Statistica Sinica **32** (2022), 1811-1833
doi:https://doi.org/10.5705/ss.202020.0264

MALLOWS MODEL AVERAGING ESTIMATOR FOR THE MIDAS MODEL WITH ALMON POLYNOMIAL WEIGHT

Hsin-Chieh Wong[1,2] and Wen-Jen Tsay[3]

[1]*National Central University,* [2]*National Chengchi University and* [3]*Academia Sinica*

Abstract: This research proposes an ordinary least squares (OLS)-based model averaging estimator using the Mallows model averaging (MMA) criterion for the MIxed DAta Sampling (MIDAS) model. We use a Vandermonde matrix to approximate the unknown weighting functions for the MIDAS model, enabling us to semiparametrically estimate each candidate model for averaging with the OLS estimator. We show that the proposed MMA estimator possesses the same asymptotic optimality properties considered in the literature under suitable regularity conditions, even though the data-generating process is much more general than the previously considered cross-sectional data structure. In addition to the simplicity of implementing the proposed MMA approach for the MIDAS model, our method delivers great numerical performance under various configurations considered in our Monte Carlo simulations.

Key words and phrases: Aggregate impact parameter, asymptotic optimality, model averaging, semiparametric MIDAS model.

1. Introduction

Improving macroeconomic forecasting is of great importance to policymakers and investors with regard to daily decision-making. Recent developments in econometric methods are making this possible by considering the presence of a huge set of real-time high-frequency economic and financial time series. Among them, Aruoba and Diebold (2010) and Giannone, Reichlin and Small (2008) propose the concept of "nowcast" to popularize the idea that one can estimate current unavailable low-frequency (usually monthly or quarterly) real economic activity using timely higher frequency (such as daily) variables.

The concept of nowcasting has attracted a growing strand of the literature based on the MIxed DAta Sampling (MIDAS) model. In fact, MIDAS methods have been proposed by Ghysels, Santa-Clara and Valkanov (2005, 2006) and

Corresponding author: Hsin-Chieh Wong, Department of Mathematics, National Central University, Taoyuan 32001, Taiwan. E-mail: wonghsinchieh@gmail.com.

Ghysels and Valkanov (2012) for forecasting volatility. Andreou, Ghysels and Kourtellos (2013), Carriero, Clark and Marcellino (2015), Clements and Galvao (2008, 2009), Engle, Ghysels and Sohn (2013), Ghysels, Sinko and Valkanov (2007), Ghysels and Valkanov (2012), Kuzin, Marcellino and Schumacher (2011, 2013), and Foroni, Marcellino and Stevanovic (2019) also consider many interesting issues using MIDAS approaches, including GDP growth, stock market volatility, U.S. inflation, and the relation between stock market volatility and macroeconomic activity, to name a few.

The key insight behind the MIDAS methodology is to weight over higher frequency explanatory variables in order to construct a new time series so as to match the length of the lower frequency dependent variable. Doing this allows one to run a parsimoniously parameterized regression of data observed at different frequencies, as advocated by Ghysels, Santa-Clara and Valkanov (2005). This also implies different estimators might be pursued under varying weighting schemes. For example, Ghysels, Santa-Clara and Valkanov (2006) employ the nonlinear least squares (NLS) estimator, because they use the *exponential* Almon weighting function to ensure all the weights are positive. Conversely, without imposing such a constraint, Chen and Tsay (2011) suggest an ordinary least squares (OLS)-based MIDAS model with a close link to the traditional autoregressive distributed lag (ADL) literature. The computational burden of the OLS-based MIDAS model is certainly much lower than that of NLS-based counterparts.

There is no doubt that the coverage of MIDAS models is tremendous. However, the issue of model uncertainty is never touched upon in the MIDAS literature. Scholars of various fields have investigated the theoretical foundations of using model selection techniques in the time series literature. For example, one might resort to information criteria, such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), or Hannan–Quinn (HQ) principle. In this paper, we propose an easy-to-implement forecast combination procedure based on the Mallows model averaging (MMA) criterion for the MIDAS model by generalizing the results in Hansen (2007). Compared with the Bayesian model averaging approaches considered in Hoeting et al. (1999) and Raftery, Madigan and Hoeting (1997), the MMA method is a frequency model averaging technique. Using simulations, Hansen (2007) demonstrates that the MMA estimator outperforms the AIC and BIC model selection methods and other averaging methods in terms of risk (expected squared error).

Hansen (2007) uses an OLS-based model averaging estimator, with the weights selected by minimizing a criterion in the spirit of Mallows' C_p (Mallows (1973)). Hansen (2008) also considers the asymptotic properties of a least squares forecast averaging method based on the MMA criteria for stationary time series observations. Since then, the MMA methodology has been applied to other regression models. For example, Wan, Zhang and Zou (2010) investigate non-nested models, Hansen and Racine (2012) study a jackknife averaging approach under heteroskedastic error settings, and Cheng and Hansen (2015) discuss factor-augmented regression models. Other works on the time series framework include Zhang, Wan and Zou (2013) and Cheng, Ing and Yu (2015).

The main objective of this research is to deal with the important issue of weighting over different candidate models constructed using various explanatory variables included in the MIDAS model, because, *a priori*, we do not know the exact model specification. In this study, we also use the Vandermonde matrix of Almon (1965) to parameterize the weighting functions for the higher frequency explanatory variables. This enables us to estimate each candidate model for averaging with the OLS estimator, as is required for the original MMA estimator of Hansen (2007). In addition to its excellent numerical performance, we show that the proposed MMA estimator possesses the same asymptotic optimality properties considered in Hansen (2007), who focuses on the regression model with a cross-sectional data structure.

The rest of the paper proceeds as follows. Section 2 introduces the semiparametric MIDAS model, providing full details of the model's features. Section 3 deals with theoretical foundations for the development of the research. Section 4 conducts simulations to demonstrate the finite-sample performance of our method. Section 5 concludes the paper. All proofs are given in the Supplementary Material.

2. Model Set-up

2.1. Semiparametric MIDAS regression

Suppose we are interested in forecasting some variable y_{t+h}, observed only at discrete times $t-1,\ t,\ t+1,\ldots$, while data on a predictor variable, $x_t^{(m)}$, are observed m times between $t-1$ and t. In financial data applications, these predictors can be generated from high-frequency shocks and behave like an autoregressive conditional heteroskedastic (ARCH) process. We intend to use the current and lagged values of $x_t^{(m)}$ to forecast y_{t+h}, in which the m superscript makes explicit the higher sampling frequency of $x_t^{(m)}$ relative to y_{t+h}. In this paper, we focus on the case of $h=1$. Our proposed method can be extended to long-horizon forecasts.

For ease of illustration, we first discuss the model with one higher frequency explanatory variable:

$$y_{t+1} = \beta_0 + \frac{A(z)}{B(z)} x_t^{(m)} + \epsilon_{t+1}, \tag{2.1}$$

where $t = 0, 1, \ldots, n-1$ and $z = L^{1/m}$, and the two components of the transfer function, $A(z)/B(z)$, that is, $A(\cdot)$ and $B(\cdot)$, have no common zeros. Characterizing the asymptotic properties of y_{t+1} with the aforementioned transfer function is normal in the time series literature, including that on the widely known ARMA model. In fact, the same function has also been suggested by Pettenuzzo, Timmermann and Valkanov (2016) for their MIDAS approach to model first- and second-moment dynamics, and Bonino-Gayoso and García-Hiernaux (2021) discuss the relationship between MIDAS polynomials and transfer functions.

Note that the transfer function can be expressed as an infinite series:

$$\frac{A(z)}{B(z)} x_t^{(m)} = \sum_{k=0}^{\infty} \xi_k L^{k/m} x_t^{(m)}, \tag{2.2}$$

and the exact value of ξ_k is unknown to the empirical users. Accordingly, we need to use a finite number K to approximate the infinite high-frequency explanatory variable in practice; that is, $K-1$ is the maximum lag length for the included predictors, and there are K higher frequency observations used for forecasting. Following the MIDAS literature, we employ equal-interval higher frequency data $\mathbf{x}_t^{(m)}(K)$ to predict y_{t+h}, where

$$\mathbf{x}_t^{(m)}(K) = (x_t^{(m)}, x_{t-1/m}^{(m)}, x_{t-2/m}^{(m)}, \ldots, x_{t-(K-1)/m}^{(m)})'. \tag{2.3}$$

Thus, we can re-express the original model as

$$y_{t+1} = \beta_0 + \boldsymbol{\xi}'(K)\mathbf{x}_t^{(m)}(K) + g_{t+1}(K) + \epsilon_{t+1}, \tag{2.4}$$

where

$$\boldsymbol{\xi}(K) = (\xi_0, \xi_1, \ldots, \xi_{K-1})' \tag{2.5}$$

and

$$g_{t+1}(K) = \sum_{k=K}^{\infty} \xi_k L^{k/m} x_t^{(m)}. \tag{2.6}$$

This research proposes estimating the coefficient ξ semiparametrically by adopting the Almon polynomial suggested by Chen and Tsay (2011) to approximate the unknown weighting function $A(z)/B(z)$. The rationale is that each candidate model thus considered for averaging can be estimated with the OLS estimator as the original MMA estimator of Hansen (2007). We show later that the combination of the Almon lag polynomial and the MMA methodology of Hansen (2007) not only delivers promising numerical performance, but also possesses the same theoretical optimality properties obtained in Hansen (2007) for cross-sectional data.

The Almon approach is based on the following $K \times p$ $\mathbf{V}(K,p)$ Vandermonde matrix:

$$\mathbf{V}(K,p) = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2^2 & \cdots & 2^{p-1} \\ 1 & 3 & 3^2 & \cdots & 3^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K & K^2 & \cdots & K^{p-1} \end{bmatrix}, \tag{2.7}$$

which is employed to transfer the $(K \times 1)$ vector of higher frequency data $\mathbf{x}_t^{(m)}(K)$ as a lower frequency $(1 \times p)$ vector of transformed regressors:

$$(\tilde{\mathbf{x}}_t^{(m)})' = (\mathbf{x}_t^{(m)}(K))'\mathbf{V}(K,p). \tag{2.8}$$

We adopt the Vandermonde matrix to approximate ξ so that (2.4) can be rewritten as

$$y_{t+1} = \beta_0 + \boldsymbol{\alpha}'(p)\tilde{\mathbf{x}}_t^{(m)} + \tilde{g}_{t+1}(K,p) + \epsilon_{t+1}, \tag{2.9}$$

where

$$\tilde{g}_{t+1}(K,p) = g_{t+1}(K) + \left(\boldsymbol{\xi}'(K) - \boldsymbol{\alpha}'(p)\mathbf{V}'(K,p)\right)\mathbf{x}_t^{(m)}(K) \tag{2.10}$$

and

$$\boldsymbol{\alpha}(p) = (\alpha_0, \ldots, \alpha_{p-1})'. \tag{2.11}$$

There are several advantages to using the Almon lag structure. First, owing to the linearity nature shown in (2.8), we can estimate this model with the OLS estimator. Second, the OLS estimator of $\boldsymbol{\alpha}(p)$ can be written as

$$\sum_{t=0}^{n-1} \left(\tilde{\mathbf{x}}_t^{(m)}(\tilde{\mathbf{x}}_t^{(m)})'\right)^{-1} \left(\tilde{\mathbf{x}}_t^{(m)}\right) y_{t+1}, \tag{2.12}$$

where the lag coefficient is approximated by a polynomial of degree $p < K$, thereby reducing the number of parameters to be estimated from K to p. Third, it delivers excellent numerical performance under various configurations considered in our Monte Carlo simulations; see Section 4.

2.2. Multiple high-frequency regressors

We now discuss our main findings, where we are interested in a MIDAS model with infinite sets of mixed-frequency predictors.

Let $(y_{t+1}, X_t^{(m)})$ be a sample, where y_{t+1} is real valued, and $X_t^{(m)} = (1, x_{1,t}^{(m)}, x_{2,t}^{(m)}, \ldots)'$ is countably infinite. For multiple high-frequency regressor cases, we have:

$$\begin{aligned} y_{t+1} &= \beta_0 + \sum_{i=1}^{\infty} \frac{A_i(z)}{B_i(z)} x_{i,t}^{(m)} + \epsilon_{t+1} \\ &= \beta_0 + \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} \xi_{i,k} L^{k/m} x_{i,t}^{(m)} + \epsilon_{t+1} \\ &= \sum_{i=0}^{\infty} \beta_i(L^{1/m}) x_{i,t}^{(m)} + \epsilon_{t+1} =: \mu_t + \epsilon_{t+1}, \end{aligned} \tag{2.13}$$

where $x_{0,t}^{(m)} = 1$, $\beta_0(L^{1/m}) = \beta_0$, and $\beta_i(L^{1/m}) = A_i(L^{1/m})/B_i(L^{1/m})$ for $i \in \mathbb{N}$.

Before displaying our theoretical findings, we present the regularity conditions for our model.

Assumption 1. $\{\epsilon_t,\, t \geq 1\}$ *is a sequence of independent and identically distributed (i.i.d.) random variables with zero mean and finite variance,* σ^2*. Moreover,* ϵ_t *is independent of* μ_s*, for all* $t, s \in \mathbb{N}$*.*

Assumption 2.

1. *Each high-frequency* $\mathbf{x}_{i,t}^{(m)}$*, for all* $i \in \mathbb{N}$*, is weakly stationary.*
2. *For all* $i \in \mathbb{N}$*,* $B_i(z) = 1 + b_{i,1}z + \cdots$ *and* $A_i(z) = 1 + a_{i,1}z + \cdots$ *are nonzero for* $|z| < 1$*. Moreover,* $A_i(e^{i\lambda})$*, for all* $i \in \mathbb{N}$*, is nonzero for* $-\pi < \lambda \leq \pi$*.*
3. $E\mu_t^2 < \infty$*, for all* $t \in \mathbb{R}^+$*.*
4. $\sum_{t=0}^{n-1} \mu_t^2/n = O(1)$*.*

Assumptions 1 and 2 ensure that $(A_i(z)/B_i(z))x_{i,t}^{(m)}$ is weakly stationary for all i. Moreover, Assumption 1 allows the high-frequency variable $x_t^{(m)}$ to exhibit the characteristics depicted by autoregressive (AR) and ARCH processes. We choose Assumption 1 to cover these important time series processes, while requiring the noise to be independent of the regressors. This condition is imposed regularly in the time series literature.

Assumption 2 contains the regularity conditions used to establish the asymptotic properties of the in-sample fit of the averaging estimator. This assumption is essential for deriving the asymptotic optimality property for our proposed method. The assumption concerning the average of μ_t^2 can be found in Shao (1997) and Wan, Zhang and Zou (2010), and Wan, Zhang and Zou (2010, Example 1) explain why the assumption of the average of μ_t^2 is reasonable.

3. The Averaging Estimator

3.1. Transformed regressors

Because we only observe a finite sample size, a finite-dimensional approximating model is estimated in practice. Therefore, we consider a sequence of approximating models $q = 1, \ldots, Q$, where the qth model uses the first k_q elements of $X_t^{(m)}$, as $0 \leq k_1 < \cdots < k_Q$. The leading case sets $k_q = q$, for $q = 1, \ldots, Q$. The qth approximating model is

$$y_{t+1} = \sum_{i=0}^{k_q} \beta_i(L^{1/m}) x_{i,t}^{(m)} + e_{q,t} + \epsilon_{t+1}, \tag{3.1}$$

where the approximation error is $e_{q,t} = \sum_{i=k_q+1}^{\infty} \beta_i(L^{1/m}) x_{i,t}^{(m)}$.

As shown in Section 2, the use of the transfer function transforms each high-frequency explanatory variable into an infinite series itself. However, we only observe a finite number of K high-frequency observations. To illustrate this phenomenon, we need to re-express the model in (3.1) as

$$\begin{aligned} y_{t+1} &= \sum_{i=0}^{k_q} \beta_i(L^{1/m}) x_{i,t}^{(m)} + e_{q,t} + \epsilon_{t+1} \\ &= \beta_0 + \sum_{i=1}^{k_q} \boldsymbol{\xi}_i'(K_i) \mathbf{x}_{i,t}^{(m)}(K_i) + E_{q,t}(K) + \epsilon_{t+1}, \end{aligned} \tag{3.2}$$

where $\mathbf{x}_{i,t}^{(m)}(K_i) = (x_{i,t}^{(m)}, x_{i,t-1/m}^{(m)}, \ldots, x_{i,t-(K_i-1)/m}^{(m)})'$, $\boldsymbol{\xi}_i(K_i) = (\xi_{i,0}, \xi_{i,1}, \ldots, \xi_{i,K_i-1})'$, and $E_{q,t}(K) = e_{q,t} + \sum_{i=1}^{k_q} \sum_{j=K_i}^{\infty} \xi_{i,j} L^{j/m} x_{i,t}^{(m)}$.

As mentioned in Section 2.1, we adopt the Vandermonde matrix to approximate ξ, yielding

$$y_{t+1} = \beta_0 + \sum_{i=1}^{k_q} \boldsymbol{\alpha}_i'(p_i) \tilde{\mathbf{x}}_{i,t}^{(m)} + \tilde{E}_{q,t}(K) + \epsilon_{t+1}, \tag{3.3}$$

where $\tilde{E}_{q,t}(K) = E_{q,t}(K) + \sum_{i=1}^{k_q} \boldsymbol{\xi}_i'(K_i) \mathbf{x}_{i,t}^{(m)}(K_i) - \sum_{i=1}^{k_q} \boldsymbol{\alpha}_i'(p_i) \tilde{\mathbf{x}}_{i,t}^{(m)}$. Remember that $\tilde{\mathbf{x}}_{i,t}^{(m)}$ is defined in (2.8), and it has K_i observations and a corresponding degree p_i; that is, the sequence $\{(K_i, p_i)\}$ is indexed by i. In matrix notation, we have

$$Y = \tilde{\mathbf{X}}_q \Gamma_q + \tilde{\mathbf{E}}_q + \boldsymbol{\epsilon}, \tag{3.4}$$

where $\Gamma_q' = (\beta_0, \boldsymbol{\alpha}_1'(p_1), \ldots, \boldsymbol{\alpha}_{k_q}'(p_{k_q}))$ is a $1 \times (1 + \sum_{i=1}^{k_q} p_i)$ vector of parameters to be estimated,

$$\tilde{\mathbf{E}}_q = (\tilde{E}_{q,1}(K), \ldots, \tilde{E}_{q,n}(K))', \tag{3.5}$$

and $\tilde{\mathbf{X}}_q = (\mathbf{1}_{n\times 1}, \tilde{\mathbf{X}}_1^{(m)}, \ldots, \tilde{\mathbf{X}}_{k_q}^{(m)})$ is the $n \times (1 + \sum_{i=1}^{k_q} p_i)$ matrix, where $\tilde{\mathbf{X}}_i^{(m)} = (\tilde{\mathbf{x}}_{i,0}^{(m)}, \tilde{\mathbf{x}}_{i,1}^{(m)}, \ldots, \tilde{\mathbf{x}}_{i,n-1}^{(m)})'$ is the $n \times p_i$ matrix, for $i = 1, \ldots, k_q$, and $\mathbf{1}_{n\times 1} = (1, 1, \ldots, 1)'$ is an $n \times 1$ matrix.

Before estimating the parameters in (3.4), we need to define $\mathbf{U}_q$

$$\mathbf{U}_q := \begin{bmatrix} 1 & \mathbf{0}_{1\times p_1} & \mathbf{0}_{1\times p_2} & \cdots & \mathbf{0}_{1\times p_{k_q}} \\ \mathbf{0}_{K_1\times 1} & \mathbf{V}(K_1, p_1) & \mathbf{0}_{1\times p_2} & \cdots & \mathbf{0}_{K_1\times p_{k_q}} \\ \mathbf{0}_{K_2\times 1} & \mathbf{0}_{K_2\times p_1} & \mathbf{V}(K_2, p_2) & \cdots & \mathbf{0}_{K_2\times p_{k_q}} \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{0}_{K_{k_q}\times 1} & \mathbf{0}_{K_{k_q}\times p_1} & \mathbf{0}_{K_{k_q}\times p_2} & \cdots & \mathbf{V}(K_{k_q}, p_{k_q}) \end{bmatrix}, \tag{3.6}$$

which is a $(1 + \sum_{i=1}^{k_q} K_i) \times (1 + \sum_{i=1}^{k_q} p_i)$ matrix. By the definition of (2.8), we now obtain

$$\mathbf{X}_q \mathbf{U}_q = \tilde{\mathbf{X}}_q, \tag{3.7}$$

where

$$\mathbf{X}_q = (\mathbf{1}_{n\times 1}, \mathbf{X}_1^{(m)}, \ldots, \mathbf{X}_{k_q}^{(m)}) \tag{3.8}$$

is the $n \times (1 + \sum_{i=1}^{k_q} K_i)$ matrix, with

$$\mathbf{X}_i^{(m)} = (\mathbf{x}_{i,0}^{(m)}(K_i), \mathbf{x}_{i,1}^{(m)}(K_i), \ldots, \mathbf{x}_{i,n-1}^{(m)}(K_i))', \tag{3.9}$$

for $i = 1, \ldots, k_q$.

3.2. An OLS-based model averaging estimator

We now introduce our estimator. Note that this research adopts an OLS-based model averaging estimator.

Let $Q = Q_n \le n$ be an integer for which $\mathbf{X}_Q$ have full rank. Assume for the moment that $\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q$ is invertible, and the least squares estimator of Γ_q is

$$\hat{\Gamma}_q = \left(\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q\right)^{-1} (\tilde{\mathbf{X}}_q)'Y, \tag{3.10}$$

for all $q \leq Q$. The following lemma verifies the invertible property of $\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q$ for all $q \leq Q$.

Lemma 1. *Suppose that $\mathbf{X}_Q$ have full rank and that K_i, for $i \in \mathbb{N}$, are finite numbers. $\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q$ is then invertible for all $q \leq Q$.*

We now introduce the model averaging estimator. Let $\mathbf{w} = (w_1, \ldots, w_Q)'$ be a weight vector in the unit simplex of $\mathbb{R}^Q$:

$$\mathcal{H}_n = \left\{ \mathbf{w} \in [0,1]^Q : \sum_{q=1}^{Q} w_q = 1 \right\}.$$

A model averaging estimator of Γ_Q is

$$\hat{\Gamma} = \sum_{q=1}^{Q} w_q \begin{pmatrix} \hat{\Gamma}_q \\ \mathbf{0} \end{pmatrix}. \tag{3.11}$$

In the qth approximating model of (3.2), we denote $\boldsymbol{\mu}_q = \tilde{\mathbf{X}}_q \Gamma_q$, so that

$$\boldsymbol{\mu} = \boldsymbol{\mu}_q + \tilde{\mathbf{E}}_q, \tag{3.12}$$

where $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_{n-1})'$ and $\tilde{\mathbf{E}}_q$ is defined in (3.5). To estimate $\boldsymbol{\mu}$ in the qth approximating model, we proceed as follows:

$$\hat{\boldsymbol{\mu}}_q = \tilde{\mathbf{X}}_q \hat{\Gamma}_q = \mathbf{P}_q Y, \quad \text{where} \quad \mathbf{P}_q = \tilde{\mathbf{X}}_q \left(\tilde{\mathbf{X}}_q' \tilde{\mathbf{X}}_q \right)^{-1} \tilde{\mathbf{X}}_q'. \tag{3.13}$$

The model averaging estimate of $\boldsymbol{\mu}$ is then computed as

$$\hat{\boldsymbol{\mu}}(\mathbf{w}) = \tilde{\mathbf{X}}_Q \hat{\Gamma} = \mathbf{P}(\mathbf{w}) Y, \tag{3.14}$$

where

$$\mathbf{P}(\mathbf{w}) = \sum_{q=1}^{Q} w_q \mathbf{P}_q \tag{3.15}$$

is the implied "hat" matrix.

To derive the asymptotic properties of the aforementioned estimator, we extend the results of Hansen (2007, Lemma 1) under the MIDAS model to investigate the properties of $\mathbf{P}(\mathbf{w})$ above. Indeed, $\mathbf{P}(\mathbf{w})$ plays an important role in computing the penalty term for the selected weights considered here. The difficulty of deriving the properties of $\mathbf{P}(\mathbf{w})$ hinges on the observation that the transformed regressors build on the Vandermonde matrix. We thus need to show the invertible property of $\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q$ in Lemma 1 before we can clarify the properties of $\mathbf{P}(\mathbf{w})$ in Lemma 2. Before showing the details of Lemma 2, we denote $\lambda_{\max}(A)$ as the largest eigenvalue of A, and define

$$\Upsilon_Q^{(p)} = \begin{bmatrix} 1+\sum_{i=1}^{k_1} p_i & 1+\sum_{i=1}^{k_1} p_i & 1+\sum_{i=1}^{k_1} p_i & \cdots & 1+\sum_{i=1}^{k_1} p_i \\ 1+\sum_{i=1}^{k_1} p_i & 1+\sum_{i=1}^{k_2} p_i & 1+\sum_{i=1}^{k_2} p_i & \cdots & 1+\sum_{i=1}^{k_2} p_i \\ 1+\sum_{i=1}^{k_1} p_i & 1+\sum_{i=1}^{k_2} p_i & 1+\sum_{i=1}^{k_3} p_i & \cdots & 1+\sum_{i=1}^{k_3} p_i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1+\sum_{i=1}^{k_1} p_i & 1+\sum_{i=1}^{k_2} p_i & 1+\sum_{i=1}^{k_3} p_i & \cdots & 1+\sum_{i=1}^{k_Q} p_i \end{bmatrix}. \tag{3.16}$$

Lemma 2. *Suppose that $\mathbf{X}_Q$ have full rank and that K_i, for $i \in \mathbb{N}$, are finite numbers. We then have*

(i) $\mathrm{tr}\,(\mathbf{P}(\mathbf{w})) = 1 + \sum_{q=1}^{Q}\sum_{i=1}^{k_q} w_q p_i \equiv k^{(p)}(\mathbf{w})$;

(ii) $\mathrm{tr}\,(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})) = \sum_{m=1}^{Q}\sum_{l=1}^{Q} w_m w_l \left[1 + \sum_{i=1}^{k_{\min\{m,l\}}} p_i\right] = \mathbf{w}'\Upsilon_Q^{(p)}\mathbf{w}$;

(iii) $\lambda_{\max}(\mathbf{P}(\mathbf{w})) \le 1$.

3.3. The Mallows criterion

This subsection presents the proposed MMA approach, and shows it possesses the same asymptotic optimality as that of Hansen (2007) for a cross-sectional data structure.

For the model averaging principle, the main problem is how to select the weights for all candidate models. Theoretically, we are concerned with a real number representing a measure of the "cost" associated with a particular decision. An acceptable function for depicting such a cost is the mean squared error (MSE) of a particular estimator. Here, we express the MSE associated with such a model averaging estimator, and then present its statistical properties.

Define the sum of the squared error as

$$L_n(\mathbf{w}) = (\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu})'(\hat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}), \tag{3.17}$$

and let

$$R_n(\mathbf{w}) = E[L_n(\mathbf{w})|\mathbf{X}^{(m)}], \tag{3.18}$$

where $\mathbf{X}^{(m)} = \{\mathbf{x}_0^{(m)}, \ldots, \mathbf{x}_{n-1}^{(m)}\}$. Note that

$$R_n(\mathbf{w}) = |\boldsymbol{\mu}'(\mathbf{I} - \mathbf{P}(\mathbf{w}))(\mathbf{I} - \mathbf{P}(\mathbf{w}))\boldsymbol{\mu}| + \sigma^2 \mathrm{tr}\,(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})), \tag{3.19}$$

so that

$$R_n(\mathbf{w}) \geq |\boldsymbol{\mu}' (\mathbf{I} - \mathbf{P}(\mathbf{w})) (\mathbf{I} - \mathbf{P}(\mathbf{w})) \boldsymbol{\mu}|, \tag{3.20}$$

and

$$R_n(\mathbf{w}) \geq \sigma^2 \mathrm{tr} (\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})) . \tag{3.21}$$

We now consider the asymptotic properties of the MMA estimator for the MIDAS model. The Mallows criterion for the model averaging estimator is

$$C_n(\mathbf{w}) =(Y - \tilde{\mathbf{X}}_Q\hat{\Gamma})'(Y - \tilde{\mathbf{X}}_Q\hat{\Gamma}) + 2\sigma^2 \mathrm{tr} (\mathbf{P}(\mathbf{w})) . \tag{3.22}$$

For the penalty term, $\mathrm{tr} (\mathbf{P}(\mathbf{w}))$, in (3.22), by item (i) of Lemma 2, the Mallows criterion becomes

$$\begin{aligned} C_n(\mathbf{w}) &= (Y - \tilde{\mathbf{X}}_Q\hat{\Gamma})'(Y - \tilde{\mathbf{X}}_Q\hat{\Gamma}) + 2\sigma^2 k^{(p)}(\mathbf{w}) \\ &= \mathbf{w}'\Psi'\Psi\mathbf{w} + 2\sigma^2\mathbf{K}'\mathbf{w}, \end{aligned} \tag{3.23}$$

where $k^{(p)}(\mathbf{w})$ is defined in Lemma 2 and represents the effective number of parameters. Here, $\hat{\boldsymbol{\epsilon}}_q$ denotes the $n \times 1$ residual vector for the qth model, where

$$\Psi = (\hat{\boldsymbol{\epsilon}}_1, \ldots, \hat{\boldsymbol{\epsilon}}_Q) \tag{3.24}$$

is the $n \times Q$ matrix collection of these residuals, and

$$\mathbf{K} = \left(1 + \sum_{i=1}^{k_1} p_i, \ldots, 1 + \sum_{i=1}^{k_Q} p_i\right)' \tag{3.25}$$

is the $Q \times 1$ vector of the number of parameters in the Q models. The Mallows criterion is used to select the weight vector $\mathbf{w}$.

There are two justifications for using the Mallows criterion. The following proposition expresses the first justification that the Mallows criterion $C_n(\mathbf{w})$ is an unbiased estimate of the expected squared error plus a constant.

Proposition 1. *Suppose that $\mathbf{X}_Q$ have full rank and that K_i, for $i \in \mathbb{N}$, are finite numbers. Assume that Assumption* 1 *holds. We then have*

$$EC_n(\mathbf{w}) = EL_n(\mathbf{w}) + n\sigma^2,$$

which means that $C_n(\mathbf{w})$ is an unbiased estimator of the expected in-sample squared error plus a constant.

Remark 1. Proposition 1 shows that $C_n(\mathbf{w})$ is an unbiased estimator for the expected in-sample squared error plus a constant. This result is consistent with Lemma 3 of Hansen (2007).

We now discuss the second justification for using the Mallows criterion. Because σ^2 is unknown in practice, we need to estimate it using sample observations. We use the largest approximating model to estimate σ^2, and the estimator is

$$\hat{\sigma}_Q^2 = \frac{1}{n - (1 + \sum_{i=1}^{k_Q} p_i)} (Y - \tilde{\mathbf{X}}_Q \hat{\Gamma}_Q)'(Y - \tilde{\mathbf{X}}_Q \hat{\Gamma}_Q), \tag{3.26}$$

where Q corresponds to the largest approximating model, and $\hat{\Gamma}_Q$ is defined in (3.10). Therefore, we define the proposed MMA criterion as

$$\hat{C}_n(\mathbf{w}) = (Y - \tilde{\mathbf{X}}_Q \hat{\Gamma})'(Y - \tilde{\mathbf{X}}_Q \hat{\Gamma}) + 2\hat{\sigma}_Q^2 k^{(p)}(\mathbf{w}), \tag{3.27}$$

and we select $\mathbf{w}$ using the following criterion:

$$\hat{\mathbf{w}} = \underset{\mathbf{w} \in \mathcal{H}_n}{\operatorname{argmin}} \, \hat{C}_n(\mathbf{w}), \tag{3.28}$$

where $\hat{\mathbf{w}}$ is the empirical Mallows selected weight vector.

The following theorem proves that the Mallows weight vector is asymptotically equivalent to the infeasible optimal weight vector, and that the empirical Mallows weight vector asymptotically minimizes the squared error. We emphasize here that these findings differ from those of Hansen (2007), because we need to transfer the higher frequency explanatory variables to match the lower frequency of the dependent variable. Therefore, the result in Theorem 1 is new, and is one possible solution to the MIDAS model using the MMA criterion.

Theorem 1. *Suppose that $\mathbf{X}_Q$ have full rank and that Q is chosen as a function of n, so that $Q \to \infty$ and $k_Q \to \infty$ as $n \to \infty$. Assume the following:*

1. *Assumptions 1 and 2 hold.*

2. *For some fixed integer $1 \leq N < \infty$, as $n \to \infty$:*

$$\Xi_n^{-2N} Q \sum_{q=1}^{Q} \left[R_n(\mathbf{w}_q^0) \right]^N \to 0, \tag{3.29}$$

   where $\Xi_n = \inf_{\mathbf{w} \in \mathcal{H}_n} R_n(\mathbf{w})$, and $\mathbf{w}_q^0$ is a $Q \times 1$ vector, in which the qth element is one and the others are zeros.

3. *For some fixed integer $N < \infty$,*

$$E[\epsilon_t^{4N}] < \infty. \tag{3.30}$$

4. *As* $n \to \infty$,

$$\frac{\left(1+\sum_{i=1}^{k_Q} p_i\right)^2}{n} = O(1). \tag{3.31}$$

We then have that as $n \to \infty$,

$$\frac{L_n(\hat{\mathbf{w}})}{\inf_{\mathbf{w}\in\mathcal{H}_n} L_n(\mathbf{w})} \xrightarrow{p} 1,$$

where $\hat{\mathbf{w}} = \operatorname{argmin}_{\mathbf{w}\in\mathcal{H}_n} \hat{C}_n(\mathbf{w})$, *and* $\hat{C}_n(\mathbf{w})$ *is defined in* (3.27).

Remark 2. Hansen (2007, Thm. 1) and Theorem 1 have certain aspects in common. First, Condition (3.30) places a bound on the conditional moments. Second, Condition (3.29) means that, because of $\Xi_n \to \infty$, there is no finite approximating model for which the bias is zero. Thus, the condition in (3.29) is the same as the one used in Wan, Zhang and Zou (2010). Additionally, Assumption 2 and (3.31) guarantee that the asymptotic optimality property holds, in which the condition in (3.31) places a constraint on the number of regressors in the largest approximating model. Generally speaking, we adopt the idea of Wan, Zhang and Zou (2010) to prove Theorem 1 of this study.

4. Simulation Study

In this section we employ the following DGP for the Monte Carlo experiments:

$$y_{t+1} = \beta_0 + \sum_{i=1}^{Q'} \beta_i \mathcal{B}_i(L^{1/m}; \boldsymbol{\theta}_i) x_{i,t}^{(m)} + \epsilon_{t+1}, \quad t = 0, 1, \ldots, n-1, \tag{4.1}$$

where $Q' \in \mathbb{N}$, and $\{\epsilon_t, t \in \mathbb{Z}\}$ is *i.i.d.* with $N(0, \sigma^2)$. Here, for $i = 1, 2, \ldots, 12$, $\mathcal{B}_i(L^{1/m}; \boldsymbol{\theta}_i) = \sum_{k=0}^{K_i'-1} B_i(k; \boldsymbol{\theta}_i) L^{k/m}$ and β_i measures the aggregate impact $x_{i,t}^{(m)}$ on y_{t+1}, provided that $\sum_{k=0}^{K_i'-1} B_i(k; \boldsymbol{\theta}_i) = 1$. In other words, we impose a restriction that the polynomial weights add up to one. Note that our theoretical findings do not impose any restriction on the polynomial weights.

We also state that each high-frequency explanatory variable is an infinite sequence, by design; that is, $K_i' = \infty$, for $i = 1, 2, \ldots, 12$. However, for ease of simulation, we choose $K' = K_i' = 100$ and $B_i(k; \boldsymbol{\theta}_i) = B(k; \boldsymbol{\theta})$, for $i = 1, \ldots, 12$ and $k = 0, 1, \ldots, 99$. To mimic the scenario usually encountered in practice, we use a finite number K, where $K < K' = 100$, to approximate the 100 ($= K'$) observations of each high-frequency explanatory variable.

Note that ϵ_t is sampled at a low frequency with sample size n, whereas the regressors $x_{i,t}^{(m)}$, for $i = 1, \ldots, 12$, are sampled 100 ($= K'$) times between t and $t-1$ from

$$x_{i,s}^{(m)} = 0.5 x_{i,s-1/m}^{(m)} + e_{i,s} \quad \text{for} \quad i = 1, 2, \ldots, 12, \tag{4.2}$$

where $s = 1/m, 2/m, \ldots, 100/m$ and $e_{i,s} \overset{i.i.d.}{\sim} N(0,1)$, for $i = 1, 2, \ldots, 12$. Thus, each higher frequency sample size is $100 \times n$. In the following Monte Carlo simulations, the lag coefficients in $B(k; \boldsymbol{\theta})$ adopt a two-parameter ($\boldsymbol{\theta} = (\theta_1, \theta_2)$) exponential Almon lag polynomial:

$$B(k; \boldsymbol{\theta}) = \frac{\exp\{\theta_1(k+1) + \theta_2(k+1)^2\}}{\sum_{k=0}^{99} \exp\{\theta_1(k+1) + \theta_2(k+1)^2\}}, \tag{4.3}$$

so that $\sum_{k=0}^{99} B(k; \boldsymbol{\theta}) = 1$. We consider two weighting schemes: $(\theta_1, \theta_2) = (7 \times 10^{-4}, -5 \times 10^{-2})$ corresponds to a fast decaying pattern, and $(\theta_1, \theta_2) = (7 \times 10^{-4}, -6 \times 10^{-3})$ reflects slowly decaying weights. The number of trials in each simulation is 10,000.

4.1. An estimator for the aggregate impact parameter

This subsection illustrates the performance of using the Almon polynomial method as a weighting function for higher frequency observations. We set $Q' = 1$, and (4.1) becomes

$$y_{t+1} = \beta_0 + \beta_1 \mathcal{B}(L^{1/m}; \boldsymbol{\theta}) x_t^{(m)} + \epsilon_{t+1}, \quad t = 0, 1, \ldots, n-1. \tag{4.4}$$

We focus on the accuracy of estimating the aggregate impact of each high-frequency variable; the results are displayed in Tables 1 and 2.

Recall from (3.7) that $\mathbf{X}_q \mathbf{U}_q = \tilde{\mathbf{X}}_q$. Note that the estimators of the aggregate impact $\boldsymbol{\beta} := (\beta_0, \beta_1)$ adopt the following formula

$$\hat{\boldsymbol{\beta}} = \mathbf{J} \mathbf{U}_1 \hat{\Gamma}_1, \tag{4.5}$$

where $\hat{\Gamma}_1 = \left(\tilde{\mathbf{X}}_1' \tilde{\mathbf{X}}_1 \right)^{-1} (\tilde{\mathbf{X}}_1)' Y$ and

$$\mathbf{J} = \begin{bmatrix} 1 \ \mathbf{0}_{1 \times K} \\ 0 \ \mathbf{1}_{1 \times K} \end{bmatrix}, \tag{4.6}$$

with $\mathbf{1}_{t \times k}$ as an all-ones matrix in which every element is equal to one. We report the mean and root of the mean squared errors (RMSE) of the estimated coefficients, (β_0, β_1), proposed by (4.5), based on 10,000 replications.

Tables 1 and 2 list the simulation results for two decaying patterns, in which every decaying pattern contains $K = 14, 34$ and $p = 4, 5$. These tables clearly show that the OLS estimators $(\hat{\beta}_0, \hat{\beta}_1)$ proposed by (4.5) are very close to the true values. Because our aggregate estimators are evaluated using (4.5), they are estimated without knowing the true weighting function. Moreover, each estimated coefficient's RMSE is found to be monotonically decreasing with the sample span n, supporting that our estimator possesses well-behaved asymptotic properties. We also document that the aforementioned RMSE increases with the value of K, but we do not find such a pattern for the magnitude of p. These results reveal that our OLS-based estimator provides excellent performance, even though the data are generated using a nonlinear exponential Almon lag model.

4.2. The relative performance of the MMA estimator

We now investigate the finite-sample performance of the proposed model averaging estimator, and compare the performance of the MMA method to that of alternative OLS-based estimators under the DGP specified in (4.1) with $Q' = 12$; that is, the DGP in this subsection is

$$y_{t+1} = \beta_0 + \sum_{i=1}^{12} \beta_i \mathcal{B}(L^{1/m}; \boldsymbol{\theta}) x_{i,t}^{(m)} + \epsilon_{t+1}, \quad t = 0, 1, \ldots, n-1. \tag{4.7}$$

Recall that $\mathbf{X}_Q$ is the original data. The first alternative OLS-based method is built on the largest candidate model, but it does not adopt the Almon polynomial. This model is called the unrestricted MIDAS model, or UR-OLS(Q), because it treats higher frequency data $x_{i,t}^{m}$ as separate regressors, and can be estimated as

$$\hat{\boldsymbol{\mu}}^{\mathrm{UR-OLS(Q)}} = \mathbf{X}_Q \left(\mathbf{X}_Q' \mathbf{X}_Q\right)^{-1} \mathbf{X}_Q' Y. \tag{4.8}$$

Note that Foroni, Marcellino and Schumacher (2015) introduce the unrestricted MIDAS model.

In addition to the proposed MMA estimator, we analyze three other model averaging methods to illustrate the advantage of using Almon polynomials for the MIDAS model. We first implement a simple combination method that assigns an equal weighting function to the higher frequency observations; we refer to this as the equal weights (EW) method. The estimate of $\boldsymbol{\mu}$ generated is $\hat{\boldsymbol{\mu}}(\mathbf{EW})$.

The second alternative method is the smoothed AIC (S-AIC) introduced by Buckland, Burnham and Augustin (1997) and embraced by Hjort and Claeskens (2003), in which the weights of the model averaging estimator (3.11) are written

Table 1. OLS estimators for the MIDAS model: Fast decay.

	$K=14$, $p=4$		$K=14$, $p=5$		$K=34$, $p=4$		$K=34$, $p=5$	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
Panel A: Time span $n=100$								
Mean	0.5002	3.0196	0.5001	3.0029	0.5005	2.9205	0.5001	3.0182
RMSE	0.1026	0.2065	0.1028	0.2066	0.1135	0.3541	0.1045	0.3204
Panel B: Time span $n=300$								
Mean	0.5006	3.0180	0.5005	3.0014	0.5004	2.9170	0.5005	3.0157
RMSE	0.0580	0.1185	0.0579	0.1174	0.0638	0.2100	0.0585	0.1783
Panel C: Time span $n=500$								
Mean	0.5000	3.0178	0.5000	3.0013	0.4997	2.9180	0.4999	3.0170
RMSE	0.0456	0.0921	0.0455	0.0904	0.0502	0.1691	0.0460	0.1373
Panel D: Time span $n=1000$								
Mean	0.5003	3.0178	0.5003	3.0012	0.5002	2.9195	0.5002	3.0183
RMSE	0.0320	0.0658	0.0319	0.0634	0.0353	0.1320	0.0324	0.0983

Notes: This table presents the mean and root of the mean squared errors (RMSE) of the estimated coefficients proposed by (4.5). The parameters for the DGP in (4.4) are $\sigma = 1$, $(\theta_1, \theta_2) = (7 \times 10^{-4}, -5 \times 10^{-2})$, and $(\beta_0, \beta_1) = (0.5, 3.0)$. The simulation studies are based on 10,000 replications.

Table 2. OLS estimators for the MIDAS model: Slow decay.

	$K=14$, $p=4$		$K=14$, $p=5$		$K=34$, $p=4$		$K=34$, $p=5$	
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
Panel A: Time span $n=100$								
Mean	0.5004	2.7035	0.5003	2.7078	0.5001	3.0171	0.5002	3.0009
RMSE	0.1043	0.3629	0.1047	0.3605	0.1025	0.3133	0.1027	0.3146
Panel B: Time span $n=300$								
Mean	0.5006	2.7020	0.5007	2.7064	0.5004	3.0140	0.5005	2.9976
RMSE	0.0593	0.3208	0.0593	0.3171	0.0578	0.1755	0.0577	0.1750
Panel C: Time span $n=500$								
Mean	0.4999	2.7017	0.5000	2.7062	0.4999	3.0152	0.4999	2.9989
RMSE	0.0464	0.3119	0.0465	0.3078	0.0454	0.1348	0.0454	0.1341
Panel D: Time span $n=1000$								
Mean	0.5002	2.7018	0.5003	2.7062	0.5002	3.0164	0.5003	3.0000
RMSE	0.0326	0.3051	0.0326	0.3008	0.0319	0.0964	0.0318	0.0950

Notes: This table presents the mean and root of the mean squared errors (RMSE) of the estimated coefficients proposed by (4.5). The parameters for the DGP in (4.4) are $\sigma = 1$, $(\theta_1, \theta_2) = (7 \times 10^{-4}, -6 \times 10^{-3})$, and $(\beta_0, \beta_1) = (0.5, 3.0)$. The simulation studies are based on 10,000 replications.

as

$$w_q(\mathrm{AIC}) = \frac{\exp\{(-1/2)\mathrm{AIC}_q\}}{\sum_{q=1}^{Q}\exp\{(-1/2)\mathrm{AIC}_q\}}, \tag{4.9}$$

where $\mathrm{AIC}_q = n\ln\hat{\sigma}_q^2 + 2q$, with

$$\hat{\sigma}_q^2 = \frac{1}{n-(1+\sum_{i=1}^{k_q} p_i)}(Y-\tilde{\mathbf{X}}_q\hat{\Gamma}_q)'(Y-\tilde{\mathbf{X}}_q\hat{\Gamma}_q), \tag{4.10}$$

and $\tilde{\mathbf{X}}_q$ is defined in (3.7).

The third alternative method is the smoothed BIC (S-BIC) proposed by Hansen (2007). As mentioned in Hansen (2007), the S-BIC is a simplified form of the Bayesian model averaging estimator, with the weights

$$w_q(\mathrm{BIC}) = \frac{\exp\{(-1/2)\mathrm{BIC}_q\}}{\sum_{q=1}^{Q}\exp\{(-1/2)\mathrm{BIC}_q\}}, \tag{4.11}$$

where $\mathrm{BIC}_q = n\ln\hat{\sigma}_q^2 + \ln(n)q$.

Tables 3 and 4 present the simulation results of the abovementioned five estimators under two decaying patterns in the higher frequency observations. The focus of our experimental design is on the relative MSE among the five estimators for estimating the true $\boldsymbol{\mu}$. The configurations of these tables are $n = 300, 500, 1000$, $K = K_i = 14, 34$, and $p = p_i = 4, 5$, for $i = 1, 2, \ldots, Q$.

The condition in (3.29) implies that there is no finite approximating model for which the bias is zero. Accordingly, the candidate models selected are all biased. To fulfill this requirement, we consider only the first six explanatory variables in the DGP to be included in the candidate models. Without loss of generality and to limit the computational burden, all the candidate models selected are nested with each other. Given such an experimental design, the model with the first six regressors becomes the largest candidate model with $Q = 6$. Under this circumstance, the weight vector for the EW method is $(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$.

The simulation results for the MMA, UR-OLS(Q), EW, S-AIC, and S-BIC methods appear in the second, third, fourth, fifth, and sixth columns, respectively, of Tables 3 and 4. For ease of comparison, we compute the relative performance of the MMA estimators against that of the other four approaches in these tables. More specifically, the basis of comparison is the relative difference defined by ("X"-MSE(MMA))/MSE(MMA), where X can be the MSE of UR-OLS(Q), EW, S-AIC, or S-BIC. They are shown in the seventh, eighth, ninth, and tenth columns, respectively.

We find that the MMA technique systematically outperforms UR-OLS(Q), except for the cases of fast decay with $(n, K, p) = (1000, 34, 4)$, revealing the advantage of transferring the high frequency variables using Almon weights. It is interesting to note that the EW method outperforms the UR-OLS(Q) method when the sample size is small, thus highlighting again the advantage of using the model averaging approach.

With regard to the performance among the family of model averaging estimators, we find that the MMA method outperforms the EW method. This observation supports the merit of our Theorem 1 in that the MMA method will pick an optimal weight for aggregating our candidate models. In addition, we observe that the performance of the proposed MMA method is very close to those of the S-AIC and S-BIC methods in most configurations considered in our experiments. Note that in the model selection literature, the Akaike and Mallows methods provide the asymptotic optimality of the out-of-sample forecast, from the seminal work of Ing and Wei (2005). However, we do not know whether, in the MIDAS model or the setting of Hansen (2007), the S-AIC method has the same asymptotic optimality as that of the MMA method. Thus, it could be fruitful to further explore the usefulness of the S-AIC and S-BIC methods under this setting.

There are another two findings to be drawn from Tables 3 and 4. First, the MMA estimator performs better for the slow decay scenario than for the fast one. Second, the MMA performs best among all methods when $p = 5$ and for the slow decay scenario. These results are consistent with the conditions imposed in Theorem 1, where all candidate models are biased, and the slow decay scenario is better at characterizing this pattern than is the fast decay one.

Furthermore, the simulation results in Tables 3 and 4 do not exhibit a systematic pattern related to the value of p, given $K = 14$ or 34. This pattern is consistent with the asymptotic properties of our estimators, which do not depend on the number of high-frequency variables, K, or the order of the Almon polynomials, p.

5. Conclusion

This research proposes an MMA approach to the well-known MIDAS model in order to fill a gap in the time series literature. We do so by establishing the asymptotic optimality of the MMA estimator under the MIDAS framework, where the high-frequency predictors can be weakly dependent or ARCH (GARCH) processes. Accordingly, we extend the coverage of Hansen (2007), who mainly considers a cross-sectional data structure. The Monte Carlo experiments

Table 3. Relative mean squared error: Fast decay.

	MSE					Relative difference (in percent)			
		(1)	(2)	(3)	(4)				
n	MMA	UR-OLS	EW	S-AIC	S-BIC	(1)	(2)	(3)	(4)
Panel A: $K_i = 14$ and $p_i = 4$ for $i = 1, \ldots, 6$									
300	1.0251	1.6526	2.5887	1.0238	1.0245	61.21	152.53	-0.13	-0.05
500	0.9159	1.2795	2.5747	0.9153	0.9153	39.70	181.12	-0.07	-0.06
1,000	0.8348	1.0012	2.5649	0.8345	0.8345	19.93	207.27	-0.03	-0.03
Panel B: $K_i = 14$ and $p_i = 5$ for $i = 1, \ldots, 6$									
300	1.0660	1.6526	2.5763	1.0681	1.0689	55.02	141.67	0.19	0.27
500	0.9323	1.2795	2.5592	0.9329	0.9329	37.24	174.51	0.06	0.07
1,000	0.8314	1.0012	2.5468	0.8315	0.8315	20.41	206.32	0.00	0.00
Panel C: $K_i = 34$ and $p_i = 4$ for $i = 1, \ldots, 6$									
300	1.6800	2.9636	3.1797	1.6761	1.6778	76.40	89.26	-0.24	-0.13
500	1.5934	2.0669	3.1835	1.5916	1.5916	29.71	99.80	-0.11	-0.11
1,000	1.5294	1.3939	3.1862	1.5289	1.5289	-8.86	108.33	-0.04	-0.04
Panel D: $K_i = 34$ and $p_i = 5$ for $i = 1, \ldots, 6$									
300	1.1517	2.9636	2.6538	1.1537	1.1545	157.32	130.42	0.17	0.24
500	1.0207	2.0669	2.6391	1.0211	1.0211	102.49	158.55	0.04	0.04
1,000	0.9228	1.3939	2.6285	0.9228	0.9228	51.05	184.83	-0.00	-0.00

Notes: This table presents the mean squared error for four model averaging methods that estimate the true $\boldsymbol{\mu}$. The parameters for the DGP in (4.7) are $\sigma = 2$, $(\theta_1, \theta_2) = (7 \times 10^{-4}, -5 \times 10^{-2})$, and $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}) = (0.5, 3.0, 2.8, 2.5, 2.0, 1.2, 1.0, 0.8, 0.6, 0.4, 0.3, 0.2, 0.1)$. The simulation studies are based on 10,000 replications. The relative difference is defined by ("X"-MSE(MMA))/MSE(MMA), where X can be the MSE of EW, S-AIC, or S-BIC, which appear in the sixth, seventh, and eighth columns, respectively.

demonstrate the promising performance of the MMA estimator relative to other estimators for the MIDAS models.

Note that the Almon polynomial method of the weighting function may also be useful for forecasting daily financial data. In fact, Mitchell (2020) employs the Almon polynomials to weight higher frequency data, yielding promising empirical results.

As with our MMA method and the U-MIDAS of Foroni, Marcellino and Schumacher (2015), the works of Ghysels, Sinko and Valkanov (2007) and Ghysels and Qian (2019) are also OLS-based MIDAS approaches. Further research should investigate the relative performance of these methods using theoretical research and numerical comparison in order to give us more information to design optimal weights for the model averaging approaches.

Table 4. Relative mean squared error: Slow decay

	MSE					Relative difference (in percent)			
		(1)	(2)	(3)	(4)				
n	MMA	UR-OLS	EW	S-AIC	S-BIC	(1)	(2)	(3)	(4)
Panel A: $K_i = 14$ and $p_i = 4$ for $i = 1, \ldots, 6$									
300	0.7353	1.4481	1.2630	0.7405	0.7506	96.93	71.76	0.71	2.07
500	0.6179	1.0427	1.2245	0.6185	0.6206	68.75	98.18	0.10	0.44
1,000	0.5295	0.7402	1.1968	0.5293	0.5294	39.79	126.00	-0.04	-0.03
Panel B: $K_i = 14$ and $p_i = 5$ for $i = 1, \ldots, 6$									
300	0.7958	1.4481	1.2820	0.8097	0.8187	81.97	61.10	1.74	2.88
500	0.6560	1.0427	1.2356	0.6598	0.6619	58.95	88.37	0.59	0.90
1,000	0.5483	0.7402	1.2014	0.5489	0.5489	35.00	119.10	0.10	0.10
Panel C: $K_i = 34$ and $p_i = 4$ for $i = 1, \ldots, 6$									
300	0.6222	2.8289	1.1611	0.6280	0.6370	354.63	86.59	0.93	2.38
500	0.5004	1.8163	1.1192	0.5013	0.5029	262.99	123.68	0.19	0.51
1,000	0.4088	1.0563	1.0890	0.4087	0.4088	158.37	166.36	-0.02	-0.01
Panel D: $K_i = 34$ and $p_i = 5$ for $i = 1, \ldots, 6$									
300	0.6710	2.8289	1.1689	0.6858	0.6939	321.61	74.21	2.21	3.41
500	0.5244	1.8163	1.1173	0.5287	0.5303	246.34	113.06	0.82	1.11
1,000	0.4129	1.0563	1.0802	0.4136	0.4137	155.81	161.59	0.17	0.18

Notes: This table presents the mean squared error for four model averaging methods that estimate the true $\boldsymbol{\mu}$. The parameters for the DGP in (4.7) are $\sigma = 2$, $(\theta_1, \theta_2) = (7 \times 10^{-4}, -6 \times 10^{-3})$, and $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}) = (0.5, 3.0, 2.8, 2.5, 2.0, 1.2, 1.0, 0.8, 0.6, 0.4, 0.3, 0.2, 0.1)$. The simulation studies are based on 10,000 replications. The relative difference is defined by ("X"-MSE(MMA))/MSE(MMA), where X can be the MSE of EW, S-AIC, or S-BIC, which appear in the sixth, seventh, and eighth columns, respectively.

It would also be interesting to address the inference problems after model averaging. A valid confidence interval for the MIDAS model using the MMA approach is certainly important for this literature. Another extension is to study the forecast averaging by extending the result in Hansen (2008) to the MIDAS model.

Supplementary Material

The online Supplementary Material shows that the MMA estimator for the MIDAS model using an Almon polynomial weight has the property of asymptotic optimality by extending the method of Wan, Zhang and Zou (2010). The proofs of Lemmas 1-2 and Proposition 1, which contain the properties of the OLS estimator, penalty term, and proposed MMA estimator, respectively, are also given.

Acknowledgments

We are grateful to Chu-Lan Kao and Chu-An Liu for their valuable discussions and comments. We are also thankful for the invaluable suggestions of the reviewer and editors of *Statistica Sinica*. The author, Wen-Jen Tsay, appreciates the financial support provided by Academia Sinica through the Academia Sinica Investigator Award AS-IA-107-H02.

References

Almon, S. (1965). The distributed lag between capital approximation and expenditures. *Econometrica* **33**, 178–196.

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

Aruoba, B. S. and Diebold, F. X. (2010). Real-time macroeconomic monitoring: Real activity, inflation, and interactions. *American Economic Review* **100**, 20–24.

Bonino-Gayoso, N. and García-Hiernaux, A. (2021). TF-MIDAS: A transfer function based mixed-frequency model. *Journal of Statistical Computation and Simulation* **91**, 1980–2017.

Buckland, S. T., Burnham, K. P. and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53**, 603–618.

Carriero, A., Clark, T. E. and Marcellino, M. (2015). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* **178**, 837–862.

Chen, Y. C. and Tsay, W. J. (2011). Forecasting commodity prices with mixed-frequency data: An OLS-based generalized ADL approach. SSRN Working Paper. Available from `http://ssrn.com/abstract=1782214`.

Cheng, X. and Hansen, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics* **186**, 280–293.

Cheng, T. C. F., Ing, C. K. and Yu, S. H. (2015). Toward optimal model averaging in regression models with time series errors. *Journal of Econometrics* **189**, 321–334.

Clements, M. P. and Galvao, A. B. (2008). Macroeconomic forecasting with mixed-frequency data. *Journal of Business and Economic Statistics* **26**, 546–554.

Clements, M. P. and Galvao, A. B. (2009). Forecasting us output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics* **24**, 1187–1206.

Engle, R. F., Ghysels, E. and Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *The Review of Economics and Statistics* **95**, 776–797.

Foroni, C., Marcellino, M. and Schumacher, C. (2015). U-MIDAS: MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* **178**, 57–82.

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

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics* **76**, 509–548.

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

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

Ghysels, E. and Qian, H. (2019). Estimating MIDAS regressions via OLS with polynomial parameter profiling. *Econometrics and Statistics* **9**, 1–16.

Ghysels, E. and Valkanov, R. (2012). Forecasting volatility with MIDAS. In *Handbook of Volatility Models and their Applications* (Edited by L. Bauwens, C. Hafner and S. Laurent), 383–401. John Wiley & Sons, Inc, Hoboken.

Giannone, D., Reichlin, L. and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics* **55**, 665–676.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica* **75**, 1175–1189.

Hansen, B. E. (2008). Least-squares forecast averaging. *Journal of Econometrics* **146**, 342–350.

Hansen, B. E. and Racine, J. S. (2012). Jackknife model averaging. *Journal of Econometrics* **167**, 38–46.

Hjort, N. L. and Claeskens, G. (2003). Frequentist model average estimators. *Journal of the American Statistical Association* **98**, 879–899.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science* **14**, 382–417.

Ing, C. K. and Wei, C. Z. (2005). Order selection for same-realization predictions in autoregressive processes. *The Annals of Statistics* **33**, 2423–2474.

Kuzin, V., Marcellino, M. and Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro area. *International Journal of Forecasting* **27**, 529–542.

Kuzin, V., Marcellino, M. and Schumacher, C. (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. *Journal of Applied Econometrics* **28**, 392–411.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **15**, 661–675.

Mitchell, R. (2020). Can daily financial data help forecast economic downturns? Unpublished paper.

Pettenuzzo, D., Timmermann, A. and Valkanov, R. (2016). A MIDAS approach to modeling first and second moment dynamics. *Journal of Econometrics* **193**, 315–334.

Raftery, A. E., Madigan, D. and Hoeting, J. A. (1997). Bayesian model averaging for regression models. *Journal of the American Statistical Association* **92**, 179–191.

Shao, J. (1997). An asymptotic theory for linear model selection (with discussion). *Statistica Sinica* **7**, 221–264.

Wan, A. T. K., Zhang, X. and Zou, G. (2010). Least squares model averaging by Mallows criterion. *Journal of Econometrics* **156**, 277–283.

Zhang, X., Wan, A. T. K. and Zou, G. (2013). Model averaging by jackknife criterion in models with dependent data. *Journal of Econometrics* **174**, 82–94.

Hsin-Chieh Wong

Department of Mathematics, National Central University, Taoyuan City 32001, Taiwan.
Department of Mathematical Sciences, National Chengchi University, Taipei 11601, Taiwan.

E-mail: wonghsinchieh@gmail.com

Wen-Jen Tsay

Institute of Economics, Academia Sinica, Taipei 11529, Taiwan.

E-mail: wtsay@econ.sinica.edu.tw

(Received July 2020; accepted February 2021)