum_{i=2}^{3}\lambda_i^a\left[E\left(X_t^a X_{t-(2i)/4}^a\right)-E\left(X_{t-2/4}^a X_{t-(2i)/4}^a\right)\right]\right]\times\frac{1}{2}\frac{1}{Var\left(X_t^a\right)}\frac{1}{1-\rho\left(X_t^a,X_{t-2/4}^a\right)}+\frac{1}{2}\beta\left(\lambda_1^a-1\right).$$

This difference is not equal to $\frac{1}{2}\beta\left(\lambda_1^a-1\right)$ because the terms involving the expectations do not cancel out. These are defined in section 3.2 and have complicated expressions that depend on several coefficients. For ARMA(1,1) models, we obtain the ABIAS by means of Monte Carlo simulations. Therefore, in order to find closed-form expressions when $r=4, m=2$, we now assume the simplest case of $X_t^{(r)}$ following a zero-mean white noise process with variance σ_ν^2 or WN(0,σ_ν^2).

Proposition 4 *Assume the conditions in Proposition 3 hold. If $X_t^{(r)}$ is a WN(0,σ_ν^2),*

$$ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right)=\frac{1}{4}\beta\left(\sum_{i=1}^{3}\lambda_i-3\lambda_0\right)\ \textit{and}\ ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right)=\frac{1}{2}\beta\left(\lambda_1-\lambda_0\right),$$

where, from (2), the MIDAS (basic) exponential weights $\lambda_i=\exp\{\theta_1 i+\theta_2 i^2\}$, $i=0,1,2,3$, are expressed as

$$\lambda_0=1;\ \lambda_1=-\sum_{i=1}^{3}\delta_i;\ \lambda_2=\sum_{i=1}^{2}\sum_{j=i+1}^{3}\delta_i\delta_j;\ \lambda_3=-\prod_{i=1}^{3}\delta_i.$$

Consequently,

$$ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right)-ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right)=\frac{1}{4}\beta\left(-\sum_{i=0}^{1}\lambda_i+\sum_{i=2}^{3}\lambda_i\right)$$

and

$$ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right)^2 - ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right)^2 = \frac{1}{16}\beta^2\left[\left(\sum_{i=1}^{3}\lambda_i - 3\lambda_0\right)^2 - 4\left(\lambda_1-\lambda_0\right)^2\right].$$

Both ABIAS increase with β in absolute value and $ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right)/ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right) = 2$ if $\lambda_2+\lambda_3 = 2$. The biases approach zero as $(\lambda_1+\lambda_2+\lambda_3)$ becomes closer to 3 and as λ_1 gets close to 1 for the FLAT-LS estimator and the MIDAS-LS estimator, respectively. Moreover, $ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right) > ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right)$ if $\lambda_2+\lambda_3 > \max\left\{\lambda_0+\lambda_1, 5\lambda_0-3\lambda_1\right\} = \max\left\{1+\lambda_1, 5-3\lambda_1\right\}$ or if $\lambda_2+\lambda_3 < \min\left\{\lambda_0+\lambda_1, 5\lambda_0-3\lambda_1\right\} = \min\left\{1+\lambda_1, 5-3\lambda_1\right\}$.

Remark 1 *The bias can be computed for some usual MIDAS exponential weights taken from the literature, e.g. Ghysels et al. (2004, 2006) and Andreou et al. (2010a, 2010b). For the decreasing weights* $\lambda_1 = 0.994, \lambda_2 = 0.977, \lambda_3 = 0.949$ $(\theta_1 = 7\times10^{-4}, \theta_2 = -6\times10^{-3})$, $ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right) = -0.02\beta$, $ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right) = -0.003\beta$, *and* $ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right)^2 - ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right)^2 = 0.000391\beta^2$. *For the hump-shaped weights* $\lambda_1 = 1.042$, $\lambda_2 = 1.070$, $\lambda_3 = 1.081$, $(\theta_1 = 5\times10^{-2}$, $\theta_2 = -8\times10^{-3})$, $ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right) = 0.048\beta$, $ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right) = 0.021\beta$, *and* $ABIAS\left(\widehat{\beta}_{FLAT-LS};\cdot\right)^2 - ABIAS\left(\widehat{\beta}_{MIDAS-LS};\cdot\right)^2 = 0.001887\beta^2$. *In both examples, the FLAT-LS has the largest ABIAS in absolute value for all* β.

As previously mentioned, the bias of the MIDAS-NLS estimator can only be computed from the Monte Carlo simulations. Just for the sake of exposition, it can be shown that if $r = 4, m = 2$ $(R = 2)$, $ABIAS\left(\widehat{\beta}_{MIDAS-NLS};\pi_{0,z},\theta^*\right)$

$$\begin{aligned}
&= \beta\left[\frac{Cov\left(\left(X_t^a+\lambda_1^aX_{t-2/4}^a\right),\left(X_t^a+\lambda_1^{a^*}X_{t-2/4}^a\right)\right)}{Var\left(X_t^a+\lambda_1^{a^*}X_{t-2/4}^a\right)}-1\right]+\frac{Cov\left(\pi_{0,z}'Z_t^a, X_t^a+\lambda_1^{a^*}X_{t-2/4}^a\right)}{Var\left(X_t^a+\lambda_1^{a^*}X_{t-2/4}^a\right)}\\
&= \beta\left[\frac{\left(1+\lambda_1^a\lambda_1^{a^*}\right)+\left(\lambda_1^{a^*}+\lambda_1^a\right)\rho\left(X_t^a,X_{t-2/4}^a\right)}{\left(1+(\lambda_1^{a^*})^2\right)+2\lambda_1^{a^*}\rho\left(X_t^a,X_{t-2/4}^a\right)}-1\right]+\frac{Cov\left(\pi_{0,z}'Z_t^a,X_t^a\right)+\lambda_1^{a^*}Cov\left(\pi_{0,z}'Z_t^a,X_{t-2/4}^a\right)}{\left(1+(\lambda_1^{a^*})^2\right)Var(X_t^a)+2\lambda_1^{a^*}Cov\left(X_t^a,X_{t-2/4}^a\right)}\\
&= \beta(\lambda_1^a-\lambda_1^{a^*})\frac{\lambda_1^{a^*}+\rho\left(X_t^a,X_{t-2/4}^a\right)}{1+(\lambda_1^{a^*})^2+2\lambda_1^{a^*}\rho\left(X_t^a,X_{t-2/4}^a\right)}+\frac{Cov\left(\pi_{0,z}'Z_t^a,X_t^a\right)+\lambda_1^{a^*}Cov\left(\pi_{0,z}'Z_t^a,X_{t-2/4}^a\right)}{\left(1+(\lambda_1^{a^*})^2\right)Var\left(X_t^a\right)+2\lambda_1^{a^*}Cov\left(X_t^a,X_{t-2/4}^a\right)}
\end{aligned} \tag{20}$$

with $\pi'_{0,z}Z_t^a = -\Lambda_1 Y_{t-1/4}^A - \Lambda_2 Y_{t-2/4}^A - \Lambda_3 Y_{t-3/4}^A - \Lambda_4 Y_{t-4/4}^A + \beta\lambda_2^a X_{t-4/4}^a + \beta\lambda_3^a X_{t-6/4}^a$. Then, if $X_t^{(r)}$ is a WN(0,σ_ν^2), (20) becomes

$$ABIAS\left(\widehat{\beta}_{MIDAS-NLS};\pi_{0,z},\theta^*\right) = \beta(\lambda_1^a - \lambda_1^{a^*})\left[\frac{\lambda_1^{a^*}}{1+(\lambda_1^{a^*})^2}\right] - \beta\frac{\Lambda_1 + \lambda_1^{a^*}\left[\Lambda_1\left(\sum_{i=1}^{2}\sum_{j=i+1}^{3}\delta_i\delta_j - \sum_{i=1}^{3}\delta_i\right) + \Lambda_2\left(1-\sum_{i=1}^{3}\delta_i\right) + \Lambda_3\right]}{2\left[1+(\lambda_1^{a^*})^2\right]}$$

because $Cov\left(-\Lambda_1 Y_{t-1/4}^A, X_t^a\right) = -\Lambda_1 E\left(Y_{t-1/4}^A X_t^a\right) = -\Lambda_1\beta\sigma_v^2$ and

$$Cov\left(-\Lambda_1 Y_{t-1/4}^A - \Lambda_2 Y_{t-2/4}^A - \Lambda_3 Y_{t-3/4}^A, X_{t-2/4}^a\right) = -\Lambda_1 E\left(Y_t^A X_{t-1/4}^a\right) - \Lambda_2 E\left(Y_t^A X_t^a\right)$$
$$-\Lambda_3 E\left(Y_{t-1/4}^A X_t^a\right) = -\Lambda_1\beta\sigma_v^2\left(\sum_{i=1}^{2}\sum_{j=i+1}^{3}\delta_i\delta_j - \sum_{i=1}^{3}\delta_i\right) - \Lambda_2\beta\sigma_v^2\left(1-\sum_{i=1}^{3}\delta_i\right) - \Lambda_3\beta\sigma_v^2.$$

Hence,

$$ABIAS\left(\widehat{\beta}_{MIDAS-NLS};\pi_{0,z},\theta^*\right)$$
$$= \beta\left[\frac{\lambda_1^{a^*}\left(\lambda_1^a - \lambda_1^{a^*}\right)}{1+(\lambda_1^{a^*})^2} - \frac{(1-\lambda_1) + \lambda_1^{a^*}\left\{(1-\lambda_1)(\lambda_1+\lambda_2) + (\lambda_1+\lambda_2)(1+\lambda_1) + (\lambda_2-\lambda_3)\right\}}{2\left[1+(\lambda_1^{a^*})^2\right]}\right]$$
$$= \beta\left[\frac{-\lambda_1^{a^*}\left(\sum_{i=1}^{3}\delta_i^2 + \lambda_1^{a^*}\right)}{1+(\lambda_1^{a^*})^2} - \frac{(1-\lambda_1) + \lambda_1^{a^*}(2\lambda_1 + 3\lambda_2 - \lambda_3)}{2\left[1+(\lambda_1^{a^*})^2\right]}\right].$$

5 Simulation Experiments

In order to assess the aggregation effects both on the bias and mean square error (mse) of the estimators of α^* and β, we conducted a simulation experiment where $10,000$ basic time series of size $n = 6,000$ of the independent variable $X_t^{(r)}$ were generated from ARMA$(1,1)$ models with $\mathcal{N}(0,1)$ white noise and with parameters that lead to positive or to alternating autocorrelation functions. The (aggregate) dependent variable Y_t^A was computed from a MIDAS regression model with decaying or hump-shaped exponential weights (Andreou *et al.* 2010a, 2010b). Several values of the regression parameters α and β were also tried and the latter was chosen so that the impact of $X_t^{(r)}$ on Y_t^A was different.

The time series $X_t^{(r)}$ were generated first with $r = 4, 6, 12$ and aggregated with $m = 2$, $m = 2, 3$ and $m = 2, 3, 4$ respectively, since r/m has to be integer. For instance, $r = 4, m = 2$ may denote the case where the independent variable and Y_t^A are respectively bimonthly and four-month-period aggregates from monthly time series, or when they are half-yearly aggregates and yearly aggregates of quarterly time series respectively. In case $r = 12, m = 3$, the independent variable and Y_t^A may be quarterly and yearly aggregates of monthly time series respectively, etc. In order to further compare the estimators, particularly the MIDAS-LS and the MIDAS-NLS, the simulation experiment was extended by considering an ARMA model as above with $r = 60$, $m = 2, 3, 4, 5, 6, 10, 12$ and a new set of hump-shaped exponential weights ($\theta_1 = 1.5 \times 10^{-2}, \theta_2 = -3 \times 10^{-4}$) which yield a longer, more pronounced hump than previously.

The bias and mse of the three estimators were computed next and some results are displayed in Tables 1 and 2 respectively (the detailed results are in Tables S1 and S2 in the Supplemental material). The asymptotic bias of the FLAT-LS and MIDAS-LS estimators was also computed from (18) and (19) respectively for the models used in the simulation experiment. This provided numerical values of such bias, enabling its assessment and the comparison of the estimators which is not possible from those equations. The results are displayed in Table 3 for one of the models used (detailed results in Table S3 of the Supplemental material).

The simulation results show that the bias (in absolute value) and mse of the three estimators grow as m increases which means that aggregation worsens their performance and the impact can be strong even for low or moderate m. Furthermore, the bias (in absolute value) and the mse for both parameters increase with the respective values. Concerning α^*, the three estimators have (approximately) the same bias and mse when $r/m = 2$. For $r/m > 2$, the MIDAS-LS and MIDAS-NLS estimators have approximately the same bias and mse, outperforming the FLAT-LS estimator.

Concerning β, for a given m, the bias and mse of the three estimators depend solely on β, i.e., they are not affected by the value of α. Moreover, the bias of the three estimators is proportional to the value of β. The bias of both the MIDAS-LS and MIDAS-NLS estimators is negative for

decaying weights and positive for hump-shaped ones. Furthermore, the FLAT-LS estimator is clearly outperformed by the other two, both in terms of bias and mse. The MIDAS-LS and MIDAS-NLS estimators have approximately the same performance for all parameters and m except for $r = 60$ and the longer-hump weights. For such weights and low m ($m \leq 6$), the latter estimator outperforms the former, even though the difference is small, but their bias and mse become approximately the same as m grows. Thus, the performance of the linear and nonlinear estimators of β based on aggregate time series tends to be the same, although a large m ($m = 10, 12$) is sometimes required. If nonlinearity is more pronounced, as in the case of the weights with a longer-hump shape, a larger m is required. Therefore, we can conclude that aggregation weakens nonlinearity in MIDAS regression models and that the effect can be severe even for moderate m. In fact, this finding agrees with the literature on the effects of aggregation on nonlinear models (Granger and Lee, 1999; Teles and Wei, 2000).

One might think that it is not easy to find in practice such large samples as we used in the simulations, but $n = 6,000$ is the length of the basic time series which are not observed, since only their aggregates are available. So, for instance, the lengths of Y_t^A and X_t^a if $r = 12$, $m = 4$ are 500 and 1500 respectively and, if $r = 60$, $m = 10$, they are 100 and 600 respectively. Thus, aggregate series are much shorter and yet they had to be kept long in the simulation because its results could not be jeopardized by a small sample size. Therefore, a large n was required. Nevertheless, the simulation experiment was further extended for some models by including smaller series such as $n = 360$, with $r = 12$, $m = 2, 3, 4$ and $n = 960$ with $r = 30$, $m = 2, 3, 5, 10$ which substantially lowers the aggregate sample sizes (for instance, if $n = 360$ and $r = 12$, $m = 4$, the lengths of Y_t^A and X_t^a are 30 and 90 respectively). Some results are in Tables 1 and 2 (the complete results are in Tables S1 and S2 of the Supplemental material). The prior conclusions remain the same, with larger bias (in absolute value) and mse of the estimator of β.

In order to check whether our findings are more general, an ARFIMA(1,d,1) model with a memory parameter $d = -0.3$ and a GARCH(1,1) model were also used as data-generating processes of the

basic time series. The results lead to the same conclusions as previously and are displayed in Tables S1 and S2 in the Supplemental material.

Concerning the FLAT-LS and MIDAS-LS estimators (linear least-squares estimators), computation of the asymptotic bias provided the same conclusions as above (Table 3).

6 Application with Macro and Financial Data

In this section, the FLAT-LS, MIDAS-LS, and MIDAS-NLS with exponential weights estimators were compared for different levels of aggregation using empirical applications with macroeconomic data as in Ghysels and Marcellino (2018) and financial data following Andreou *et al.* (2013). In the former application, the variable of interest is the quarterly US economic growth (g) from 1985Q1 until 2012Q4 where, following the cited authors, each regression includes a single determinant (X). The list of variables, published by the Conference Board, includes the index of coincident indicators (ci) and its four main components: non-farm payroll employees (e), index of industrial production (ip), personal income less transfer payments (inc), and manufacturing and trade sales (sa), all observed on a monthly basis from 1985M1 until 2012M12. Just as with the US real GDP, all the variables were transformed into growth rates through the log-difference operator. More details about the data can be found in Ghysels and Marcellino (2018). The aforementioned list also includes financial variables observed on a daily basis: returns of the NASDAQ 100 index ($nasdaq$, with $6,783$ observations starting in 1987), Chicago Board Options Exchange volatility index (vix, with $6,000$ observations starting in 1990), ICE BofA option-adjusted spreads (oas, with $4,228$ observations starting in 1997), and 5-year breakeven inflation rate ($t5y$, with $2,608$ observations starting in 2003). The data were retrieved from the Federal Reserve Bank of St. Louis (FRED), at https://fred.stlouisfed.org (NASDAQ100, VIXCLS, BAMLC0A0CM, and T5YIE respectively).

The main motivation of this paper was to assess the estimation biases of the FLAT-LS, MIDAS-LS, and MIDAS-NLS estimators once we aggregate data from its basic, highest frequency, into a lower frequency. With this in mind, we extended the original Ghysels and Marcellino's empirical

Table 1. Bias computed from a simulation experiment for the three estimators

	$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6,$				$\phi_0 = 1, \phi_1 = -0.7, \omega_1 = 0.5,$				$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6,$			
	$\theta_1 = 7\times10^{-4}, \theta_2 = -6\times10^{-3},$				$\theta_1 = 7\times10^{-4}, \theta_2 = -6\times10^{-3},$				$\theta_1 = 5\times10^{-2}, \theta_2 = -8\times10^{-3},$			
	$\sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10$				$\sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10$				$\sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10$			
$r = 6, n = 6,000$												
$m = 2$												
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	8.468	−0.211	8.548	−0.395	8.396	−0.206	8.406	−0.386	10.741	0.200	10.860	0.375
MLS	8.427	−0.018	8.467	−0.035	8.396	−0.019	8.405	−0.035	10.673	0.084	10.734	0.158
MNLS	8.427	−0.018	8.467	−0.035	8.396	−0.019	8.405	−0.035	10.673	0.084	10.734	0.158
$m = 3$												
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	9.819	−0.211	9.898	−0.395	9.746	−0.206	9.757	−0.386	10.091	0.200	10.210	0.375
MLS	9.819	−0.055	9.897	−0.103	9.746	−0.045	9.757	−0.084	10.091	0.141	10.210	0.265
MNLS	9.819	−0.055	9.897	−0.103	9.746	−0.045	9.757	−0.084	10.091	0.141	10.210	0.265
$r = 12, n = 6,000$												
$m = 2$												
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	5.054	−0.825	5.472	−1.547	4.592	−0.805	4.618	−1.506	7.504	−0.253	8.432	−0.473
MLS	4.652	−0.020	4.736	−0.039	4.587	−0.023	4.609	−0.043	6.620	0.081	6.779	0.153
MNLS	4.652	−0.020	4.735	−0.039	4.588	−0.022	4.609	−0.042	6.619	0.082	6.778	0.153

Note: FLT, MLS and MNLS denote the FLAT-LS, MIDAS-LS and MIDAS-NLS estimators respectively.

Table 1 (cont.). Bias computed from a simulation experiment for the three estimators

$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6,$ $\theta_1 = 7\times10^{-4}, \theta_2 = -6\times10^{-3},$ $\sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10$	$\phi_0 = 1, \phi_1 = -0.7, \omega_1 = 0.5,$ $\theta_1 = 7\times10^{-4}, \theta_2 = -6\times10^{-3},$ $\sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10$	$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6,$ $\theta_1 = 5\times10^{-2}, \theta_2 = -8\times10^{-3},$ $\sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10$

$r = 12, n = 6,000$

$m = 3$

	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	7.515	−0.825	7.937	−1.547	7.054	−0.805	7.080	−1.506	9.778	−0.253	10.705	−0.473
MLS	7.226	−0.063	7.413	−0.117	7.035	−0.044	7.042	−0.080	9.143	0.132	9.526	0.248
MNLS	7.226	−0.063	7.413	−0.117	7.035	−0.043	7.042	−0.079	9.143	0.132	9.526	0.248

$m = 4$

	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	8.854	−0.825	9.273	−1.546	8.392	−0.805	8.419	−1.506	10.599	−0.253	11.527	−0.473
MLS	8.683	−0.123	8.969	−0.230	8.390	−0.094	8.413	−0.174	10.217	0.153	10.812	0.286
MNLS	8.683	−0.123	8.969	−0.230	8.390	−0.094	8.413	−0.174	10.217	0.153	10.812	0.286

$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6, \sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10, \theta_1 = 1.5 \times 10^{-2}, \theta_2 = -3 \times 10^{-4}, r = 60, n = 6,000$

	$m = 2$				$m = 5$				$m = 12$			
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	19.740	0.377	21.841	0.707	12.458	0.377	14.559	0.707	12.468	0.377	14.569	0.707
MLS	17.320	0.033	17.361	0.061	10.371	0.124	10.677	0.234	11.147	0.297	12.108	0.559
MNLS	17.331	0.031	17.318	0.059	10.362	0.121	10.646	0.228	11.148	0.297	12.097	0.559

Note: FLT, MLS and MNLS denote the FLAT-LS, MIDAS-LS and MIDAS-NLS estimators respectively.

Table 1 (cont.). Bias computed from a simulation experiment for the three estimators

$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6, \sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10, \theta_1 = 8.5 \times 10^{-2}, \theta_2 = -3 \times 10^{-3}, r = 30, n = 960$

	$m = 2$				$m = 5$				$m = 10$			
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	16.050	1.813	21.793	3.401	16.608	1.813	22.351	3.401	16.608	1.813	22.352	3.401
MLS	9.769	0.190	10.024	0.356	12.040	0.725	13.760	1.369	14.723	1.431	18.799	2.687
MNLS	9.788	0.183	9.975	0.345	12.029	0.721	13.782	1.359	14.723	1.431	18.799	2.687

Note: FLT, MLS and MNLS denote the FLAT-LS, MIDAS-LS and MIDAS-NLS estimators respectively.

exercise by modeling the growth rates (g) in a yearly frequency ($r = 12$, $t = 1985, 1986, \ldots, 2012$) which results from the sum of the values of the corresponding yearly quarters. In order to study the impact of temporal aggregation on estimation, we considered each determinant X in its basic monthly frequency ($m = 1$), bimonthly ($m = 2$), quarterly ($m = 3$), and four-month period ($m = 4$). Concerning the financial determinants, the frequency of GDP growth rates is quarterly (or yearly) and the frequency of X is monthly and weekly (or quarterly and monthly). The main results for the estimated slope are in Table 4 (the complete results can be found in Tables S4 to S44 in the Supplemental material and show all the estimated coefficients, including the MIDAS-NLS exponential weights).

The estimated intercept is robust to the level of aggregation m and non-farm payroll employees (e) is the macro determinant with the largest estimated slope. Not surprisingly, the FLAT-LS estimator is the same for every m. Furthermore, the MIDAS-LS and -NLS estimates of β are sometimes very close or almost coincident for $m = 1$, depending on r. However, even when they are different for $m = 1$, they tend to get closer to each other as m increases and become (nearly) coincident, especially for the macro variables. The FLAT-LS estimates are always different from the other two and the differences can be large (in absolute value). These empirical results agree with

Table 2. MSE computed from a simulation experiment for the three estimators

ϕ_0=1, ϕ_1=0.4, ω_1=−0.6, θ_1=7×10^{-4}, θ_2=−6×10^{-3}, σ^2_ν=1, σ^2_ε=1, α=10				ϕ_0=1, ϕ_1=−0.7, ω_1=0.5, θ_1=7×10^{-4}, θ_2=−6×10^{-3}, σ^2_ν=1, σ^2_ε=1, α=10				ϕ_0=1, ϕ_1=0.4, ω_1=−0.6, θ_1=5×10^{-2}, θ_2=−8×10^{-3}, σ^2_ν=1, σ^2_ε=1, α=10			
$r = 6, n = 6,000$												
$m = 2$												
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	71.748	0.045	73.109	0.157	70.497	0.043	70.673	0.149	115.403	0.040	117.981	0.141
MLS	71.038	0.001	71.714	0.002	70.493	0.001	70.651	0.002	113.949	0.008	115.238	0.026
MNLS	71.038	0.001	71.714	0.002	70.493	0.001	70.651	0.002	113.949	0.008	115.2238	0.026
$m = 3$												
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	96.438	0.045	97.013	0.157	94.992	0.043	95.196	0.149	101.859	0.040	104.283	0.141
MLS	96.440	0.004	97.995	0.011	94.997	0.002	95.204	0.008	101.856	0.020	104.279	0.071
MNLS	96.440	0.004	97.995	0.011	94.997	0.002	95.204	0.008	101.856	0.020	104.279	0.071
$r = 12, n = 6,000$												
$m = 2$												
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	26.137	0.682	31.798	2.396	21.155	0.650	21.544	2.282	56.754	0.065	72.362	0.226
MLS	21.764	0.002	22.559	0.003	21.061	0.002	21.267	0.004	43.979	0.008	46.163	0.025
MNLS	21.761	0.002	22.557	0.003	21.062	0.002	21.270	0.004	43.970	0.008	46.149	0.025

Note: FLT, MLS and MNLS denote the FLAT-LS, MIDAS-LS and MIDAS-NLS estimators respectively.

Table 2 (cont.). MSE computed from a simulation experiment for the three estimators

	ϕ_0=1, ϕ_1=0.4, ω_1=−0.6, θ_1=7×10^{-4}, θ_2=−6×10^{-3}, σ^2_ν=1, σ^2_ε=1, α=10				ϕ_0=1, ϕ_1=−0.7, ω_1=0.5, θ_1=7×10^{-4}, θ_2=−6×10^{-3}, σ^2_ν=1, σ^2_ε=1, α=10				ϕ_0=1, ϕ_1=0.4, ω_1=−0.6, θ_1=5×10^{-2}, θ_2=−8×10^{-3}, σ^2_ν=1, σ^2_ε=1, α=10			
$r = 12, n = 6,000$												
$m = 3$												
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	57.085	0.683	64.824	2.396	49.834	0.650	50.352	2.282	96.052	0.065	115.881	0.226
MLS	52.399	0.005	55.266	0.016	49.503	0.003	49.616	0.008	83.787	0.019	91.079	0.064
MNLS	52.396	0.005	55.262	0.016	49.504	0.003	49.619	0.008	83.787	0.019	91.078	0.064
$m = 4$												
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	78.988	0.682	87.856	2.396	70.503	0.650	71.091	2.282	112.783	0.065	134.135	0.226
MLS	75.567	0.016	80.652	0.055	70.413	0.010	70.826	0.032	104.594	0.025	117.346	0.085
MNLS	75.567	0.016	80.652	0.055	70.413	0.010	70.826	0.032	104.594	0.025	117.346	0.085
$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6, \sigma^2_\nu = 1, \sigma^2_\varepsilon = 1, \alpha = 10, \theta_1 = 1.5 \times 10^{-2}, \theta_2 = -3 \times 10^{-4}, r = 60, n = 6,000$												
	$m = 2$				$m = 5$				$m = 12$			
	$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$		$\beta = 4$		$\beta = 7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	409.048	0.144	539.352	0.506	176.864	0.144	276.572	0.506	117.199	0.144	276.869	0.506
MLS	303.181	0.010	305.121	0.014	112.784	0.020	121.924	0.061	132.092	0.092	164.074	0.320
MNLS	301.987	0.004	301.948	0.006	112.097	0.017	120.529	0.056	131.946	0.091	163.339	0.319

Note: FLT, MLS and MNLS denote the FLAT-LS, MIDAS-LS and MIDAS-NLS estimators respectively.

Table 2 (cont.). MSE computed from a simulation experiment for the three estimators

$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6, \sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \alpha = 10, \theta_1 = 8.5 \times 10^{-2}, \theta_2 = -3 \times 10^{-3}, r = 30, n = 960$

	$m=2$				$m=5$				$m=10$			
	$\beta=4$		$\beta=7.5$		$\beta=4$		$\beta=7.5$		$\beta=4$		$\beta=7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	395.600	3.343	951.546	11.755	413.815	3.343	976.168	11.755	413.837	3.343	976.197	11.755
MLS	115.466	0.102	140.066	0.260	183.844	0.496	309.927	2.090	303.559	2.139	642.478	7.517
MNLS	106.601	0.052	121.166	0.156	179.409	0.472	297.944	2.002	303.559	2.139	642.478	7.517

Note: FLT, MLS and MNLS denote the FLAT-LS, MIDAS-LS and MIDAS-NLS estimators respectively.

Table 3. Asymptotic bias — linear least-squares estimators

$\phi_0 = 1, \phi_1 = 0.4, \omega_1 = -0.6, \sigma_\nu^2 = 1, \sigma_\varepsilon^2 = 1, \theta_1 = 7 \times 10^{-4}, \theta_2 = -6 \times 10^{-3}, \alpha = 10,\ r = 12$

	$m=2$				$m=3$				$m=4$			
	$\beta=4$		$\beta=7.5$		$\beta=4$		$\beta=7.5$		$\beta=4$		$\beta=7.5$	
	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β	α^*	β
FLT	4.230	−0.784	3.943	−1.470	6.693	−0.784	6.405	−1.470	8.031	−0.784	7.743	−1.470
MLS	4.528	−0.003	4.501	−0.006	6.928	−0.016	6.847	−0.029	8.190	−0.043	8.043	−0.081

Note: FLT, MLS and MNLS denote the FLAT-LS, MIDAS-LS and MIDAS-NLS estimators respectively.

those found with the simulations. Note also that the differences between the FLAT-LS estimator and the other two depend on the determinant used in the regression.

For the index of coincident indicators (ci), the MIDAS-LS and MIDAS-NLS estimates gain statistical significance for larger m. These are larger (smaller) than the FLAT-LS estimate for $m = 4$ ($m = 2$). On the contrary, for the non-farm payroll employees (e), the MIDAS-LS and MIDAS-NLS estimated slopes decrease with m but are always greater than that found with the FLAT-LS estimator. For the index of industrial production (ip), the MIDAS-LS (MIDAS-NLS)

increases (decreases) with m and again both are always greater than the FLAT-LS. The MIDAS-LS and MIDAS-NLS estimators are not statistically significant in the case of personal income (inc) for all m. Finally, the slope of the FLAT-LS estimator is slightly greater than the other two for the determinant manufacturing and trade sales (sa). On the other hand, for the NASDAQ returns ($nasdaq$) and the breakeven inflation rate ($t5y$), the estimated slopes are the same for the MIDAS-LS and MIDAS-NLS and are very different from the FLAT-LS estimate, unlike the other two financial variables. In fact, for the volatility index (vix), the FLAT-LS and MIDAS-LS estimates are roughly the same, whereas for the option-adjusted spreads (oas), any pairwise similarity depends on the specific aggregation case. Finally, when these four determinants are weekly aggregated, the slope is not statistically significant for the MIDAS-LS and MIDAS-NLS estimators.

7 Concluding Remarks

This paper discussed the effects of temporal aggregation in the context of a MIDAS regression. Applying methods available for commonly used time series models, we derived the true specification of the model under temporal aggregation. Unlike existing studies, we determined the correct number of lags of all components (dependent and independent variables and the moving average error term), thus showing that the FLAT-LS, MIDAS-LS, and the MIDAS-NLS estimators are biased. We also concluded that assuming aggregate weights with the usual exponential form is an additional source of bias of the latter estimator.

Our Monte Carlo results show that the bias (in absolute value) and mse of the three estimators increase as the order of aggregation m grows which means that aggregation worsens the estimators' performance. Furthermore, the FLAT-LS estimator, which fully aggregates the independent variables, has always the largest bias and mse. The performance of the other two estimators is usually the same for any m but, when nonlinearity is more pronounced, the MIDAS-NLS is the best, although by a small difference, for low m. However, when m grows, their performance tends to be

Table 4. Estimation results of the slope for monthly and daily basic X

	g^A	X^a	FLT	MLS	MNLS
$X = ci$ (monthly)	yearly	bimonthly	0.695^{***} (0.069)	0.266 (0.628)	0.351 (0.476)
	yearly	quarterly	0.695^{***} (0.069)	0.684^{**} (0.318)	0.683^{**} (0.309)
	yearly	four-month period	0.695^{***} (0.069)	0.811^{***} (0.289)	0.811^{***} (0.289)
$X = e$ (monthly)	yearly	bimonthly	0.837^{***} (0.119)	3.092^{***} (1.089)	3.374^{***} (1.136)
	yearly	quarterly	0.837^{***} (0.119)	3.037^{***} (0.810)	3.190^{***} (0.666)
	yearly	four-month period	0.837^{***} (0.119)	1.933^{***} (0.490)	2.127^{***} (0.504)
$X = ip$ (monthly)	yearly	bimonthly	0.373^{***} (0.045)	0.444^{*} (0.230)	0.625^{***} (0.184)
	yearly	quarterly	0.373^{***} (0.045)	0.548^{***} (0.153)	0.598^{***} (0.145)
	yearly	four-month period	0.373^{***} (0.045)	0.611^{***} (0.117)	0.583^{***} (0.118)
$X = inc$ (monthly)	yearly	bimonthly	0.547^{***} (0.114)	0.100 (0.260)	0.186 (0.203)
	yearly	quarterly	0.547^{***} (0.114)	0.277 (0.250)	0.258 (0.232)
	yearly	four-month period	0.547^{***} (0.114)	0.349 (0.247)	0.349 (0.247)
$X = sa$ (monthly)	yearly	bimonthly	0.413^{***} (0.052)	0.343 (0.224)	0.343^{**} (0.159)
	yearly	quarterly	0.413^{***} (0.052)	0.244 (0.148)	0.334^{**} (0.122)
	yearly	four-month period	0.413^{***} (0.052)	0.378^{***} (0.113)	0.378^{***} (0.113)
$X = nasdaq$ (daily)	quarterly	monthly	0.00018^{***} (5.051×10^{-5})	0.00011^{**} (4.786×10^{-5})	0.00011^{**} (4.786×10^{-5})
	quarterly	weekly	0.00018^{***} (5.051×10^{-5})	0.00011^{*} (6.076×10^{-5})	8.527×10^{-5} (5.317×10^{-5})
	annually	quarterly	0.0003^{***} (0.0001)	0.0005^{***} (0.0001)	0.0006^{***} (0.0002)
$X = vix$ (daily)	quarterly	monthly	-0.0004^{***} (0.00008)	-0.0004^{**} (0.0002)	-0.00026 (0.0002)
	quarterly	weekly	-0.0004^{***} (0.00008)	-0.00069^{*} (0.0003)	-8.409×10^{-5} (0.0001)
	annually	quarterly	-0.001^{**} (0.0006)	-0.001^{***} (0.0004)	-0.001^{**} (0.0004)

Table 4. (cont.) Estimation results of the slope for monthly and daily basic X

	g^A	X^a	FLT	MLS	MNLS
$X = oas$ (daily)	quarterly	monthly	-0.004^{***} (0.0005)	-0.009^{***} (0.003)	-0.004^{*} (0.003)
	quarterly	weekly	-0.004^{***} (0.0005)	-0.003 (0.017)	-5.718×10^{-8} (1.974×10^{-5})
	annually	quarterly	-0.018^{***} (0.003)	-0.013^{**} (0.005)	-0.011^{**} (0.005)
$X = t5y$ (daily)	quarterly	monthly	0.007^{***} (0.001)	0.003 (0.004)	0.003 (0.004)
	quarterly	weekly	0.007^{***} (0.001)	-0.0004 (0.016)	6.378×10^{-8} (4.005×10^{-5})
	annually	quarterly	0.024^{*} (0.012)	0.019^{***} (0.003)	0.020^{***} (0.004)

Notes: *,**,*** stand for significance at 10%, 5%, 1% levels, respectively; standard errors in parentheses; ci: index of coincident indicators; e: non-farm payroll employees; ip: index of industrial production; inc: personal income; sa: manufacturing and trade sales; nasdaq: returns of the nasdaq 100 index; vix: CBOE volatility index; oas: ICE BofA option-adjusted spreads; t5y: 5-year breakeven inflation rate; FLT, MLS and MNLS denote the FLAT-LS, MIDAS-LS and MIDAS-NLS estimators respectively.

the same which shows that aggregation weakens nonlinearity in MIDAS regressions and that the effect can be severe for moderate or high m. Thus, using a linear or a nonlinear MIDAS regression with aggregate time series is equivalent in what concerns parameter estimation.

Moreover, for simple aggregate models, we compared the analytical closed-form expressions of the bias of the FLAT-LS and MIDAS-LS estimators and concluded that the former is the least accurate. The results found were illustrated with real data where the dependent variable is the quarterly US economic growth and each covariate is a monthly component of the index of coincident indicators or a keystone financial variable observed at a daily frequency.

We showed the class of MIDAS regression models is not closed under temporal aggregation of an arbitrary number of p independent variables. An important extension to this paper is to

study the properties of the estimators when p is "large". As in Babii *et al.* (2022), for these high-dimensional mixed-frequency regressions we can model the MIDAS weights linearly, e.g., with Legendre polynomials. It would also be interesting to incorporate time-varying coefficients in the linear regression model, possibly in a multivariate framework. This alternative approach can follow from the recent work by Chan *et al.* (2023), which consider Gaussian state space models with missing data, and Chan *et al.* (2024) that nowcast the US GDP with time-varying parameter MIDAS models. We leave that for future work.

The comparison of the forecasting performance using the existing misspecified ARDL-MIDAS model versus the correct model we proposed is of particular interest, but beyond the scope of this paper. This research is currently ongoing. An important aspect to be dealt with in such a forecasting problem is the fact that the true model we derived includes rolling overlapping aggregates of the dependent variable on the right-hand side.

Acknowledgments

The authors are very grateful to the Editor and to two anonymous referees whose valuable suggestions and comments helped to improve the paper. The first author acknowledges financial support from FCT - Fundação para a Ciência e a Tecnologia (National Science and Technology Foundation) through grant UIDB/00315/2020.

Disclosure Statement

The authors report there are no competing interests to declare.

Supplemental Material

Online Appendix Supplemental material: The Effects of Temporal Aggregation on MIDAS Regressions (.pdf file)

References

Andreou, E., Ghysels, E., and Kourtellos, A. (2010a), "Regression Models With Mixed Sampling Frequencies," *Journal of Econometrics*, 158, 246–261, doi: 10.1016/j.jeconom.2010.01.004.

——— (2010b), "Forecasting With Mixed-Frequency Data," *University of Cyprus Discussion Paper, Department of Economics, 2010-10.*

——— (2013), "Should Macroeconomic Forecasters Use Daily Financial Data and How?," *Journal of Business and Economic Statistics*, 31, 240–251, doi: 10.1080/07350015.2013.767199.

Armesto, M.T., Engemann, K.M., and Owyang, M.T. (2010), "Forecasting With Mixed Frequencies," *Federal Reserve Bank of St. Louis Review*, 92, 521–536.

Babii, A., Ghysels, E., and Striaukas, J. (2022), "Machine Learning Time Series Regressions With an Application to Nowcasting," *Journal of Business and Economic Statistics*, 40, 1094–1106, doi: 10.1080/07350015.2021.1899933.

Breitung, J., and Swanson, N.R. (2002), "Temporal Aggregation and Spurious Instantaneous Causality in Multiple Time Series Models," *Journal of Time Series Analysis*, 23, 651–665, doi:10.1111/1467-9892.00284.

Chan, J.C.C., Poon, A., and Zhu, D. (2023), "High-dimensional conditionally gaussian state space models with missing data," *Journal of Econometrics*, 236, 105468, doi: 10.1016/j.jeconom.2023.05.005.

——— (2024), "Time-Varying Parameter MIDAS Models: Application to Nowcasting US real GDP", SSRN https://ssrn.com/abstract=4802887.

Clements, M.P., and Galvão, A.B. (2008), "Macroeconomic Forecasting With Mixed-Frequency Data: Forecasting Output Growth in the United States," *Journal of Business and Economic Statistics*, 26, 546–554, doi: 10.1198/073500108000000015.

——— (2009), "Forecasting US Output Growth Using Leading Indicators: An Appraisal Using MIDAS Models," *Journal of Applied Econometrics*, 24, 1057–1217, doi: 10.1002/jae.1075.

Drost, F,C., and Nijman, T.E. (1993), "Temporal Aggregation of Garch Processes," *Econometrica*, 61, 909–927, doi: 10.2307/2951767.

Foroni, C., and Marcellino, M. (2013), "A Survey of Econometric Methods for Mixed-Frequency Data," *Norges Bank Working Paper*.

——— (2014), "A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates," *International Journal of Forecasting*, 30, 554–568, doi: 10.1016/j.ijforecast.2013.01.0

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 A*, 178, 57–82, doi: 10.1111/rssa.12043.

Foroni, C., Marcellino, M., and Stevanović, D. (2018), "Mixed Frequency Models With MA Components," *European Central Bank, Working Paper Series No 2206*.

Ghysels, E. (2016), "Macroeconomics and the Reality of Mixed Frequency Data," *Journal of Econometrics*, 196, 294–314, doi: 10.1016/j.jeconom.2016.04.008.

Ghysels, E., and Marcellino, M. (2018), "*Applied Economic Forecasting Using Time Series Methods"*, New York: Oxford University Press.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004), "The MIDAS Touch: Mixed Data Sampling Regression Models," *UCLA: Finance*, Retrieved from https://escholarship.org/uc/item/9mf223rs.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2005), "There is a Risk-Return Tradeoff After All," *Journal of Financial Economics*, 76, 509–548, doi: 10.1016/j.jfineco.2004.03.008.

——— (2006), "Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies," *Journal of Econometrics*, 131, 59–95, doi: 10.1016/j.jeconom.2005.01.004.

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

Ghysels, E., and Wright, J. (2009), "Forecasting Professional Forecasters," *Journal of Business and Economic Statistics*, 27, 504–516, doi: 10.1198/jbes.2009.06044.

Granger, C.W.J., and Lee, T.H. (1999), "The effect of aggregation on nonlinearity", *Econometric Reviews*, 18, 259–269, doi: 10.1080/07474939908800445.

Hotta, L.K., and Neto, J.C. (1993), "The Effect of Aggregation on Prediction in Autoregressive Integrated Moving-Average Models," *Journal of Time Series Analysis*, 14, 261–269, doi: 10.1111/j.1467-9892.1993.tb00143.x.

Lütkepohl, H. (1987), "*Forecasting Aggregated Vector ARMA Processes*," Berlin: Springer.

——— (2009), "Forecasting Aggregated Time Series Variables: A Survey," *European University Institute Working Paper, Department of Economics*.

Mamingi, N. (2017), "Beauty and Ugliness of Aggregation Over Time: A Survey," *Review of Economics*, 68, 205–227, doi: 10.1515/roe-2017-0027.

Marcellino, M. (1999), "Some Consequences of Temporal Aggregation in Empirical Analysis," *Journal of Business and Economic Statistics*, 17, 129–136, doi: 10.1080/07350015.1999.10524802.

Schumacher, C. (2016), "A Comparison of MIDAS and Bridge Equations," *International Journal of Forecasting*, 32, 257–270, doi: 10.1016/j.ijforecast.2015.07.004.

Silvestrini, A., and Veredas, D. (2008), "Temporal Aggregation of Univariate and Multivariate Time Series Models: A Survey," *Journal of Economic Surveys*, 22, 458–497, doi: 10.1111/j.1467-6419.2007.00538.x.

Stram, D.O., and Wei, W.W.S. (1986), "Temporal Aggregation in the ARIMA Process," *Journal of Time Series Analysis*, 7, 279–292, doi: 10.1111/j.1467-9892.1986.tb00495.x.

Teles, P., and Wei, W.W.S. (2000), "The effects of temporal aggregation on tests of linearity of a time series," *Computational Statistics and Data Analysis*, 34, 91–103, doi: 10.1016/S0167-9473(99)00072-9.

Wei, W.W.S. (2006), "*Time Series Analysis: Univariate and Multivariate Methods,"* 2nd ed., New York: Pearson.