Estimating MIDAS Regressions via OLS with Polynomial Parameter Profiling*

Eric Ghysels† Hang Qian‡
UNC Chapel Hill MathWorks

First draft: April, 2016
This version: September 10, 2016

Abstract

A typical MIDAS regression involves estimating parameters via nonlinear least squares, unless U-MIDAS is applied - which involves OLS - the latter being appealing when the sampling frequency differences are small. In this paper we propose to use OLS estimation of the MIDAS regression slope and intercept parameters combined with profiling the polynomial weighting scheme parameter(s). The use of Beta polynomials is particularly attractive for such an approach. The new procedure shares many of the desirable features of U-MIDAS, while it is not restricted to small sampling frequency differences.

Keywords: Mixed frequency data, MIDAS regressions, profile likelihood
JEL classification: C13, C22, C52, C53

*We thank Massimiliano Marcellino, Jurij Plazzi and Ross Valkanov for helpful discussions.
†UNC Chapel Hill, NC 27599, USA, email: `eghysels@email.unc.edu`
‡55 Centre Street, Natick, MA 01760, U.S.A. email: `matlabist@gmail.com`

1 Introduction

The idea of profiling likelihood functions has a long history in the statistics literature, see e.g. Patefield (1977), Barndorff-Nielsen (1983), Barndorff-Nielsen and Cox (1994, p. 94), among others. It is appealing both from a conceptional and computational point of view. Suppose that y_t are i.i.d. random variables, with density $f(y;\delta,\theta)$ where our objective is to estimate δ and θ. Given a sample of size T, the log-likelihood is $\mathcal{L}(\delta,\theta) = \sum_{t=1}^{T} \log f(y;\delta,\theta)$. In many situations the log-likelihood is difficult to maximize over the entire parameter space, but if we fix $\theta = \bar{\theta}$ the maximization of the log-likelihood with respect to only δ given $\bar{\theta}$ becomes fairly straightforward. In such cases it is attractive to rewrite the log-likelihood as $\mathcal{L}_{T,\delta}(\theta)$ - referred to as the profile log-likelihood function which is optimized with respect to δ for a given θ - and to reformulate the maximization problem as $(\hat{\delta},\hat{\theta}) = \underset{\theta}{\operatorname{argmax}} \mathcal{L}_{T,\delta}(\theta)$.

The profile likelihood not only offers computational attractive solutions, but can also be viewed as a function which shares many appealing properties of the standard likelihood function and is therefore a statistical object of interest to derive tests (see e.g. van der Vaart (2002) for a detailed discussion). Hence, the curvature of the profile log-likelihood can be used to construct asymptotic inference about δ. Profile log-likelihood functions are often used in situations where δ is a low-dimensional parameter of interest and θ a higher-dimensional nuisance parameter. This applies in particular to semiparametric settings. Murphy and van der Vaart (2000) give a general justification for using a semiparametric profile likelihood function as an inferential tool.

Recently, econometric models that take into account the information in datasets with data sampled at different frequencies have attracted substantial attention. Mixed-data sampling (MIDAS) regression models provide an attractive tool to handle time series data sampled at different frequencies. They are parsimonious specifications based on distributed lag polynomials, which flexibly deal with data sampled at different frequencies and provide a direct forecast of a low-frequency variable conditional of lagged low frequency and high frequency data (see e.g. Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Santa-Clara, and Valkanov (2006) and Clements and Galvão (2009), among others).[1] MIDAS regression models are typically estimated via nonlinear least squares (NLS), see e.g. Andreou, Ghysels, and Kourtellos (2010). One important exception is so called U-MIDAS, proposed by Foroni, Marcellino, and Schumacher (2015) who study a variant of MIDAS which does not resort to functional distributed lag polynomials and is particularly appealing for situations where the difference in sampling frequencies is small - a prominent example being a mixture of quarterly and monthly data with only three high frequency observations associated with each low frequency data point. There unrestricted regression cases, hence U-MIDAS, only involve OLS.

In this paper we propose a profile log-likelihood approach to MIDAS regressions where we let θ be the parameters governing the MIDAS polynomial and δ the intercept and slope parameters of the regression. The new estimation approach has many attractive properties - some that are totally unexpected and novel to both the MIDAS and profile log-likelihood literature. For a given $\theta = \bar{\theta}$, the profile approach reduces the estimation problem to a series of rather straightforward linear regression model OLS estimators $\hat{\delta}(\bar{\theta})$. Since MIDAS polynomial parameters θ are low-dimensional (as well) - sometimes even one-dimensional - it is fairly easy to select $\bar{\theta}$ across a comprehensive grid, which makes the estimation computationally simple.[2]

The remarkable and unexpected result is that in cases where U-MIDAS is attractive, we find that using our new estimation approach applied to standard MIDAS regressions is computationally *faster*, than the OLS approach of Foroni, Marcellino, and Schumacher (2015). One may wonder why this is the case. The insight is that U-MIDAS - say in a quarterly/monthly setting - is a regression model with say 3, 6 or 9 lags of monthly data and lagged quarterly data projected on current or future quarterly data whereas the profiling

[1]Recent surveys of the growing literature include Armesto, Engemann, and Owyang (2010), Andreou, Ghysels, and Kourtellos (2011) and Foroni and Marcellino (2013).

[2]Some of the early roots of the ideas developed in this paper appeared in Ghysels, Plazzi, and Valkanov (2011) who used profiling in a MIDAS quantile regression (the published version – Ghysels, Plazzi, and Valkanov (2016) –uses standard nonlinear estimation). Engle, Ghysels, and Sohn (2013) also used profiling to decide on lag selection in a GARCH-MIDAS model.

estimator only involves a *single* regressor (ignoring intercepts in both cases) of pre-filtered high frequency data (using the MIDAS polynomial given $\bar{\theta}$). Hence, U-MIDAS applies OLS to 3, 6 or 9 high frequency regressors and profiling applies OLS to a single regressor. The latter is computationally faster. Of course, U-MIDAS becomes unattractive when the mixtures involve a large number of high frequency data such as daily data in a quarterly prediction model. Our profile estimator remains computationally efficient, statistically appealing and above all is simple and avoids the numerical pitfalls of the sometimes ill-behaved NLS.

We also find that a correctly specified profile MIDAS model, has smaller R^2s than U-MIDAS regressions. However, the profile estimator has higher predictive power, in particular its out-of-sample forecast MSE is significantly lower. This suggests that U-MIDAS suffers from over-fitting. The unrestricted parameters can be trained to exhibit high in-sample R^2s, but cannot outperform a parsimonious model in out-of-sample prediction. This observation also applies to situations where U-MIDAS is not appealing. Namely, in the context of regression-based volatility forecasting we find that a so called Corsi model, which requires only OLS estimation, is outperformed by MIDAS with profiling in terms of out-of-sample predictions.

The paper is organized as follows. Section 2 introduces the MIDAS and U-MIDAS models of interest. Next we cover the asymptotics of profile and OLS estimators. A final section covers numerical efficiency.

2 MIDAS regressions and profiling

It will be convenient to focus on a mixture of two frequencies, respectively high and low. In terms of notation, $t = 1, \ldots, T$ indexes the low frequency time unit, and m is the number of times the higher sampling frequency appears in the same basic time unit (assumed fixed for simplicity). For example, for quarterly GDP growth and monthly indicators as explanatory variables, $m = 3$. The low frequency variable will be denoted by y_t^L, whereas a generic high frequency series will be denoted by $x_{t-j/m}^H$ where $t - j/m$ is the j^{th} (past) high frequency period with $j = 0, 1, \ldots$. For a quarter/month mixture one has x_t^H, $x_{t-1/3}^H$, $x_{t-2/3}^H$ as the last, second to last and first months of quarter t.

2.1 A Primer on MIDAS

MIDAS regressions are essentially tightly parameterized, reduced form regressions that involve processes sampled at different frequencies. The response to the higher-frequency explanatory variable is modeled using highly parsimonious distributed lag polynomials, to prevent the proliferation of parameters that might otherwise result, as well as the issues related to lag-order selection.

The basic single high frequency regressor MIDAS model for h-step-ahead (low frequency) forecasting, with high frequency data available up to x_t^H is given by:

$$y_{t+h}^L = a_h + b_h C(L^{1/m}; \theta_h) x_t^H + \varepsilon_{t+h}^L, \tag{1}$$

where $C(L^{1/m}; \theta) = \sum_{j=0}^{j_{max}-1} c(j; \theta) L^{j/m}$, and $C(1; \theta) = \sum_{j=0}^{j_{max}-1} c(j; \theta) = 1$. This model is often called DL-MIDAS, where DL refers to distributed lag. The parameterization of the lagged coefficients of $c(j; \theta)$ in a parsimonious way is one of the key MIDAS features. Various other parsimonious polynomial specifications have been considered, including (1) beta polynomial, (2) Almon lag polynomial specifications, (3) step functions, among others. Ghysels, Sinko, and Valkanov (2006) provide a detailed discussion.[3]

While it is not an absolute necessity, it is surely convenient if we restrict θ to a one-dimensional space. For this reason, it is worth to focus on the class of so called Beta polynomials introduced by Ghysels, Santa-Clara, and Valkanov (2006). It is based on the

[3]Various software packages including the MIDAS Matlab Toolbox (Ghysels (2013)), the R Package *midasr* (Ghysels, Kvedaras, and Zemlys (2016)) and EViews cover a variety of polynomial specifications.

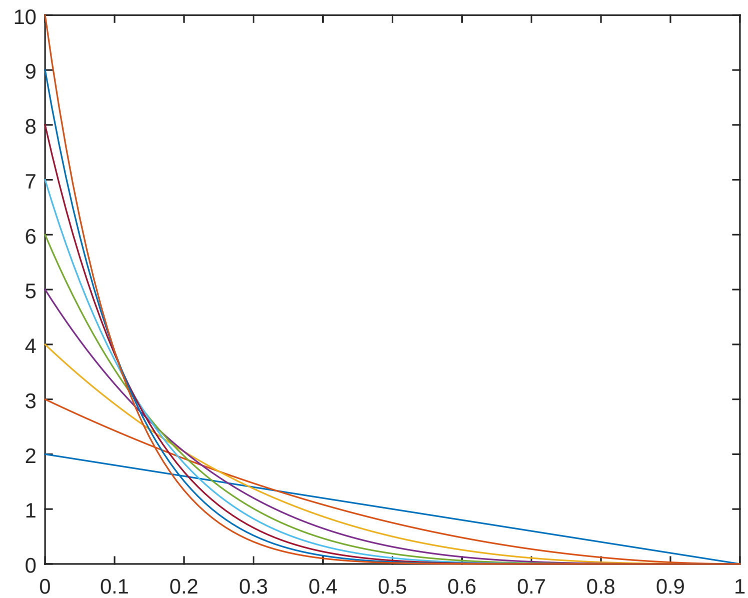

Figure 1: Plot of Beta polynomial appearing in equation (3) with $\theta_1 = 1$ and $\theta_2 = 2, \ldots, 10$

Beta probability density function which involves two parameters:

$$c(j; \theta_1, \theta_2) = \frac{f(\frac{j}{j^{max}}, \theta_1; \theta_2)}{\sum_{j=0}^{j^{max}-1} f(\frac{j}{j^{max}}, \theta_1; \theta_2)}, \tag{2}$$

$$f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}, \tag{3}$$

and $\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx$. One attractive specific case of the MIDAS Beta polynomial involves only one parameter, namely setting $\theta_1 = 1$ and estimating the single parameter θ_2 with the restriction that it be larger than one, yields single-parameter downward sloping weights more flexible than exponential or geometric decay patterns. In Figures 1 and 2 we plot the Beta function, not normalized as in equation (3), for a grid of parameters. In the former the range is 2 through 10 with increments of one, in the latter 10 through 90 with increments of 10. In both figures we observe a wide range of downward sloping curves. When we examine equation (2) - which is ultimately used in MIDAS regressions - we observe that there is one more implicit parameter which affects the actual weight being attributed

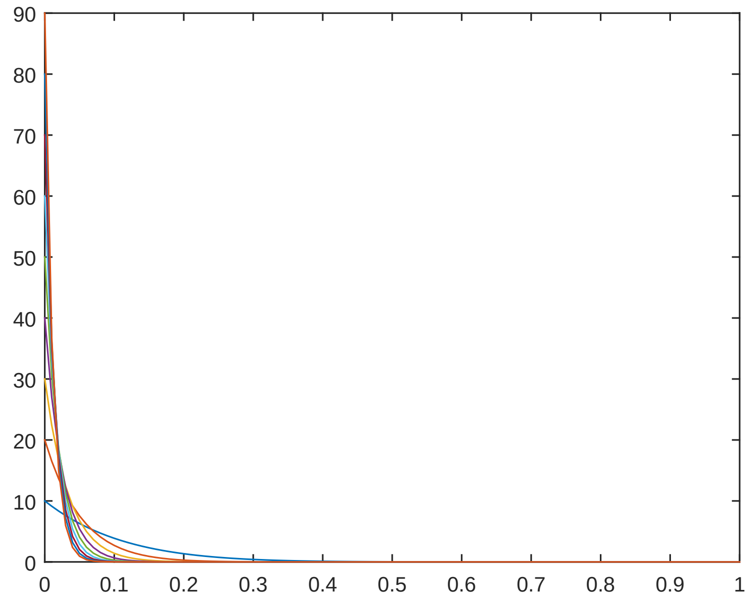

Figure 2: Plot of Beta polynomial appearing in equation (3) with $\theta_1 = 1$ and $\theta_2 = 10, \ldots,$ 90

to the high frequency data. Indeed, the lag length j^{max} also plays a key role. The Beta polynomial is defined on the $[0, 1]$ interval. Fixing j^{max} amounts to setting the equally spaced discrete tick size of the weights being computed. Combined with the normalization in equation (2) implies that one has very different lag coefficients. Take for example $\theta_1 = 1$, $\theta_2 = 5$. Then with $j^{max} = 10$ (perhaps a plausible value for quarterly/monthly data mixtures) we have $c(1; \theta_1, \theta_2) = 0.3947$, whereas with $j^{max} = 100$ (perhaps a plausible value for quarterly/daily data mixtures) $c(1; \theta_1, \theta_2) = 0.0488$, ceteris paribus. How do we decide on the maximal lag in the MIDAS polynomial? It might be tempted to use say an information criterion as is typically done in ARMA models. However, the number of lags in the high frequency polynomial is not affecting the number of parameters. Hence, the usual penalty functions such as those in the Akaike, Schwarz or Hannan-Quinn criteria will not apply. The only penalty of picking j^{max} too large is that we require more (high frequency) data at the beginning of the sample as the weights typically vanish to zero with j^{max} too large. Picking j^{max} too small is more problematic. This issue has been discussed extensively in the standard literature on distributed lag models, see e.g. Judge, Hill, Griffiths, Lütkepohl, and Lee (1988, Chapters 8 & 9). Needless to say that the choice is often fairly natural and depends on the sampling frequencies being involved. As alluded to earlier, $j^{max} = 10$ appears suitable for monthly high frequency data, but not for daily.

We noted earlier that it is not an absolute necessity to restrict θ to be one-dimensional. The parameterization of the lagged coefficients of $c(j;\theta)$ in a parsimonious way is one of the key MIDAS features.

The procedure discussed in this paper is not limited to the DL-MIDAS model appearing in equation (1). For example, adding lagged dependent - i.e. low frequency - variables yields the class of ADL-MIDAS regressions. To streamline the notation we will assume for convenience and no loss of generality in the remainder of the paper that $h = 1$. Henceforth we will also drop the subscripts h as a consequence. Assuming an autoregressive augmentation of order p, the model can be written as:

$$y_{t+1}^L = a + \sum_{j=1}^{p} \rho_j y_{t-j+1}^L + bC(L^{1/m};\theta)x_t^H + \varepsilon_{t+1}^L. \tag{4}$$

Once again, we can proceed as before, i.e. fix the parameters of the MIDAS polynomial and the resulting equation is a dynamic regression. Profiling can also be used for other MIDAS-type regressions, including (1) smooth transition MIDAS regression models proposed by Galvão (2013), (2) the Markov-Switching MIDAS models of Guérin and Marcellino (2013) and (3) the MIDAS quantile regression models studied by Ghysels, Plazzi, and Valkanov (2016), to name a few.

Foroni, Marcellino, and Schumacher (2015) study the performance of a variant of MIDAS which does not resort to functional distributed lag polynomials. Suppose m is small, like equal to three - as in quarterly/monthly data mixtures. Instead of estimating $bC(L^{1/m};\theta)$ in equation (4) let us estimate the individual lags separately - hence the term unrestricted -

yielding the following MIDAS regression:

$$y_{t+1}^L = a + \sum_{j=1}^{p} \rho_j y_{t-j+1}^L + \sum_{j=0}^{j^{max}-1} c_j x_{t-\frac{j}{m}}^H + \varepsilon_{t+1}^L, \tag{5}$$

which implies that in addition to the parameters a and ρ_j we estimate $j^{max} - 2$ additional parameters. With $m = 3$ and $(j^{max} - 1)$ small, like say up to 4 (annual lags) and large enough to make the error term ε_{t+1}^L uncorrelated, then, all the parameters in the U-MIDAS model can be estimated by simple OLS.

2.2 Estimators and their Asymptotic Properties

We are interested in comparing regressions (4) and (5). For simplicity we will assume that the errors of both regressions are i.i.d. Gaussian mean zero and variance σ^2. Let (Δ, Θ) be the parameter space associated with the MIDAS regression (4), more specifically we let $\delta = (a, \rho_1, \ldots, \rho_p, b, \sigma^2)$, and for some interior point $(\delta_0, \theta_0) \in (\Delta, \Theta)$ with $\delta_0 \neq 0$, we have:

$$E\left[y_{t+1}^L | y_{t-j}^L, x_{t-j/m}^H; j = 0, 1, \ldots\right] = a_0 + \sum_{j=1}^{p} \rho_{0j} y_{t-j+1}^L + b_0 C(L^{1/m}; \theta_0) x_t^H, \tag{6}$$

meaning that the MIDAS regression model is correctly specified with parameters $\delta_0 = (a_0, \rho_{01}, \ldots, \rho_p, b_0, \sigma_0^2)$, and θ_0.[4] For any given θ we obtain the linear regression model:

$$Y \equiv \begin{bmatrix} y_{p+1}^L \\ \vdots \\ \\ \vdots \\ y_T^L \end{bmatrix}, \qquad X(\theta) \equiv \begin{bmatrix} 1 & y_p^L & \ldots & y_1^L & x_p(\theta) \\ \vdots & & & & \vdots \\ & & & & \\ \vdots & & & & \vdots \\ 1 & y_{T-1}^L & \ldots & y_{T-p}^L & x_{T-1}(\theta) \end{bmatrix},$$

[4]The restriction $\delta_0 \neq 0$ is somewhat too strong. What is important is that $b_0 \neq 0$, because the parameters θ will not be identified. The latter unnecessarily - at least in the context of the present paper - complicates the asymptotics.

where $x_t(\theta) = C(L^{1/m};\theta)x_t^H$ and we assume that enough high frequency data x_t^H are available at the start of the sample. We can therefore consider a standard linear regression, namely:

$$Y = X(\theta)\beta + \varepsilon, \tag{7}$$

and add the assumption that the errors are homoskedastic and all ε_t are i.i.d. $N(0,\sigma^2)$. The log-likelihood which obviously depends on the entire parameter space (δ,θ), namely:

$$\mathcal{L}([y_t^L, x_t^H];\delta,\theta) = \sum_{t=h+1}^{T} \log f([y_t^L, x_t^H];\delta,\theta)),$$

where $\delta = (\beta,\sigma^2)$. As is typical in profile likelihood let us consider first fix $\theta = \bar{\theta}$ which yields the easier to evaluate:

$$\begin{aligned} \mathcal{L}_{T,\theta=\bar{\theta}}([y_t^L, x_t^H];\delta_{\bar{\theta}}) &= \sum_{t=h+1}^{T} \log f([y_t^L, x_t^H];\delta_{\bar{\theta}}) \\ &\approx -.5(Y - X(\bar{\theta})\beta_{\bar{\theta}})'(Y - X(\bar{\theta})\beta_{\bar{\theta}}), \end{aligned} \tag{8}$$

which yields the OLS estimator as maximand:

$$\hat{\beta}_{\bar{\theta}} = (X(\bar{\theta})'X(\bar{\theta}))^{-1}X(\bar{\theta})'Y.$$

Note that we have *profiled* out the MIDAS polynomial parameter(s) θ, to obtain a closed form solution estimator. Substituting $\hat{\beta}_{\bar{\theta}}$ into the likelihood function we obtain:

$$\begin{aligned} \mathcal{L}([y_t^L, x_t^H];\delta,\theta) &\approx -.5(Y - X(\theta)(X(\theta)'X(\theta))^{-1}X(\theta)'Y)'(Y - X(\theta)(X(\theta)'X(\theta))^{-1}X(\theta)'Y) \\ &\approx -.5Y'M(\theta)Y, \end{aligned}$$

where $M(\theta) = (I - X(\theta)(X(\theta)'X(\theta))^{-1}X(\theta)')$ the standard idempotent matrix associated with linear regression models. A bit of logical deduction shows that θ which minimizes

$Y'M(\theta)Y$ yields the estimators which maximize the log-likelihood $\mathcal{L}_T([y_t^L, x_t^H]; \delta, \theta)$. The optimization with respect to θ can be further simplified, namely:

$$\begin{aligned}
\max_\theta(-.5Y'M(\theta)Y) &= \min_\theta(Y'Y - Y'X(\theta)(X(\theta)'X(\theta))^{-1}X(\theta)'Y) \\
&= \max_\theta(Y'X(\theta)(X(\theta)'X(\theta))^{-1}X(\theta)'Y), \qquad (9)
\end{aligned}$$

where the above maximization is under the constraint that the regressors $X(\theta)$ are selected within the class of polynomials - Beta or other - with weights adding up to one (call this the class of feasible regressors). Hence, we have derived a remarkably simple MIDAS regression estimator, which amounts to:

$$\begin{aligned}
\hat{\theta} &= \max_\theta(Y'X(\theta)(X(\theta)'X(\theta))^{-1}X(\theta)'Y), \qquad (10) \\
\hat{\beta}(\hat{\theta}) &= (X(\hat{\theta})'X(\hat{\theta}))^{-1}X(\hat{\theta})Y.
\end{aligned}$$

The procedure is intuitively fairly straightforward. Given a set of regressors, we have the best linear regression estimates in closed form, and we select the regressors such as to have the best fit within the class of feasible regressors. If we write $(X(\theta)'X(\theta))^{-1}$ as $(\Lambda(\theta)^{-1/2})'(\Lambda(\theta)^{-1/2})$, then we can rewrite (10) as a Euclidean norm maximization, namely:

$$\hat{\theta} = \max_\theta \|\Lambda(\theta)^{-1/2}X(\theta)'Y\|, \qquad (11)$$

which amounts to finding - not surprisingly - the largest scaled covariance between Y and the parameter-driven regressors. Estimating only one parameter in the Beta polynomial has the advantage that it involves univariate optimization and we do not have any saddle or deep valley areas, from which numerical algorithm can hardly escape. Moreover, its gradient is analytically available so that large-scale search algorithms for a global optimum are applicable. Finally, in this case θ falls into $(1, \infty)$, so we may use the invariance property of MLE and one-to-one transform this parameter such that it falls in a compact region, say

$[0,1]$, which facilitates numerical search for a bounded problem.

Against the background of the profile MIDAS regression estimator, we now consider the U-MIDAS model (5) which can be written in matrix form as:

$$Y \equiv \begin{bmatrix} y^L_{p+1} \\ \vdots \\ \\ \vdots \\ y^L_T \end{bmatrix}, \qquad X^H \equiv \begin{bmatrix} 1 & y^L_p & \dots & y^L_1 & x^H_p & \dots & x^H_{p-(j^{max}-1)} \\ \vdots & & & & & & \vdots \\ \\ \vdots & & & & & & \vdots \\ 1 & y^L_{T-1} & \dots & y^L_{T-p} & x^H_{T-1} & \dots & x^H_{T-1-(j^{max}-1)} \end{bmatrix},$$

which yields:

$$Y = X^H \beta_U + \tilde{\varepsilon}, \tag{12}$$

where $\beta_U = (a, \rho_1, \dots, \rho_p, c_0, c_1, \dots, c_{j^{max}-1})$. Since the slope coefficient in the MIDAS regression has the property that $b = \sum_{j=0}^{(j^{max}-1)} c_j$, it is easy to see that the profile MIDAS is a constrained estimation as:

$$\begin{aligned} 0 &= c_0 - [\sum_{j=0}^{(j^{max}-1)} c_j] \times c(0, \theta), \\ \vdots \quad & \quad \vdots \qquad \vdots \\ 0 &= c_{j^{max}-1} - [\sum_{j=0}^{(j^{max}-1)} c_j] \times c(j^{max}-1, \theta), \end{aligned}$$

which amounts to $j^{max}-1$ restrictions (one less that the number of high frequency coefficients in the U-MIDAS regression). Hence, we can write these restrictions as:

$$R(\theta)\beta_U = 0 \tag{13}$$

where $R(\theta)$ is of dimension $(p+2) \times (p+2+(j^{max}-1))$. Using standard linear regression arguments, see e.g. Judge, Hill, Griffiths, Lütkepohl, and Lee (1988), we can then also write that:

$$\hat{\beta}(\theta) = \hat{\beta}_U - (X^{H\top} X^H)^{-1} R(\theta)^\top [R(\theta)(X^{H\top} X^H)^{-1} R(\theta)^\top]^{-1} (R(\theta)\hat{\beta}_U), \qquad (14)$$

which also implies the following relationship between the asymptotic variances of the respective estimators:

$$\begin{aligned} V(\hat{\beta}(\theta)) &= \sigma^2 (X^{H\top} X^H)^{-1} \qquad (15) \\ &\quad -\sigma^2 (X^{H\top} X^H)^{-1} R(\theta)^\top [R(\theta)(X^{H\top} X^H)^{-1} R(\theta)^\top]^{-1} R(\theta)(X^{H\top} X^H)^{-1}, \end{aligned}$$

where the matrix $C = \sigma^2 (X^{H\top} X^H)^{-1} R(\theta)^\top [R(\theta)(X^{H\top} X^H)^{-1} R(\theta)^\top]^{-1} R(\theta)(X^{H\top} X^H)^{-1}$ is positive semi-definite if the constraints in (13) are valid (see Fomby, Hill, and Johnson (1984)). This means that properly specified MIDAS regressions will be asymptotically more efficient than the OLS estimation of the U-MIDAS regression which yields asymptotic variance of the parameters equal to $\sigma^2 (X^{H\top} X^H)^{-1}$.

Besides asymptotic efficiency of the estimators, we also need to discuss numerical efficiency issues - which will be discussed next.

3 Examples

It will be useful to consider examples to appraise the performance of the profile estimators introduced in the previous section. The first is inspired by the work of Foroni, Marcellino, and Schumacher (2015) and covers cases where m is small and so called unrestricted MIDAS, dubbed U-MIDAS by the authors, is attractive. What is of interest here is assessing how the OLS estimation of MIDAS regression parameters compares to the de facto nonlinear estimation via profiling. The second case is when m is large and we use the example of volatility forecasting using MIDAS-type regressions, as suggested by Ghysels, Santa-Clara, and Valkanov (2006), as a motivating example.

3.1 Profiling when U-MIDAS is attractive

Suppose m is small, like equal to three - as in quarterly/monthly data mixtures and by implication j^{max} is small. We consider a simplified MIDAS regression:

$$y_{t+1}^L = \sum_{j=0}^{j^{max}-1} c_j x_{t-\frac{j}{m}}^H + \varepsilon_{t+1}^L, t = 1, \ldots, T, \tag{16}$$

where the intercept, low frequency autoregressive and exogenous terms are not essential to our analysis and thus omitted. U-MIDAS assumes that $c_j, j = 0, \ldots, j^{max}-1$ are unrestricted and estimated by OLS, whereas the Beta-polynomial MIDAS model tightly parameterizes weights as $c_j \propto (1 - \frac{j}{j^{max}})^{\theta-1}$. The profiling approach estimates the proportionality constant (i.e., the slope term) by OLS conditional on θ, and subsequently optimizes θ.

We compare the computational complexity of the two approaches in terms of floating point operations (FLOPs).[5] Since the sample size is T and the number of regressors is j^{max}, the OLS estimator of U-MIDAS has the complexity $O((j^{max})^2T) + O((j^{max})^3)$. Specifically, the required FLOPs amount to $2(j^{max})^2T + 2j^{max}T + \frac{2}{3}(j^{max})^3 + 2(j^{max})^2$. For the profile MIDAS regression, computing the normalized Beta polynomial takes about $5j^{max}$ FLOPs, and generating the weighted average of the high-frequency variables requires $2j^{max}T$ FLOPs. Estimating the proportionality constant by univariate OLS roughly costs $4T$ FLOPs. Lastly, it takes another $4T$ FLOPs to compute the squared residuals, which will be minimized to obtain an estimator of θ. We may resort to either grid search or an iterative method such as golden section search. If the grid size (or the number of iterations) equals S, then the algorithm has the complexity $O(j^{max}ST)$, or more specifically, $S(5j^{max} + 2j^{max}T + 8T)$ FLOPs. Generally speaking, univariate optimization is computationally simple. Our experience is that $S < 10$ can yield a good estimator for θ.

Since the complexity of the profile MIDAS estimator increases linearly with j^{max}, while that of the U-MIDAS estimator rises with $(j^{max})^3$, we expect that profiling is computationally more efficient than U-MIDAS when j^{max} is moderately large.

[5] By convention, a floating-point addition, subtraction, multiplication, or division counts towards a FLOP.

In our first Monte Carlo experiment, the simulated monthly data follow an AR(1) process with the autocorrelation 0.8, and the quarterly data are generated by the Beta-polynomial MIDAS.[6] We consider $T = \{50, 300, 2000\}, j^{max} = \{6, 9, 12\}$ and $\theta = \{2, 10\}$, whose combinations lead to 18 scenarios. The MIDAS parameter is optimized via golden section search with the number of iterations $S = 7$.[7] The Monte Carlo experiment involves 1000 replications. In each scenario and replication, we simulate monthly/quarterly data, run U-MIDAS and profile MIDAS regressions, compute the squared bias of the estimated weights, calculate the in-sample R-squared (R^2), make out-of-sample prediction for six months and record the mean squared errors (MSE). Specifically, those indicators are defined as

$$Bias^2 = \sum_{j=0}^{j^{max}-1} (\hat{c}_j - c_j)^2,$$

$$R^2 = 1 - \sum_{t=1}^{T} (\hat{y}_t{}^L - y_t^L)^2 / \sum_{t=1}^{T} y_t^2,$$

$$MSE = \frac{1}{r} \sum_{t=T+1}^{T+r} (\hat{y}_t{}^L - y_t^L)^2,$$

where $\hat{c}_j$ is the estimated weight and $\hat{y}_t{}^L$ is the fitted/predicted value.

The simulation results are presented in Table 1. The profile estimator has smaller $Bias^2, R^2$ and MSE compared to U-MIDAS in all scenarios, regardless of the sample size and the number of lags. On average, the squared bias of U-MIDAS is 422% larger and its forecast MSE is 7.1% higher. Though the profile regression represents the correctly specified model, its R^2 is smaller than that of U-MIDAS. For example, in the scenario $T = 50, j^{max} = 6, \theta = 2$, the R^2 are 0.673 and 0.701 respectively. Despite its modesty in goodness-of-fit, the profile estimator has higher predictive power: its forecast MSE equals 1.063, which is significantly

[6] The autocorrelation parameter has little impact on the simulation results. We also tried autocorrelations of 0.1 and 0.9, and the results are similar.

[7] For the convenience of FLOPs counting, the number of functional evaluations is limited to 7. We could run more iterations until convergence, and the profile estimator would be numerically identical to the estimator without profiling. However, our simulation results suggest that the coarsely tuned parameter has already performed well in forecast. We did not observe improvement with more iterations.

Table 1: Comparison between profile estimator and U-MIDAS

Scenarios			SquaredBias		R-Squared		ForecastMSE		FLOPs (thousand)	
T	j^{max}	θ	Profile	U-MIDAS	Profile	U-MIDAS	Profile	U-MIDAS	Profile	U-MIDAS
50	6	2	0.094	0.414	0.673	0.701	1.063	1.160	7.3	5.3
			(0.002)	(0.005)	(0.003)	(0.003)	(0.020)	(0.021)		
50	6	10	0.142	0.415	0.726	0.750	1.035	1.128	7.3	5.3
			(0.003)	(0.005)	(0.002)	(0.002)	(0.020)	(0.022)		
50	9	2	0.073	0.548	0.644	0.696	1.042	1.219	9.5	10.8
			(0.001)	(0.006)	(0.003)	(0.003)	(0.020)	(0.023)		
50	9	10	0.143	0.550	0.710	0.753	1.068	1.259	9.5	10.8
			(0.003)	(0.005)	(0.003)	(0.002)	(0.021)	(0.025)		
50	12	2	0.064	0.683	0.600	0.684	1.017	1.318	11.7	18.5
			(0.001)	(0.006)	(0.004)	(0.003)	(0.018)	(0.025)		
50	12	10	0.119	0.673	0.700	0.763	1.046	1.339	11.7	18.5
			(0.003)	(0.006)	(0.003)	(0.002)	(0.019)	(0.025)		
300	6	2	0.038	0.159	0.675	0.680	0.999	1.012	42.3	30.5
			(0.001)	(0.002)	(0.001)	(0.001)	(0.018)	(0.018)		
300	6	10	0.066	0.160	0.724	0.728	1.016	1.030	42.3	30.5
			(0.001)	(0.002)	(0.001)	(0.001)	(0.018)	(0.018)		
300	9	2	0.029	0.205	0.643	0.652	0.994	1.025	55.0	61.6
			(0.001)	(0.002)	(0.001)	(0.001)	(0.018)	(0.019)		
300	9	10	0.057	0.205	0.713	0.719	0.936	0.953	55.0	61.6
			(0.001)	(0.002)	(0.001)	(0.001)	(0.018)	(0.018)		
300	12	2	0.025	0.246	0.614	0.627	1.005	1.040	67.7	103.8
			(0.000)	(0.002)	(0.001)	(0.001)	(0.018)	(0.019)		
300	12	10	0.053	0.245	0.701	0.712	0.993	1.027	67.7	103.8
			(0.001)	(0.002)	(0.001)	(0.001)	(0.018)	(0.019)		
2000	6	2	0.014	0.062	0.678	0.678	0.976	0.978	280.3	202.2
			(0.000)	(0.001)	(0.000)	(0.000)	(0.019)	(0.019)		
2000	6	10	0.037	0.062	0.726	0.727	1.001	1.002	280.3	202.2
			(0.000)	(0.001)	(0.000)	(0.000)	(0.019)	(0.019)		
2000	9	2	0.011	0.078	0.645	0.646	1.003	1.009	364.4	406.7
			(0.000)	(0.001)	(0.001)	(0.001)	(0.018)	(0.018)		
2000	9	10	0.042	0.078	0.713	0.714	1.031	1.033	364.4	406.7
			(0.000)	(0.001)	(0.000)	(0.000)	(0.019)	(0.019)		
2000	12	2	0.009	0.094	0.615	0.617	0.999	1.004	448.5	683.5
			(0.000)	(0.001)	(0.001)	(0.001)	(0.018)	(0.018)		
2000	12	10	0.041	0.093	0.701	0.703	0.986	0.990	448.5	683.5
			(0.000)	(0.001)	(0.000)	(0.000)	(0.017)	(0.018)		

The data are generated by MIDAS Beta polynomials according to the scenarios specified by the first three columns. The Monte Carlo experiment involves 1000 replications, and the average squared bias of estimated weights, R-squared and forecast MSE are reported, with standard errors in parentheses. The last two columns report the runtime counting of FLOPs, which is close to, but slightly larger than, the theoretic numbers due to programming overhead costs.

lower than U-MIDAS MSE 1.160. The results suggest that U-MIDAS suffers from overfitting.The unrestricted parameters can be trained with a high in-sample R^2, but cannot outperform a parsimonious model in out-of-sample prediction.[8]

If we fix j^{max} and increase T, both the OLS and profile estimators are consistent. The fourth and fifth columns of Table 1 witness monotone decrease of the squared bias with the sample size. For instance, the squared bias of the profile (U-MIDAS) estimator drops from 0.094 (0.414) to 0.038 (0.159), and further to 0.014 (0.062), as $T = 50, 300, 2000$, respectively. Simulation results confirm that the profile estimator is more efficient than the OLS estimator. There are notable differences in their finite-sample predictive behaviors. In the smaller-sample scenarios, the discrepancy between the two approaches looms larger, and the profile estimator demonstrates its advantage in prediction. For example, when $T = 50$, the MSE of U-MIDAS is 18.4% higher. The advantage stretches to the scenarios of $T = 300$. Not until $T = 2000$ does U-MIDAS catch up in its predictive power. With a large sample size, the unrestricted parameters in U-MIDAS can mimic the shape the true Beta polynomial, as evidenced by smaller bias in the estimated weights. Therefore, the predictive power increases. Table 1 shows that MSEs are similar between the two methods when $T = 2000$.

The number of lags is also a key factor to predictive performance. Simulation results indicate that U-MIDAS is overshadowed by the profile estimator in the scenarios of large j^{max} and small T. When $T = 50, j^{max} = 12, \theta = 2$, the MSE of the profile estimator is as low as 1.017, while U-MIDAS has 1.318, or a 29.5% increase.

Table 1 also reports the runtime FLOPs. The number of lags critically determines the complexity of the matrix algorithm. When $j^{max} = 6$, U-MIDAS requires smaller FLOPs and thus is attractive. However, as j^{max} rises the FLOPs increase so quickly that the computational advantage of U-MIDAS is lost at $j^{max} = 9$. As for the scenarios of $j^{max} = 12$,

[8]Overfitting of U-MIDAS is a theoretical necessity, because the unrestricted weights can always replicate the estimated weights by the Beta polynomial. However, the unrestricted weights overreact to in-sample data so as to minimize the squared residuals.

U-MIDAS falls behind the profiling approach, which only requires two thirds of FLOPs compared to U-MIDAS.

One may suspect that the comparison shown in Table 1 is unfair, as the profile MIDAS model represents the true data generating process. In our second Monte Carlo experiment, we generate mixed frequency data by an exponential Almon polynomial but fit the model by the Beta polynomial. The weights of the former are given by $c_j \propto e^{\theta j^2}$, which is essentially a Gaussian density with a thinner tail than the Beta polynomial. In that case, the MIDAS model is misspecified and the profile estimator is biased and inconsistent. However, If we fix j^{max} and increase T, the OLS estimator of U-MIDAS still satisfies the Gauss-Markov assumptions, and thus remains unbiased and consistent. Despite such handicap in statistical properties, the misspecified profile estimator is blessed with smaller variance due to tightly regularized weights. As MSE inflates with both bias and variance, it is of interest to evaluate the predictive performance of a misspecified model.

The monthly data are simulated in the same manner, while the quarterly data are generated by the exponential Almon polynomial with $\theta = \{-0.02, -0.002\}$. Then we estimate and forecast using U-MIDAS and the profile MIDAS under the Beta polynomial. The FLOPs are exactly the same as those in the previous experiment, and thus omitted. In addition to $Bias^2, R^2$ and MSE, Table 2 also reports the bias in level, which is defined as $Bias = \sum_{j=0}^{j^{max}-1} (\hat{c}_j - c_j)$. As the reported statistics are averaged among the 1000 Monte Carlo repetitions, the sample analogue shown in column 4 and 5 for an unbiased (consistent) estimator should be close to (asymptotically close to) zero, and statistically insignificant.

Results in Table 2 confirm that U-MIDAS has little bias in level even if the sample size is small. As T increases, bias rapidly vanishes to zero. In contrast, the misspecified profile estimator has substantially larger bias, which slowly decays, but not to zero. It is enlightening to compare the squared bias of the estimated weights. Surprisingly, $Bias^2$ of the profile estimator is significantly smaller, especially when the sample size is small or moderate. The mean squared bias of U-MIDAS is 295% higher. This is because more

Table 2: Comparison between misspecified profile estimator and U-MIDAS

Scenarios			Bias		Squared Bias		R-Squared		MSE	
T	j^{max}	θ	Profile	U-MIDAS	Profile	U-MIDAS	Profile	U-MIDAS	Profile	U-MIDAS
50	6	-0.020	-0.022	-0.003	0.113	0.414	0.645	0.679	1.059	1.160
			(0.003)	(0.004)	(0.001)	(0.005)	(0.003)	(0.003)	(0.020)	(0.021)
50	6	-0.002	-0.030	0.001	0.133	0.415	0.634	0.677	1.036	1.128
			(0.003)	(0.004)	(0.001)	(0.005)	(0.003)	(0.003)	(0.020)	(0.022)
50	9	-0.020	-0.010	0.002	0.083	0.548	0.628	0.683	1.045	1.219
			(0.004)	(0.004)	(0.001)	(0.006)	(0.003)	(0.003)	(0.020)	(0.023)
50	9	-0.002	-0.029	-0.003	0.092	0.550	0.595	0.663	1.063	1.259
			(0.004)	(0.004)	(0.001)	(0.005)	(0.004)	(0.003)	(0.020)	(0.025)
50	12	-0.020	-0.010	-0.015	0.068	0.683	0.604	0.686	1.018	1.318
			(0.004)	(0.005)	(0.001)	(0.006)	(0.004)	(0.003)	(0.018)	(0.025)
50	12	-0.002	-0.028	-0.010	0.074	0.673	0.562	0.658	1.051	1.339
			(0.004)	(0.005)	(0.001)	(0.006)	(0.004)	(0.003)	(0.019)	(0.025)
300	6	-0.020	-0.018	-0.002	0.092	0.159	0.652	0.659	1.004	1.012
			(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.018)	(0.018)
300	6	-0.002	-0.024	0.000	0.115	0.160	0.643	0.656	1.034	1.030
			(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.018)	(0.018)
300	9	-0.020	-0.012	-0.000	0.052	0.205	0.628	0.637	0.996	1.025
			(0.001)	(0.002)	(0.000)	(0.002)	(0.001)	(0.001)	(0.018)	(0.019)
300	9	-0.002	-0.021	-0.002	0.072	0.205	0.609	0.622	0.946	0.953
			(0.001)	(0.002)	(0.000)	(0.002)	(0.002)	(0.002)	(0.018)	(0.018)
300	12	-0.020	-0.005	-0.001	0.033	0.246	0.616	0.630	1.005	1.040
			(0.002)	(0.002)	(0.000)	(0.002)	(0.001)	(0.001)	(0.018)	(0.019)
300	12	-0.002	-0.016	0.000	0.054	0.245	0.575	0.591	0.995	1.027
			(0.002)	(0.002)	(0.000)	(0.002)	(0.002)	(0.002)	(0.018)	(0.019)
2000	6	-0.020	-0.013	-0.000	0.084	0.062	0.655	0.658	0.983	0.978
			(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.019)	(0.019)
2000	6	-0.002	-0.023	0.001	0.114	0.062	0.647	0.655	1.025	1.002
			(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.019)	(0.019)
2000	9	-0.020	-0.013	-0.000	0.043	0.078	0.629	0.631	1.007	1.009
			(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.019)	(0.018)
2000	9	-0.002	-0.019	-0.000	0.069	0.078	0.611	0.616	1.031	1.033
			(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.019)	(0.019)
2000	12	-0.020	-0.004	0.000	0.021	0.094	0.618	0.620	1.001	1.004
			(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.018)	(0.018)
2000	12	-0.002	-0.014	-0.000	0.048	0.093	0.577	0.581	0.985	0.990
			(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.017)	(0.018)

The data are generated by MIDAS exponential Almon polynomials according to the scenarios specified by the first three columns, but the MIDAS model is estimated under the Beta polynomial. The Monte Carlo experiment involves 1000 replications, and the average bias and squared bias of estimated weights, R-squared and forecast MSE are reported, with standard errors in parentheses.

unrestricted parameters induce bigger OLS estimator variance, which in turn inflates the squared bias for U-MIDAS. Such variability in estimation further translates to poorer out-of-sample forecast performance. On average, U-MIDAS MSE is 6.6% higher compared to the misspecified profile estimator. For example, when $T = 50, j^{max} = 12, \theta = -0.02$, the forecast MSE of U-MIDAS amounts to 1.318, while the profile estimator only has 1.018.

Our empirical findings do not conflict with the asymptotic theory. The elegance of unbiasedness and consistency in U-MIDAS exhibit in our large T and small j^{max} scenarios, in which U-MIDAS does slightly outperform the misspecified profile estimator. When $T = 2000, j^{max} = 6, \theta = -0.02$, the squared bias of the profile estimator amounts to 0.084, while that of U-MIDAS is as low as 0.062. The forecast MSE of the profile estimator is 0.983, while that of U-MIDAS is 0.978. In conclusion, a misspecified profile estimator could be slightly inferior only if the sample size is really large and the number of lags is small. In a typically monthly/quarterly mixed frequency application, it is not feasible to collect thousands of quarterly observations and therefore we cannot set a high expectation on the unrestricted OLS estimator. However, the simulation results assure that the profile estimator works reliably and exceeding well in most scenarios.

3.2 Profiling when U-MIDAS is unattractive

We now turn to a situation where m is not small as for example in the case of predicting so called realized volatility (RV) over some future horizon h using daily realized volatility. We fit models using the S&P 500 realized volatility series from the Oxford-Man Realized Library data source.[9] We consider two types of models. The first is a DL-MIDAS, namely:

$$V_t^{t+h} = a + bC(L,\theta)V_{t-j-1}^{t-j} + \varepsilon_t^h, t = 1, \ldots, T, \tag{17}$$

[9]See, http://realized.oxford-man.ox.ac.uk/ for further details.

where V_t^{t+h} is the logarithmic RV from t to $t+h$ using intra-daily squared returns.[10] The second model is the so called HAR model of Corsi (2009), which is a popular application of MIDAS with step functions (cfr. Ghysels, Sinko, and Valkanov (2006) and Forsberg and Ghysels (2006)) involving daily, weekly and monthly realized volatility. In particular:

$$V_t^{t+h} = a + c_d V_{t-1}^t + c_w V_{t-5}^t + c_m V_{t-22}^t + \varepsilon_t^h, t = 1, \ldots, T, \tag{18}$$

i.e. one uses the last day (with slope c_d), last week (hence c_w) and last month (c_m) to predict future RV.

To facilitate comparison, the number of lags in the MIDAS Beta polynomial is fixed at 22, so that both models utilize information of the previous month to forecast future RV. Note that this puts the DL-MIDAS model at a disadvantage, since it can easily handle more lags, but we prefer to keep the lag length the same as in the Corsi model to avoid potential differences in our results due to different information sets. The two models differ in the weights assigned to the past realized volatility.

We provide the details of the excercise in Appendix A and only provide a brief description here. Namely, we consider 16 scenarios by varying $h = \{1, 5, 10, 22\}$ (RV of the next day/week/bi-week/month), as well as the estimation sample size $T = \{500, 1500, 2500, 3500\}$. After parameter estimation, we evaluate the forecast MSE for the next 200 periods.

The DL-MIDAS model can be estimated with or without profiling, and we report the results for both. Ideally, they should produce exactly the same estimators, if there weren't any frictions in the numerical optimization. We intentionally create some friction. For the profile estimator, we do not fully optimize θ, but resort to a coarse grid search with only five logarithmically spaced values, namely $1.5, 2.7, 4.7, 8.4, 15$. If the profile estimator still thrives in such hardship, it will be valuable in practice: simple to implement, low in computational complexity, and suitable for parallel computing on big data.

[10] Note that one could also consider MIDAS regressions using directly the intra-daily squared returns - but as Ghysels, Santa-Clara, and Valkanov (2006) note there are no forecasting gains to do so.

Table 3: DL-MIDAS regression with and without profiling, and Corsi regression

Scenarios		Estimator θ		R-Squared			Forecast MSE			FLOPs		
h	T	PF	NP	PF	NP	CO	PF	NP	CO	PF	NP	CO
1	500	15.0	11.7	0.510	0.511	0.509	0.341	0.339	0.344	1.6E5	3.8E6	3.0E4
1	1500	8.4	11.0	0.704	0.706	0.704	0.358	0.364	0.363	4.7E5	9.4E6	8.9E4
1	2500	8.4	9.2	0.738	0.738	0.739	0.383	0.383	0.392	7.9E5	2.1E7	1.5E5
1	3500	8.4	9.1	0.695	0.695	0.692	0.357	0.355	0.374	1.1E6	2.6E7	2.1E5
5	500	8.4	9.4	0.572	0.574	0.569	0.215	0.213	0.222			
5	1500	8.4	8.8	0.780	0.780	0.778	0.168	0.169	0.170			
5	2500	8.4	6.8	0.804	0.805	0.806	0.262	0.261	0.281			
5	3500	8.4	6.8	0.762	0.763	0.762	0.260	0.261	0.285			
10	500	8.4	6.8	0.541	0.543	0.537	0.234	0.232	0.243			
10	1500	8.4	7.5	0.763	0.763	0.760	0.147	0.146	0.147			
10	2500	4.7	6.2	0.787	0.789	0.789	0.297	0.289	0.306			
10	3500	4.7	5.8	0.746	0.747	0.746	0.300	0.294	0.327			
22	500	4.7	3.9	0.439	0.439	0.438	0.312	0.313	0.321			
22	1500	8.4	6.4	0.690	0.691	0.691	0.131	0.127	0.130			
22	2500	4.7	5.3	0.720	0.721	0.725	0.433	0.430	0.452			
22	3500	4.7	4.9	0.684	0.684	0.686	0.429	0.428	0.439			

Corsi and DL-MIDAS models with and without profiling are fitted by S&P 500 realized volatility series, with the sample size T. Then future realized volatilities of horizon h are predicted. The estimated MIDAS coefficient, R^2 and forecast MSE are reported. PF = profile estimation; NP = non-profile estimation; CO = Corsi regression. FLOPs are invariant to h, so entries corresponding to h other than one are omitted to avoid duplication.

Table 3 presents R^2, MSE and FLOPs of the DL-MIDAS model estimated with and without profiling, as well as the Corsi model estimated by OLS. In short, we have two major findings: 1) the profile estimator is as good as the non-profile estimator in terms of in-sample fit and out-of-sample forecast performance, but the profile estimator is substantially faster; and 2) the profile estimator outperforms Corsi estimator in prediction, though the Corsi estimator is the fastest.

First, we compare the DL-MIDAS regression with and without profiling. The estimated parameters differ substantially due to coarsely discretized values. In one scenario, the non-profile estimator of θ equals 11.7, while the best available choice for the profile estimator is 15. In another scenario, the non-profile approach yields 11.0, to which the closest match by the profile estimator is 8.4. Given such discrepancy in polynomial weights, the estimated intercept and slope parameters have to be adjusted to sub-optimality.

Surprisingly, there is no apparent goodness-of-fit loss, and the predicted values are similar. The mean absolute difference in R^2 is 0.0007, and the non-profile MSE is only 0.6% lower in prediction. We are using the real data with no knowledge on the true data generating process. Table 3 demonstrates that a crudely tuned profile estimator does not necessarily yield an inferior predictive model.

The profiling method is clearly faster. The non-profile estimator is obtained by nonlinear numerical optimization, while the profiling approach resorts to the analytic optimizer, namely OLS, conditional on the specified θ value. In this application, the nonlinear optimization routine only receives three choice variables, and already observe huge complexity increases by an order of magnitude compared to profiling. The FLOPs of the non-profile estimator are 22.3 times higher on average. If we adopted an ADL-MIDAS model that contains autoregressive and exogenous terms, there would be more variables to be numerically optimized, and the non-profile estimation would be increasingly expensive in computation. In contrast, profiling always resorts to analytic optimizers for all variables except for the scalar θ.

Second, we compare the profile estimator with the Corsi model OLS estimator. The former has slightly higher R^2, and the mean absolute difference is 0.002. However, the predictive power of MIDAS is higher than that of the Corsi model; the latter has 3.5% higher in forecast MSE. For example, to predict next-day RV with 3500 observations, the profile estimator has the MSE 0.357, while Corsi MSE is 0.374. Empirical results suggest that classifying the past RV by day, week and month offers good predictions, but the DL-MIDAS outperforms Corsi regression. Also, it appears that the sample size has an impact on the predictive performance. Estimated using 500 observations, the profile estimator has 2.7% lower in MSE compared to Corsi regression. When the sample size increases to 3500, the advantage increases to 5.9% lower in MSE.

There is no doubt that the Corsi model estimator is the fastest, because it only requires OLS estimation with three predictors. Nevertheless, the profile estimator is second in computational efficiency, as it only involves several evaluations of OLS with one predictor. Also, the profile estimator can be straightforwardly parallelized in computing, which may give it an advantage in big data applications.

4 Conclusions

MIDAS regressions are a useful tool to address the challenges associated with data of mixed sampling frequency which are commonly encountered in many fields of science. The profiling methods introduced in this paper resolve an unpleasant trade-off hitherto faced by practitioners. Either one uses U-MIDAS, which involves simple OLS but is limited to frequency mixtures with only small sampling differences - like the most prominent example of monthly and quarterly data. Or, one uses standard MIDAS regressions involving nonlinear estimation. The new profiling estimation methods presented in the current paper make the computations both simple and fast, regardless of the sampling schemes involved.

MIDAS regressions obviously apply to cases involving potentially a large set of mixed frequency regressors. This leads to variable selection issues. Marsilli (2014) suggests to combine the MIDAS regressions and the LASSO estimator put forward by Tibshirani (1996). Unfortunately, Marsilli (2014) uses standard MIDAS involving nonlinear estimation - which makes applications with large numbers of regressors challenging. Our profiling approach would make the implementation of MIDAS-LASSO computationally more feasible. Similarly, Siliverstovs (2015) proposes what he calls MIDASSO, a combination of U-MIDAS and LASSO. While this is computationally more attractive, it has the same limitations as U-MIDAS.

Finally, as far as mixed frequency data is concerned, the idea of profiling is not limited to MIDAS regressions. In fact, a hint of using profiling methods appeared in Engle, Ghysels, and Sohn (2013) - who introduced GARCH-MIDAS volatility models (see also Asgharian, Hou, and Javed (2013), Conrad and Loch (2015), among others for further discussion of GARCH-MIDAS models). By the same token, one could apply profiling to dynamic conditional correlation type models, such as DCC-MIDAS - see e.g. Asgharian, Christiansen, and Hou (2015), Baele, Bekaert, and Inghelbrecht (2010), Colacito, Engle, and Ghysels (2011), Connor and Suurlaht (2013), among others. Moreover, since profiling methods are very attractive in the context of semi-parametric estimation, it is also natural to apply them to semi-parametric MIDAS regressions, see Chen and Ghysels (2011).

References

Andreou, E., E. Ghysels, and A. Kourtellos, 2010, "Regression Models With Mixed Sampling Frequencies," *Journal of Econometrics*, 158, 246–261.

Andreou, E., E. Ghysels, and A. Kourtellos, 2011, "Forecasting with mixed-frequency data," in *Oxford Handbook of Economic Forecasting*, ed. by C. M. P., and D. F. Hendry, pp. 225–245.

Armesto, M. T., K. M. Engemann, and M. T. Owyang, 2010, "Forecasting with mixed frequencies," *Federal Reserve Bank of Saint Louis Review*, 92, 521–536.

Asgharian, H., C. Christiansen, and A. J. Hou, 2015, "Macro-finance determinants of the long-run stock–bond correlation: The DCC-MIDAS specification," *Journal of Financial Econometrics*, forthcoming.

Asgharian, H., A. J. Hou, and F. Javed, 2013, "The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH-MIDAS Approach," *Journal of Forecasting*, 32, 600–612.

Baele, L., G. Bekaert, and K. Inghelbrecht, 2010, "The determinants of stock and bond return comovements," *Review of Financial Studies*, 23, 2374–2428.

Barndorff-Nielsen, O., 1983, "On a formula for the distribution of the maximum likelihood estimator," *Biometrika*, 70, 343–365.

Barndorff-Nielsen, O., and D. R. Cox, 1994, *Inference and asymptotics*. Chapman and Hall.

Chen, X., and E. Ghysels, 2011, "News Good or Bad and Its Impact on Volatility Predictions over Multiple Horizons," *Review of Financial Studies*, 24–81, 46.

Clements, M. P., and A. B. Galvão, 2009, "Forecasting US output growth using Leading Indicators: An appraisal using MIDAS models," *Journal of Applied Econometrics*, 24, 1187–1206.

Colacito, R., R. F. Engle, and E. Ghysels, 2011, "A component model for dynamic correlations," *Journal of Econometrics*, 164, 45–59.

Connor, G., and A. Suurlaht, 2013, "Dynamic stock market covariances in the Eurozone," *Journal of International Money and Finance*, 37, 353–370.

Conrad, C., and K. Loch, 2015, "Anticipating Long-Term Stock Market Volatility," *Journal of Applied Econometrics*, 30, 1090–1114.

Corsi, F., 2009, "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 7, 174–196.

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

Fomby, T., R. C. Hill, and S. Johnson, 1984, *Advanced Economic Methods*. New York: Springer-Verlag.

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

Foroni, C., and M. G. Marcellino, 2013, "A survey of econometric methods for mixed-frequency data," Available at SSRN 2268912.

Forsberg, L., and E. Ghysels, 2006, "Why do absolute returns predict volatility so well?," *Journal of Financial Econometrics*, 6, 31–67.

Galvão, A. B., 2013, "Changes in predictive ability with mixed frequency data," *International Journal of Forecasting*, 29, 395–410.

Ghysels, E., 2013, "Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regression Models," Available at `http://www.mathworks.com/matlabcentral/fileexchange/45150-midas-matlab-toolbox`.

Ghysels, E., V. Kvedaras, and V. Zemlys, 2016, "Mixed Frequency Data Sampling Regression Models: The R Package midasr," *Journal of Statistical Software*, 72, 1–35.

Ghysels, E., A. Plazzi, and R. Valkanov, 2011, "Conditional skewness of stock market returns in developed and emerging markets and its economic fundamentals," Swiss Finance Institute Research Paper - 11-06.

Ghysels, E., A. Plazzi, and R. Valkanov, 2016, "Why invest in emerging markets? The role of conditional return asymmetry," *Journal of Finance*, (forthcoming).

Ghysels, E., P. Santa-Clara, and R. Valkanov, 2004, "The MIDAS touch: Mixed Data Sampling Regressions," Discussion paper UNC and UCLA.

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

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

Guérin, P., and M. Marcellino, 2013, "Markov-switching MIDAS models," *Journal of Business and Economic Statistics*, 31, 45–56.

Judge, G. G., R. C. Hill, W. Griffiths, H. Lütkepohl, and T.-C. Lee, 1988, *Introduction to the Theory and Practice of Econometrics*. John Wiley & Sons, New York.

Marsilli, C., 2014, "Variable Selection in Predictive MIDAS Models," Banque de France Working Paper.

Murphy, S. A., and A. van der Vaart, 2000, "On profile likelihood," *Journal of the American Statistical Association*, 95, 449–465.

Patefield, W., 1977, "On the maximized likelihood function," *Sankhyā: The Indian Journal of Statistics, Series B*, 39, 92–96.

Siliverstovs, B., 2015, "Short-term forecasting with mixed-frequency data: A MIDASSO approach," KOF Working Paper.

Tibshirani, R., 1996, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.

van der Vaart, A., 2002, "Lecture 9 - Maximum and Profile Likelihood," in *Lectures on Probability Theory and Statistics: Ecole D'Eté de Probabilités de Saint-Flour XXIX-1999*, ed. by E. Bolthausen, E. Perkins, and A. van der Vaart. Springer Science.

A Appendix: Data and estimation of MIDAS and Corsi model

Daily RVs obtained from intra-day 5-minute squared returns of S&P 500 are downloaded from Oxford-Man Realized Library. There are 4156 observations over a 16 year period, from January 2000 to August 2016. RV of horizon h equals the sum of h daily RVs.

We consider 16 scenarios by varying $h = \{1, 5, 10, 22\}$ and $T = \{500, 1500, 2500, 3500\}$. When $T = 500$, we have 7 non-overlapping estimation samples. The first one uses observations $t = 23, \ldots, 522$, and the second one uses $t = 523, \ldots, 1022$, and so on. Similarly, when $T = 1500$ we have 2 non-overlapping samples. The scenarios of $T = 2500, 3500$ have only one sample.

In each sample, we estimate the Corsi model and MIDAS regressions with and without profiling. If a scenario has more than one sample, we report the average R^2 (Column 5-7 of Table 3). However, we report $\hat{\theta}$ corresponding to the last sample instead of the average (Column 3 and 4 of Table 3), because $\hat{\theta} \in \{1.5, 2.7, 4.7, 8.4, 15\}$ for the profile estimator and Column 3 should reflect one of those five values.

After parameter estimation, we compute out-of-sample predictions for 200 periods and compute the forecast MSE. If a scenario has more than one sample, we report the average MSE (Column 8-10 of Table 3).