Improved boosted two parameter Breitung estimator: An application to the crude oil prices

[1]Talha Omer, [1]Kristofer Månsson, [1]Pär Sjölander, [2]B.M.Golam Kibria
[1]Department of Economics, Finance and Statistics, Jönköping University, P.O. Box 1026, SE-551 11 Jönköping, Sweden
[3]Department of Mathematics and Statistics, Florida International University, Miami, FL, USA
Email:[1]talha.omer@ju.se;[1]kristofer.mansson@ju.se;[1]par.sjolander@ju.se; [3]kibriag@fiu.edu

Abstract: Breitung and Roling (2015) presented a superiority of the nonparametric approach for estimation of mixed-frequency forecast. This approach remained outer performed than the usual parametric approach. In this paper, a boosted two parameter Breitung estimator by using the Lipovetsky and Conklin (2005) is proposed. Monte Carlo Simulation experiment suggests that boosted two parameters with an additional predicting parameter provides more reliable, and efficient approximation to the actual lag distribution than the one parameter Breitung nonparametric estimator. One parameter and two parameter Breitung estimators are applied to judge the predictive power of monthly Brent crude oil prices (low frequency variable) of the world on the bases of various daily stock market index (high frequency variable) indicators. In our real-time forecasting application, we find that the two parameter Breitung estimator performed well.

Keywords: MIDAS, nonparametric, crude oil, Two-parameter, Forecast

1. Introduction

Generally, time series regression model contains data sampled at the same frequencies. The idea to construct regression model that combines the data with dissimilar frequencies is first time introduced by Ghysels et al., (2004), and well known as MIDAS regression. It is very gradually popular now a days to employ variables observed at different sampling frequencies for economic forecasting and nowcasting, for huge and messive literature, (see Ghysels et al., 2006; Andreo et al., 2011; Monteforte and Moretti, 2013; Banbura et al., 2013; Modugno, 2013). In these types of exercises several HF predictor observations (4 quarters, say) with a single low frequency observation (1, annual). This condition induces the problem that how to forecast the large number of intra period observation for the low frequency process. The simplest way to handle this problem to take the averages of the high frequency observation for an aggregative high frequency predictor for the low frequency dependent variable. Correspondingly, high frequency intra-period observation should be weighted equally while developing the predictors. One of the appealing criteria for this is to plan a more flexible weighting scheme for intra-period observations. One of the examples is to assign larger weights to the recent observation of the daily financial data to forecast the monthly variable (inflation rates). Estimation of the weights scheme by using the least square techniques, however, get untrustworthy estimates due to the multicollinearity among regressors and large number of estimated parameters. The MIDAS approach advocated by Ghysels et al., (2006) is an appropriate and well-designed solution to overcome these problems.

Moreover, MIDAS parametric approach is very useful and popular, but still have an important disadvantage. While, applying the parametric approach, shape of the lag distribution is ruled by an arbitrary class of function like the beta function or an exponential polynomial. Parsimonious specification of this type of parametric function may collapse to get a precise approximation of the lag distribution, in real life empirical applications (see Breitung and Rolling, 2015). By considering the important drawback of standard MIDAS parametric approach, Breitung and Rolling (2015) suggested the nonparametric approach to purely imposes some degree of smoothness to the lag distribution. This nonparametric approach is identical to fitting cubic splines for approximating an unknown functional form. Related approaches to this are very famous in applied statistics. For massive literature on this (see, e.g. Hodrick and Prescott, 1997; or Kalaba and Tesfatsion, 1989).

The SLS estimator (Breitung and Rolling (2015)) has the different interpretation as it can be treated as ridge estimator. If multicollinearity exists, then one of the drawbacks of this ridge estimator is worst quality of fit of regression model and non-satisfaction of the orthogonality relation, so the interpretation of results from this ridge estimator can be unreliable. In this article, we modified the SLS estimator by using the Lipovetsky and Conklin (2005) into two parameter SLS estimator. The resultant two parameter SLS estimator works on the same principle suggested by Breitung and Rolling (2015) but with an additional parameter.

Next section will cover the MIDAS approach outline and benefits of nonparametric approach, our proposed estimators and their comparison criteria. Rest of the article as follows, third section illustrates the simulation, design of experiments and in sample estimation performance. Fourth section represents the predictive performance of the two estimators and their predictive performance on real life application. Whereas, the last section will represent the concluding remarks.

2. Estimation procedure of MIDAS

We can study the regression model that combines the low frequency (LF) y_t with the high frequency (HF) $x_{t,j}$ (regressors), where $t = 1, \ldots\ldots, T$, is the LF time index (say, annually) and j is the intra-period HF (say, monthly) observation with $j = 1, \ldots\ldots, n_t$ $(n_t \epsilon \{1, \ldots 16\}\ in\ our\ example)$. For more simplification, we assume that the time index j runs in the opposite direction; will be, the pair (t, n_t) depicts the first and $(t, 0)$ represents the final observation of the quarter t. We can write the LF and HF in the form linear regression

$$y_{t+\varepsilon} = \alpha_0 + \sum_{j=0}^{p} \beta_j x_{t,j} + \varepsilon_{t+h} \qquad (1)$$

where the ε_{t+h} is uncorrelated with the HF variable $x_{t,0} \ldots., x_{t,p}$. For the generalization commitment, we retain ourselves to a single regressor $x_{t,j}$ so that we can treat β_j as a scaler. Furthermore, the lag-length p is assumed to be smaller than the minimum number of intra-period observations (i.e. $p < \min(n_t) = 18$ in our example with monthly observation) . In Midas approach (Ghysels et al. (2007); Andreou et al. (2011)), we can write the coefficients as $\beta_j = \alpha_1 \omega_j(\theta)$, where the weights are produced (exponential polynomial) can be written as:

$$\omega_j(\theta) = \frac{exp(\theta_1 j + \cdots + \theta_k j^k)}{\sum_{i=0}^{p} \exp((\theta_1 i + \cdots + \theta_k i^k)} \qquad (2)$$

The vector of k hyperparameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)'$ is unknown and estimated by using the beta distribution (Ghysels et al., 2007). The weight $(\omega_j(\theta))$ are always lies in the interval $[0,1]$, as the sum of the all weights equal to 1. Then with the given specification, we can write the model as

$$y_{t+\varepsilon} = \alpha_0 + \alpha_1 \sum_{j=0}^{p} \omega_j(\theta) + \varepsilon_{t+h} \qquad (3)$$

In this background, parameter $\alpha_0, \alpha_1\ and\ \theta_1, \theta_2, \ldots, \theta_k$ can be estimated by using the non-linear least square (NLS). Furthermore, MIDAS approach can also be driven as a restricted ordinary least square (OLS) regression where ‘reduced-form parameters’ in the restricted OLS regression can be written as

$$y_{t+\varepsilon} = \alpha_0 + \beta' x_{t,\blacksquare} + \varepsilon_{t+h} \qquad (4)$$

where regression coefficients on different lags are ($\beta = (\beta_1, \beta_2, \dots, \beta_p)'$) and the regressors on different lags $x_{t,\blacksquare} = (x_{t,0}, x_{t,1}, \dots, x_{t,p})'$ are restricted by the non-linear function $\beta = g(\alpha_1, \theta)$ with the j_{th} component $\beta_j = \alpha_1 \omega_j(\theta)$. Similarly, we can minimize the parameters with the objective, which can be written as

$$S(\alpha, \theta) = \sum_{t=1}^{T} \left[y_{t+\varepsilon} - \alpha_0 - x_{t,\blacksquare}' g(\alpha_1, \theta) \right] \qquad (5)$$

The important part of the $g(\alpha, \theta)$ function is that, it portrays the vector of β that is high dimensional vector and it depends on the low dimensional vector of the parameters (θ). In this research editorial, we propose a two-parameter (SLS) alternative nonparametric approach by using the Lipovetsky and Conklin (2005) technique. This approach doesn't enforce a particular functional form but only specifies that the coefficients β_j is a smooth function of j in the mean that the absolute values of the second differences

$$\nabla^2 \beta_j = \beta_j - 2\beta_{j-1} + \beta_{j-2} \ \ for\ j = 2, \dots, p$$

are small. Where, $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ can be derived by minimizing the penalized least square function and it can be written as,

$$\tilde{S}(\alpha_0, \beta) = \sum_{t=1}^{T} (y_{t+\varepsilon} - \alpha_0 - \beta' x_{t,\blacksquare})^2 + \lambda \sum_{j=2}^{p} (\nabla^2 \beta_j)^2 \qquad (6)$$

where λ is a prespecified smoothing parameter. One of the key feature of this objective function it provides a trade-off between goodness of fit

In the next section we will first present the $\hat{\beta}_\lambda$ by minimizing the equation (6) with respect to smoothing parameter λ and will get the two parameters SLS estimator by using the Lipovetsky and Conklin (2005).

2.1 Midas two parameter approach

By minimizing the equation (6) we can get the SLS estimator

$$\hat{\beta}(\lambda) = (X'X + \lambda D'D) X' y^h \qquad (7)$$

where $r = X'y^h, C = X'X,$ G= $D'D$ and $X = \left[x_{1,\blacksquare}, x_{2,\blacksquare}, \dots, x_{T,\blacksquare}\right]'$, $y^h = [y_{1+h}, y_{2+h}, \dots, y_{T+h}]'$,

and $x_{t,\blacksquare} = \left[x_{t,0}, x_{t,1}, \dots, x_{t,p}\right]'$. Constant term α_0 is ignored in the regression. If we minimize the equation (5), then the objective function is equivalent to

$$\tilde{S}(\gamma) = \left(y^h - \tilde{X}_1\gamma_1 - \tilde{X}_2\gamma_2\right)' \left(y^h - \tilde{X}_1\gamma_1 - \tilde{X}_2\gamma_2\right) + \lambda \gamma_2' \gamma_2 (8)$$

We get the given equation after minimizing the equation 8 with respect to γ_2

$$\tilde{\gamma}_2 = (\tilde{X}_2^{*'} \tilde{X}_2^{*} + \lambda \mathrm{I}_{p-1})^{-1} \tilde{X}_2^{*'} y^h \qquad (9)$$

where $\tilde{X}_2^{*} = M_1 \tilde{X}_2$ and $M_1 = I_T - \tilde{X}_1 (\tilde{X}_1' \tilde{X}_1)' \tilde{X}_1'$. Equation (9) will become the unrestricted OLS estimator if $\lambda = 0$. Same type of resulting form can be received from the equation (1) by taking the sum of square of residuals and minimization with respect to the $\beta_1, \beta_2, \dots, \beta_p$. The equation (1) can be written in the form of

$$y^h = X\beta + \varepsilon\ ,\ \ \varepsilon \sim (0, \sigma^2 I_n) \qquad (10)$$

where y^h is a dependent variable (LF) variable and X is the matrix of HF variable at different time-period h.

In MIDAS for standardized variable, $r = X' y^h\ whereas\ C = X'X$, denote the vector of correlation between LF and HF and correlation matrix of the HF regressors, respectively. Then the unrestricted OLS for the MIDAS corresponds to minimize the sum of square of deviations

$$S^2 = \|\varepsilon\|^2 = \left(y^h - X\beta\right)' \left(y^h - X\beta\right) = 1 - 2\beta' r + \beta' C B, \qquad (11)$$

If we minimize equation (11) with respect to vector of β, then it can be written as

$$C\beta = r, \qquad (12)$$

with the solution of equation (12)

$$\hat{\beta} = C^{-1} r, \qquad (13)$$

Equation (13) is a simple OLS estimator for the regression coefficient estimates,

$$R^2 = 1 - S^2 = \beta' r = \beta' C \beta \qquad (14)$$

whereas, we can represent the equation (12) relation as

$$X'\varepsilon = 0 \qquad (15)$$

We can derive the ridge estimator (Hoerl and Kennard (1970)) in a usual way and it can be written as,

$$\hat{\beta}_\lambda = \left(C + \lambda I_{p-1}\right) r. \qquad (16)$$

Equation (16) represents the same characteristics as available in equation (9) and it is denoted one parameter SLS estimator and this estimator can be treated as ridge estimator. This SLS estimator can come-up with lack of quality of fit of regular MIDAS regression. By using the Lipovetsky (2006) technique we can generalize the estimator given in equation (16) as follows.

$$\hat{\beta}_q(\lambda) = q(X'X + \lambda D'D)X'y^h \qquad (17)$$

where

$$D = \begin{bmatrix} 1 & -2 & 1 & 0 & . & 0 \\ 0 & 1 & -2 & 1 & . & 0 \\ . & . & . & . & . & . \\ 0 & . & . & 1 & -2 & 1 \end{bmatrix} (p-1) \times (p+1)$$

$y^h = [y_{1+h}, y_{2+h} \dots, y_{T+h}]', X = [x_1, x_2, \dots., x_T], and\ x_t = \left[x_{t,0}, x_{t,1}, \dots, x_{t,p}\right]$. We ignore the constant term α_0 for the simplicity. We will compare the proposed SLS estimator with the mentioned Breitung estimator under the MSE criterion. The MSE of the $\tilde{\beta}$ which is an estimator of β will be

$$MSE(\tilde{\beta}) = E\left[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\right] = Var(\tilde{\beta}) + Bias(\tilde{\beta})Bias(\tilde{\beta})'$$

where

$Var(\tilde{\beta}) = E\left[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\right]$

And

$$Bias(\tilde{\beta}) = E(\tilde{\beta}) - \beta$$

The scaler mean square error (mse) is obtained applying the trace operator:

$\text{mse}(\tilde{\beta}) = tr\left(Var(\tilde{\beta})\right) + \left[Bias(\tilde{\beta})\right]'\left[Bias(\tilde{\beta})\right]$

if the two estimators $\tilde{\beta}_1\ and\ \tilde{\beta}_2$ are given, $\tilde{\beta}_2$is considered to be superior $\tilde{\beta}_1$ in the mean of MSE criterion if and only if $MSE\left(\tilde{\beta}_1\right) - MSE\left(\tilde{\beta}_2\right) \geq 0$. If $MSE\left(\tilde{\beta}_1\right) - MSE\left(\tilde{\beta}_2\right)$ is a non-negative (n. n. d) definite matrix then $\text{mse}\left(\tilde{\beta}_1\right) - \text{mse}\left(\tilde{\beta}_2\right) \geq 0$, proved by Theobald (1974). The reverse condition can't necessarily hold true. Consequently, on the basis of this, MSE criterion is considered to be stronger and comparison on the basis of MSE criterion can be more appropriate.

2.2 Comparison of two estimators

In this subsection we will compare SLS estimator with our two-parameter proposed estimator. By considering the spectral decomposition of
decomposition of $C = P\Lambda P$, where Λ is a diagonal matrix whose diagonal elements are eigen values of $X'X$ matrix and P is $p \times p$ matrix whose elements are the eigenvectors of $X'X$ matrix, then the, OLS, SLS and our proposed estimators can be written as

$$\hat{\beta} = (P\Lambda^{-1}P')r, \qquad (18)$$

$$\tilde{\beta}(\lambda) = (P\Lambda P' + \lambda DD')^{-1} r, \qquad (19)$$

$$\tilde{\beta}_q(\lambda) = q(P\Lambda P' + \lambda DD')^{-1} r. \qquad (20)$$

Then the MSE of these estimators will be

$$MSE\left(\hat{\beta}\right) = \sigma^2\, P\Lambda^{-1}P', \qquad (21)$$

$$MSE(\tilde{\beta}(\lambda)) = \sigma^2\, G_\lambda + \left[Bias(\tilde{\beta}_\lambda)\right]'\left[Bias(\tilde{\beta}_\lambda)\right], \qquad (22)$$

$$MSE(\tilde{\beta}_q(\lambda)) = q^2\sigma^2\, G_\lambda + \left[Bias\left(\tilde{\beta}_q(\lambda)\right)\right]'\left[Bias\left(\tilde{\beta}_q(\lambda)\right)\right], \qquad (23)$$

whereas we define,

$G_\lambda = (P\Lambda P' + \lambda DD')'P\Lambda P'(P\Lambda P' + \lambda DD')^{-1}$, $Bias\left(\tilde{\beta}(\lambda)\right) = -\lambda(P\Lambda P' + \lambda DD')\beta$ and $Bias\left(\tilde{\beta}_q(\lambda)\right) = [q(P\Lambda P' + \lambda DD')^{-1}P\Lambda P' - DD']\,\beta$.

Furthermore, we will compute the smoothing parameter λ by minimizing the AIC (see., Breitung and Rolling (2015)) and whereas we will compute the additional parameter q optimal by minimizing MSE of the estimator with respect to q.

Furthermore, for selection criteria, we can say that the matrix $\Lambda = diag(k_1, \ldots, k_p)$ is diagonal matrix whose diagonal elements are the eigenvalues of XX' and T be $p \times p$ matrix whose elements are the eigenvectors of matrix XX' fulfilling $T'XX'T = \Lambda, T'T = DD'$. Afterwards, the original model can be written in canonical form

$$y^h = Z\alpha + \varepsilon_{t+h} \qquad (24)$$

where $Z = XT, \alpha = T'\beta \text{ and } Z'Z = T'XX'T = \Lambda$. Then, $\hat{\alpha}_q(\lambda) = T'\tilde{\beta}_q(\lambda)$ and $MSE\left(\hat{\alpha}_q(\lambda)\right) = T'MSE(\tilde{\beta}_q(\lambda))T$, then the MSE of the estimator $\hat{\alpha}_q(\lambda)$ can be written as

$$MSE\left(\hat{\alpha}_q(\lambda)\right) = q^2\sigma^2(\Lambda + \lambda DD')^{-1}\Lambda(\Lambda + \lambda DD')^{-1} + [q(\Lambda + \lambda DD')^{-1}\Lambda - DD']\alpha\alpha'[q(\Lambda + \lambda DD')^{-1}\Lambda - DD']'$$

We can get the optimal values of the λ and q, by minimizing the above equation with respect to λ and q, respectively.

3. Design of experiment, simulation and estimation performance

In this section, we will compare the small sample properties of nonparametric variant of SLS estimator and two parameter SLS estimators of the MIDAS regression. For comparative study, we generate data similarly to the Andreou et al., (2010) given as

$$y_{t+h} = \beta_0 + \sum_{j=0}^{p}\beta_j x_{t,j} + \varepsilon_{t+h} \qquad (25)$$

$$\beta_j = \alpha_1\omega_j(\theta) \qquad (26)$$

$$\varepsilon_{t+h} \sim N_{iid}(0, 0.125) \qquad (27)$$

where $t = 1,2,\ldots,T$, $\beta_0 = 0.5, and\ \omega_j(.)$is a weighting function which can be chosen in several specifications, details are presented below. The high frequency regressor is generated by the $AR(1)$ process given below

$$x_{t,j} = \alpha_0 + \varrho x_{t,j-1} + \varepsilon_{j,t}, \qquad \varepsilon_{j,t} \sim iid\ N(0,1) \qquad (28)$$

where $j = 0,1,\ldots,p$, $x_{t,j-p-k} = x_{t-1,j-k}$ for all $k > 0$. Correspondingly, $x_{t,j}$ denotes the j_{th} lag of the $AR(1)$ series $x_{t,0}$. As suggested by Andreou et al., (2010), we will take $\alpha_0 = 0.5$ and $\varrho = 0.9$.

We choose $T \sim \{100,200,400\}$ different sample sizes and high frequencies lags are $p + 1 \sim \{20,40,60\}$. The scale parameter given in equation (10) is chosen as $\alpha_1 \sim (0.2, 0.3, 0.4)$, respectively to model of small, medium and large signal to noise ratio. Furthermore, we replicated this experiment 5000 times by considering the Monte Carlo simulation.

At first step, we selected the different functions and respective equations are given in the next section. At second step, we considered the ratio of mean MSE of boosted two parameter SLS estimator nonparametric approach with the SLS estimator nonparametric approach. Assuming the SLS nonparametric approach superiority on the usual parametric approach we only compared the SLS with the boosted two parameter SLS nonparametric approach. The ratio given in the Table 1 illustrates that two-parameter estimator is more superior and perform better than the SLS. Further, results can be verified in Table 1.

3.1. Results

All of the analysis Monte Carlo Simulation we considered a lag distribution that fits into the nonparametric and parametric MIDAS framework equally (Breitung and Roling, 2015). First experiment was conducted by using the exponentially weight function with

$$\omega_j(\theta) = \frac{\exp\,(\theta_1 j + \theta_2 j^2)}{\sum_{i=0}^{p} \exp\,(\theta_1 i + \theta_2 i^2)}, j = 0,1 \dots p$$

where we follow Andreou et al., (2010) and set $\theta_1 = 7 \times 10^{-4}$ and $\theta_2 = -6 \times 10^{-3}$. The Table 1 shows the MSE ratio of the proposed nonparametric SLS estimator with an additional parameter relative to the SLS estimator.

We done the second experiment by considering the hump shaped weight give as

$$\omega_j(\theta) = \frac{\exp\,(\theta_1 j + \theta_2 j^2)}{\sum_{i=0}^{p} \exp\,(\theta_1 i + \theta_2 i^2)}, j = 0,1 \dots p$$

We govern the parameter such that the weighting function reaches maximum at $j = 6, j = 10,\ j = 16$, when lags are 20, 40 and 60 lags are retained, respectively. We chosen $\theta_1 = 8 \times 10^{-2}\ and\ \theta_2 = \frac{\theta_1}{10}, \theta_2 = \frac{\theta_1}{20}, and\ \theta_2 = \frac{\theta_1}{30}$, respectively.

The third experiment is run by utilizing the sign changing weight function given as

$$\omega_j(c_1, c_2) = \frac{c_1}{p+1}\left[\sin(c_2 + \frac{j2\pi}{p})\right]$$

Where $c_2 = 1 \times 10^{-2}, c_1 = 5, c_1 = 2.5, and\ c_1 = 5/3$, for the lags 20, 40, and 60 respectively, we the constraint that these weights sum to unity.

Table 1: In sample MSE ratios

High frequency lags		$p+1=20$			$p+1=40$			$p+1=60$		
Exp. declining weights mse($\hat{\alpha}_q(\lambda)$)/MSE(SLS_1)	**T**	$\alpha_1 = 0.2$	$\alpha_1 = 0.3$	$\alpha_1 = 0.4$	$\alpha_1 = 0.2$	$\alpha_1 = 0.3$	$\alpha_1 = 0.4$	$\alpha_1 = 0.2$	$\alpha_1 = 0.3$	$\alpha_1 = 0.4$
	100	0.7271	0.6172	0.6021	0.7212	0.7023	0.5023	0.8223	0.7523	0.7012
	$\hat{q}_{opt}$	12.23	8.2402	6.7200	5.960	4.960	2.1300	4.3211	3.2111	2.1601
	AIC	2.555	1.640	2.7500	2.540	1.630	1.7401	2.542	1.542	0.740
	200	0.8231	0. 8112	0.8004	0.9312	0.8921	0.8108	0.9024	0.9016	0.8015
	$\hat{q}_{opt}$	8.230	7.250	5.380	3.250	5.510	1.5401	1.2301	1.2012	1.430
	AIC	2.230	3.121	2.450	2.230	1.455	2.440	2.220	1.335	1.311
	400	0.2561	0.2667	0.3222	0.3562	0.3210	0.3001	0.2901	0.2871	0.341
	$\hat{q}_{opt}$	6.600	2.140	2.510	5.660	6.320	4.120	5.519	5.001	2.791
	AIC	2.920	2.031	1.150	2.920	1.020	1.004	2.920	3.021	3.140
Humped-shaped weights mse($\hat{\alpha}_q(\lambda)$)/MSE(SLS_1)	**100**	0.2830	0.2139	0.2097	0.367	0.2012	0.2311	0.1778	0.450	0.026
	$\hat{q}_{opt}$	16.280	13.210	11.390	17.25	11.000	8.170	12.070	10.520	8.460
	AIC	1.555	1.640	1.750	2.540	2.630	2.740	3.542	2.542	1.740
	200	0.180	0.190	0.130	0.190	0.170	0.160	0.210	0.130	0.120
	$\hat{q}_{opt}$	18.99	17.620	15.740	15.74	10.49	17.74	10.85	8.690	17.92
	AIC	2.23	2.231	2.456	2.235	3.352	2.441	2.221	1.325	2.441
	400	0.201	0.180	0.170	0.234	0.420	0.420	0.512	0.53	0.440
	$\hat{q}_{opt}$	17.92	15.34	14.30	19.885	17.42	15.33	10.11	17.69	16.17
	AIC	2.921	3.031	3.150	5.920	4.020	3.140	3.920	3.021	3.140
Sign-changing weights mse($\hat{\alpha}_q(\lambda)$)/MSE(SLS_1)	**100**	0.950	0.931	1.431	1.391	1.155	0.980	0.980	0.981	0.954
	$\hat{q}_{opt}$	1.131	2.610	1.011	1.021	1.312	1.101	1.501	1.141	1.053
	AIC	1.556	1.640	1.750	1.540	2.630	2.740	2.542	2.542	2.742
	200	0.981	0.661	1.100	0.762	0.730	0.791	0.637	0.508	0.410
	$\hat{q}_{opt}$	1.031	1.011	1.101	1.942	1.231	1.150	3.000	3.000	2.000
	AIC	1.439	1.431	1.654	1.530	1.590	1.540	1.235	3.325	2.441
	400	0.701	1.710	2.021	0.261	0.123	0.745	0.191	0.147	0.360
	$\hat{q}_{opt}$	2.321	2.079	2.021	2.163	2.224	1.511	2.460	2.123	2.201
	AIC	3.4224	2.8403	3.256	3.425	3.860	4.222	3.121	4.861	4.260

Note: The entries are the MSE ratios of nonparametric two parameter SLS estimator relative to the nonparametric SLS estimator . Data are generated according to the equation 25-28. Both of the estimators are computed by minimal choice of smoothing parameter λ by using the Breitung and Roling (2015), modified AIC technique. The number of replications is 5000.

The results of Table 1 are summarized as follows. First, the proposed nonparametric SLS estimator with an additional parameter q evidently dominates the SLS nonparametric approach in-sample estimation for all the weights function. It is evidently noted that, proposed SLS estimator outperformed the SLS estimator in sample estimation in hump shaped weight function. For the hump shaped weights, the proposed SLS estimator yields more accurate estimates and less MSE.
Concerning the exponentially declining weights, the performance of the nonparametric MIDAS approach also improves as the sample size T increases which is for sure expected. The same is applicable for all the weight functions, except some point in sign changing weights, where nonparametric approach with SLS estimator evidently dominates when the sample size was 200. One thing we more noted that, larger value of $\hat{q}_{opt}$ yields more efficient results. But from the result in Table 1, it is noted that as the sample size increased, we can see the decline in the $\hat{q}_{opt}$ value. We believe that the for large sample size $\hat{q}_{opt}$ will be equal to 1 and proposed estimator can be equal to the SLS estimator at some point. In these results, we adopted the same criteria by minimizing the AIC similar to Breitung and Roling (2015), nonparametric approach.
In Table 1, we were concerned for estimating the weights β_j in equation (26) and we observed the domination of the proposed SLS estimator in-sample estimation. Now at next step, we move to evaluate the accuracy of out-of-sample forecast made by the nonparametric, by using the SLS and proposed SLS estimator in the sample model. For this, we partitioned the whole sample into in sample estimation, comparing observation $1,2,\dots.T^e$, and a forecasting sample consisting of observation $T^e+1,2,\dots T$. In this exercise we set $T_e = T/2$.
The estimation sample is used to achieve a baseline estimate of the weights β_j from the sample as follows

$$\hat{y}_{t+h} = \sum_{j=0}^{p} \hat{\beta}_{j:T_0} x_{t,j} + \hat{u}_{t+h}, for\ t = 1, \dots . T^e$$

Whereas the given baseline estimates, one step ahead forecast of the target variable (dependent variable) is given as

$$\hat{y}\, T^e + 1 \,\vdots\, T^e = \sum_{j=0}^{p} \hat{\beta}_{j:T^e} x T^e, j$$

We then include the next period in the estimation sample while dropping the first period, the with the same model and we obtain the next one-step- ahead forecast.

4. Empirical application: Forecasting the Monthly Brent crude oil price by using the daily AMEX oil Index

Daily predictors

Crude oil is a type of energy and chemical material source, plays vital role in global commodity market and in the expansion of worldwide economy. In modern years, relationship between crude oil markets and stock markets has appealed extensive attention, and it is proved by the large number of studies that crude oil markets are attentively related to the stock markets; see e.g., (Wang and Liu, 2016; Pan et al; 2016; Broadstock and Filis, 2014; Mensi et al., 2013; Creti et al., 2013; Du and He, 2015) .
In our empirical application, we mentioned earlier we emphasis on the most reliable daily indicator stock market indices. In the empirical literature, due to close associations between crude oil market and stock markets, an enormous number of scholars use stock market indices information to forecast crude oil prices see (Guglielmo et al., 2015; Chen, 2014; Liu et al., 2015). Recently, Zhang and Wang (2019), forecasted the WTI and Brent crude oil prices monthly (LF variable) on the basis of daily and weekly (HF variable) of four

stock markets indices. Amex Oil indices daily indicator was remained the superior indices to forecast the crude oil price.
In our empirical exploration, we therefore focus on the Amex oil indices as a daily predictors of crude oil price. We thus examine the predictive power of the crude oil prices employing the SLS (non-parametric) and two parameters proposed SLS MIDAS framework.

4.1 Preliminary analysis In-sample results of MIDAS regressions

The target variable is the monthly Brent crude oil price, ranged from May 1993 to March 2017 and our daily predictor is the Amex oil stock index. Amex oil stock index high frequency predictor was used to forecast the Brent crude oil (Zhang and Wang, 2019). They suggested that the, Amex oil stock index (daily predictor) is superior and powerful predictor as compare to all other index for predicting the Brent crude oil price (LF variable). During these analyses. we observed that the number of daily observations in a month was not consistent, months have had different lags (from 28 up to 31 days) and some observation was missing due to different holidays, and weekends. We judged the number of observation ranges from 18 to 22. As we don't have thumb rule to resolve this problem, earlier researcher solved these in different ways. Andreou et al., (2013) considered the 22 trading days in a month for checking the predictive power of financial time series models. So, we fixed the 18 observations in a month, we run the model after estimating the missing observation by linear interpolation methods. Table 2 represents the estimation results of for the parametric MIDAS regression using the Amex oil stock index as the predictor where

$$\omega_j(\theta) = \frac{\exp\,(\theta_1 j + \theta_2 j^2)}{\sum_{i=0}^{p} \exp\,(\theta_1 i + \theta_2 i^2)}$$

The estimated parameter for forecast horizon $h = 1$ and lag lengths $p \epsilon \{10, 20, 40, 60\}$ are reported along with their t-statistics. Excitingly, a low R^2 for the 0.037 reported for the Amex index daily predictor for lag length 10, Whereas, the R^2 is increased as the lag length increased, which shows that the daily Amex stock index is useful predictor to forecast the Brent crude oil prices. Unrestricted OLS estimation and simulation results already suggests using the hump shaped weight function, and the nonparametric one parameter SLS estimator, and two parameter SLS estimator smooth out these erratic unrestricted estimates.

Table 2. In sample estimation of MIDAS regression: $y_{t+h} = \alpha_0 + \alpha_1 \sum_{j=0}^{p} \omega_j(\theta) x_{t-j} + u_{t+h}$

	$p+1=10$	p+1=20	p+1=40	p+1=60
α_0	150.9 (5.459)	152.9 (5.553)	153.1 (5.566)	153.4 (5.570)
α_1	0.0321 (4.352)	0.0425 (5.594)	0.0559 (0.218)	0.1057 (13.92)
θ_1	131.5 (0.002)	80.14 (0.322)	80.37 (0.393)	80.42 (0.652)
θ_2	-6.629 (-0.002)	-2.2679 (-0.319)	-2.275 (-0.390)	-2.280 (-0.647)
$R^2(SLS)$	0.03789	0.0986	0.1743	0.2935
$R^2(Proposed\ estimator)$	0.03959	0.1197	0.2034	0.3512

4.3 Out of Sample Forecast

Out of sample forecast by Kuzin et al. (2009), strategy is executed to direct forecast the monthly crude oil prices on the bases of Amex oil index daily predictor. For this approach, we regress the future values of the dependent variable (mathematically denoted as y_{t+h}) on current or past values of the regressor. We now crack to forecast monthly Brent crude

oil prices of the market. First step, we split the sample into an estimation sample $t = 1, \dots \dots, T^e$ and a forecasting sample $t = T^e + 1, \dots \dots T$, where $n_f = T - T^e$ denote the number of forecast. The estimation sample runs from May 1993 to December 2012, and the forecasting exercising executed from January 2013 to March 2017. We assess forecasts according to the root mean squared forecast error.
Table 3 shows the forecasting accuracy of the usual nonparametric SLS estimator relative to the proposed two parameter SLS estimator for different choice of lag lengths. From the Table 3, it can be observed that the proposed two parameter SLS estimator produces more accurate forecast as compare to the usual nonparametric SLS estimator for the MIDAS regression with small RMSE value . All of the forecast were uninformative provided that the RMSE is close to the estimated standard deviation (80.02) of the target variable (Brent crude oil prices).

Table 3: Out of sample forecast comparison

horizon		$\boldsymbol{p+1 = 10}$	**p+1=20**	**p+1=40**	**p+1=60**
1	$\boldsymbol{RMSE(SLS)}$	79.50	79.66	79.36	78.69
	$\boldsymbol{RMSE(Proposed)}$	79.48	79.36	79.11	77.78

5. Conclusion

Now a days, MIDAS regression is widely used in financial time series models. In Midas regression, we have to combine the LF and HF variables at different frequencies, and it can induce the difficulty and unreliable results while estimating the autoregressive lag distribution where LF observations are considered as an effective sample size. By considering the nonparametric SLS estimator (Breitung and Rolling (2015)) that gives the advantage over the usual NLS estimator because of its smoothness. But this estimates can lose predicting power when we have larger lag length, we proposed an alternative estimator with an additional parameter q, and we choose the MSE as a performance criteria to verify its utility over SLS estimator. Monte Carlo simulation and real application suggests that the our proposed estimator with and additional parameter substantially produces the lower MSE as compare to the usual SLS estimator. To assess the predictive power of daily indicators for the Brent Crude oil prices, applied nonparametric SLS estimator and proposed SLS estimator to a sample of 24-year data. It turns out that Amex Index is a useful one-month-ahead forecast predictor for the Brent crude oil prices. The estimated lag distribution covers around 30 days and is hump shaped with a maximum 6 to 9 lags.

References

Andreou, E., Ghysels, E., & Kourtellos, A. (2011). Forecasting with mixed-frequency data. In *The Oxford handbook of economic forecasting*.

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

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting* (Vol. 2, pp. 195-237). Elsevier.

Breitung, J., & Roling, C. (2015). Forecasting inflation rates using daily data: A nonparametric MIDAS approach. *Journal of Forecasting*, *34*(7), 588-603.

Broadstock, D. C., & Filis, G. (2014). Oil price shocks and stock market returns: New evidence from the United States and China. *Journal of International Financial Markets, Institutions and Money*, *33*, 417-433.

Chen, S. S. (2014). Forecasting crude oil price movements with oil-sensitive stocks. *Economic Inquiry*, *52*(2), 830-844.

Creti, A., Joëts, M., & Mignon, V. (2013). On the links between stock and commodity markets' volatility. *Energy Economics*, *37*, 16-28.

Dias, A., & Embrechts, P. (2004). Dynamic copula models for multivariate high-frequency data in
finance. *Manuscript, ETH Zurich*, *81*.

Du, L., & He, Y. (2015). Extreme risk spillovers between crude oil and stock markets. *Energy Economics*, *51*, 455-465.

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

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, *131*(1-2), 59-95.

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

Hamilton, J. D. (2008). Oil and the Macroeconomy. *The new Palgrave dictionary of economics*, *2*.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1-16.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67.

Kalaba, R., & Tesfatsion, L. (1989). Time-varying linear regression via flexible least squares. *Computers & Mathematics with Applications*, *17*(8-9), 1215-1245

Kuzin, V., Marcellino, M. G., & Schumacher, C. (2009). MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area.

Lipovetsky, S. (2006). Two-parameter ridge regression and its convergence to the eventual pairwise model. *Mathematical and Computer Modelling*, *44*(3-4), 304-318.

Lipovetsky, S., & Conklin, W. M. (2005). Ridge regression in two-parameter solution. *Applied Stochastic Models in Business and Industry*, *21*(6), 525-540.
Liu, L., Ma, F., & Wang, Y. (2015). Forecasting excess stock returns with crude oil market data. *Energy Economics*, *48*, 316-324.

Mensi, W., Beljid, M., Boubaker, A., & Managi, S. (2013). Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold. *Economic Modelling*, *32*, 15-22.

Modugno, M. (2013). Now-casting inflation using high frequency data. *International Journal of Forecasting*, *29*(4), 664-675.

Monteforte, L., & Moretti, G. (2013). Real-time forecasts of inflation: The role of financial variables. *Journal of Forecasting*, *32*(1), 51-61.

Pan, Z., Wang, Q., Wang, Y., & Yang, L. (2018). Forecasting US real GDP using oil prices: A time-varying parameter MIDAS model. *Energy Economics*, *72*, 177-187.

Theobald, C. M. (1974). Generalizations of mean square error applied to ridge regression. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(1), 103-106.

Wang, Y., & Liu, L. (2016). Crude oil and world stock markets: volatility spillovers, dynamic correlations, and hedging. *Empirical Economics*, *50*(4), 1481-1509.

Zhang, Y. J., & Wang, J. L. (2019). Do high-frequency stock market data help forecast crude oil prices? Evidence from the MIDAS models. *Energy Economics*, *78*, 192-201.