Multivariate mixed frequency data sampling model (m-MIDAS)[1]

Fok Ming Cheung

This section introduces the mathematic specification of the model, which has been extended from the univariate model to multi-variables with the same frequencies and articulates all necessary conditions to perform optimization. Relevant derivations can be found in the appendix [A]. Lastly, details about the estimation package are available in appendix [B].

A general review of the model

Since economic data generally comes with different frequencies, especially financial data are usually available daily whereas Macroeconomic data are only accessible quarterly. In view of this, the conventional approach is to aggregate high-frequency observations by taking simple average into lower frequency samples. This approach has been strongly criticized for compromising valuable information contained within the high-frequency observation, for example, Ghysel, Santa-Clara and Valkanov (2002) had shown that the application of linear regression model consistently underestimates the impact of macroeconomic variables on financial volatility. Therefore, an estimation approach which considers the non-linear relationship between data in different frequencies is needed in order to reliably examine the validity of economic theories.

Since the (univariate) Mixed frequency data sampling (MIDAS) was firstly introduced in Ghysels, Santa-Clara, and Valkanov (2004), this model has been widely applied in different studies, for example Engle, Ghysels and Sohn (2013) examined the linkage of macroeconomic data and stock index volatility, Asimakopoulos (2017) evaluate the relationship of dividend-price ratio and dividend growth predictability etc. Later, the model has been further modified. Carriero, Clark and Marcellino (2015) developed the Bayesian MIDAS to forecast quarterly GDP growth by using a range of monthly economic indicators, Bacchiocchi, Bastianin, Missale and Rossi (2018) constructed the MIDAS-Structural Vector autoregressive (MIDAS-SVAR) to study the impact of financial volatility and monetary policy on the gross capital inflow in the US market, Ghysels, Santa-Clara and Valkanov (2006) found that high-frequency realized volatility to be the best predictor of future (daily) volatility by using the MIDAS-GARCH model, and the MIDAS-GARCH has been further developed as the Dynamical Component Correlation MIDAS (DCC-MIDAS) by Colacito, Engle and Ghysels (2011) to study the determination of the short-run as well as long-term correlation between 10 year bond and Manufacturing and Retails portfolios in different countries. Yet, most of these studies are limited to univariate setting, and the sole purpose of this paper is to extend the mode to multivariate cases.

Apparently, the MIDAS approach suffers from parameters proliferation. As such, notations used in this paper are slightly different from those used in Ghysel (2002) for the reason of avoiding confusions. Low-

[1] This is an extraction of my thesis - *Is dividend policy expectation or errors driven? A new perspective of asymmetric information with mixed frequency evidence*

frequency data, which are sampled by a step size of t , are denoted as$Y_t^{(L)}$; whereas high frequency regressors, denoted as $X_t^{(h)}$, are sampled at $1/t$. For example, if $Y_t^{(L)}$ is a monthly series, then $L = 1/12$ and $X_t^{(h)}$ contains $h = 22$ observations if it is a daily series. Consider if there are N regressors, each regressors contains J lags and P lags of variable Y_t, then the regression model is written as:

$$Y_t^{(L)} = \beta_0 + \sum_{i=1}^{N}\sum_{j=1}^{J} B_i f(\theta) X_{i,t-j}^{(h)} + \sum_{p=1}^{P} \rho_p Y_{t-p}^{(L)} + \epsilon_t \tag{1}$$

, where $E(\epsilon_t) = N(0, \Sigma)$ and Σ denotes the co-variance matrix.

Superficially, the model resembles the format of any ordinary linear regression models. However, as the dimension on both sides of the equation does not match so one cannot simply apply least-square optimization. The key of mixed frequency regression lies in the specification of the polynomial function, $f(\theta)$. Ghysels (2012) suggested several options for the specification, which range from (i) the exponential Almon lag, (ii) beta function. (iii) the step function to (iv) the unrestricted cases. Although these functions are generally concave, one should consider the applicability of these specifications on a case-by-case basis. This paper focuses on only one specification, the exponential Almon lag function, as literatures, like Asimakopoulos. etal (2017), suggested that the Almon lag function has higher flexibility in generating different shapes of the function which is relatively suitable to volatile financial data. To alleviate the influence of parameter proliferation, it has been proposed that the exponential Almon lag function will be controlled by two hyper-parameters (θ_1, θ_2).

$$\omega(j, \theta_1, \theta_2) = \frac{e^{\theta_1 j + \theta_2 j^2}}{\sum_{j=1}^{J} e^{\theta_1 j + \theta_2 j^2}} \tag{2}$$

, where j denotes the number of lags. Intuitively, one can notice two particular features associated with the function. On the one hand, as the exponential function constructs or explodes for 'too large' or 'too small' parameters, a significant but not exaggerated weight (ω)[2] can only be generated by hyperparameters bounded in the domain $[-1,1]$, such that $\theta \in [-1,1]$. On the other hand, the function embodies the 'memory loss' feature, meaning that the weight of each observation decays according to the time lags, i.e. $\lim_{j\to\infty} \omega(\theta, j) = 0$.

Like ordinary least square, the NLS estimators minimize the sum of squared residuals[3]. By rearranging equation (1), such that

$$E\left[\left(Y_t^{(L)} - \left(\beta_0 + \sum_{i=1}^{N}\sum_{j=1}^{J} \beta_i \omega(\theta) X_{i,t-j}^{(h)} + \sum_{p=1}^{P} \rho_p Y_{t-p}^{(L)}\right)\right)^2\right] = E[\epsilon_t^2] = 0 \tag{3}$$

, and

[2] One can easily verify that $\lim_{\theta\to-\infty} \omega(\theta, k) = 0$ and $\lim_{\theta\to\infty} \omega(\theta, k) = \infty$. In either case, the function fails to generate significant and reasonable weights.

[3] Brandon (2010) provided details about NLS using Quasi-maximum likelihood estimation (QMLE) and Generalized method of moments(GMM). But these approaches are covered in this paper.

$$\hat{\theta} = \underset{\hat{\theta}}{argmin} \sum_{t=1}^{T} \epsilon_t^2 \tag{4}$$

Although the NLS estimator resembles with the OLS estimator, the estimation method is different. According to Davidson and MacKinnon (2004)[4], the NLS estimators (θ) are asymptotically identified, which means

$$\alpha(\hat{\theta}) = \underset{t\to\infty}{\text{plim}} \frac{1}{t}(\epsilon_t^2). \tag{5}$$

, where $\alpha(\hat{\theta}) = 0$ if and only if $\hat{\theta} = \theta_0$. And the estimators must be asymptotically consistent and normal so that $\underset{t\to\infty}{\text{plim}} \frac{1}{t}\left(\omega(\theta_0) - \omega(\hat{\theta})\right) = 0$. The importance of both criteria becomes apparent when extending to multivariate analysis, if regressors are linearly dependent then the asymptotic estimators cannot be uniquely identified, thereby, estimators are not consistent. A preliminary proof of relevant discussion is provided in Appendix [A] and summary of the estimation package is also available in Appendix [B].

Because of the polynomioal function, Ghysels and Marcellino (2018) stated that standarad statistical tests of distributed lag terms and slope-parameters ($\widehat{\beta_\iota}$) do not apply in MIDAS model. Since the polynomial function maps all high-frequency data into low-frequency vectors, marginal changes in the number of lags in high-frequency observation (h) are not penalized by common information criterion functions, like Akaike (AIC), Schwarz (BIC) etc..However, these tests remain valid when examining lagged low-frequency dependent variables, $y_{t-p}^{(L)}$. Furthermore, the slope-parameters ($\widehat{\beta_\iota}$) depends on the hyperparameters ($\hat{\theta}$) so there are no standard tests to examine $H_0: \beta_i = 0$. Rather, Davies (1987) suggested the t-value should be computed by simulating the distribution of $\widehat{\beta_\iota}$ over a finite range of θ so that $t_{max} = \sup_i t(\widehat{\beta_\iota}(\theta))$.

[4] Chapter 6 section 2 -3

Appendix [A] <u>Proof of multi-collinearity and inconsistent estimator</u>

A set of NLS estimators (θ_0) is said to be unique if there is no another set of estimators ($\hat{\theta}$) which can satisfy equation (3). For simplicity, we denote $Y_t^{(L)}$ as $X(\theta_0)$ and $X(\hat{\theta}) = \beta_0 + \sum_{i=1}^{N}\sum_{j=1}^{J} B_i f(\hat{\theta}) X_{i,t-j}^{(h)} + \sum_{p=1}^{P} \rho_p Y_{t-p}^{(L)}$. Then, equation (3) can be rearranged as

$$\alpha(\hat{\theta}) = \underset{t\to\infty}{\text{plim}} \frac{1}{t}\left(X(\theta_0) - X(\hat{\theta}) + \epsilon_t\right)^2$$

$$= \underset{t\to\infty}{\text{plim}} \frac{1}{t}\left(\sum_{i=1,j-1}^{N,J} \beta_i x_{i,t-j}^{(H)}\, \omega_i(\theta_0 - \hat{\theta}) + \epsilon_t\right)^2$$

, where $\alpha(\hat{\theta}) = 0$ if $\hat{\theta} = \theta_0$. Consider that one regressor can be represented as a linear combination of other regressors, for example $x_{a,t} = \psi x_{b\neq a,t}$. Then

$$= \underset{t\to\infty}{\text{plim}} \frac{1}{t}\left(\sum_{j=1}^{J} \beta_1 x_{1,t-j}^{(H)}\, \omega_1(\theta_0 - \hat{\theta}) + \cdots + \sum_{j=1}^{J} \beta_a x_{a,t-j}^{(H)}\, \omega_a(\theta_0 - \hat{\theta}) + \sum_{j=1}^{J} \beta_b x_{b,t-j}^{(H)}\, \omega_b(\theta_0 - \hat{\theta}) + \cdots + \epsilon_t\right)^2$$

$$= \underset{t\to\infty}{\text{plim}} \frac{1}{t}\left(\sum_{j=1}^{J} \beta_1 x_{1,t-j}^{(H)}\, \omega_1(\theta_0 - \hat{\theta}) + \cdots + \psi \sum_{j=1}^{J} (\beta_a + \beta_b)(\omega_a(\theta_0 - \hat{\theta}) + \omega_b(\theta_0 - \hat{\theta})) x_{b,t-j}^{(H)} + \cdots + \epsilon_t\right)^2$$

, where $\omega_b(\hat{\theta}^*) = (\omega_a(\theta_0 - \hat{\theta}) + \omega_b(\theta_0 - \hat{\theta}))$. Then there is another set of estimators ($\hat{\theta}^*$) which can also satisfy the condition $\alpha(\hat{\theta}^*) = 0$. So, $\hat{\theta}^* = \theta_0$ but $\hat{\theta}^* \neq \hat{\theta}$ violates the condition of uniqueness and consistency.

Appendix [B] <u>Introduction to m-MIDAS estimation package</u>

In the appendix, section (1) provides a general summary of key functions involved in the estimation, and section (2) illustrates the technical procedure behind the estimation algorithm.

I. Summary table of functions

<table>
<tr><th>Function name</th><th>Input/ Description</th><th>Output</th></tr>
<tr><td>mMIDAS_master</td><td>- This is the main function</td><td></td></tr>
<tr><td>m1MIDAS_ADL</td><td>- This function generates the estimation result.

- Inputs:
<table>
<tr><td>[DataY,DataYdate];
[DataX,DataXdate]</td><td>- Explanatory variable and regressors.
- Date vector</td></tr>
<tr><td>Xlag;
Ylag</td><td>- No. of lags in high-and-low freq. variables</td></tr>
<tr><td>Horizon = 3;</td><td>- MIDAS lead/lag specification</td></tr>
<tr><td>[EstStart, EstEnd];</td><td>- The start and end date of the estimation</td></tr>
</table>
※</td><td>Estimation result</td></tr>
<tr><td>mMixFreqData.m</td><td>This function aligns the date of both high-and-low freq. variables given the specified lag numbers</td><td>- 3-dimensional matrix of regressors $[X^H_{(N,H,t-j)}, Y^L_{(T,t-p)}]$
- Vector of explanatory variable, Y^L_t</td></tr>
<tr><td>MIDAS_Estimate.m</td><td>- This is the main estimation algorithm which depends on the pre-specified polynomial function</td><td></td></tr>
<tr><td>midas_X.m</td><td>- This function maps high-frequency data into low-frequency vector for a given pair of hyperparameters (θ_1, θ_2).</td><td>- Matrix of $\omega(\theta_1,\theta_2)X^H_{(N,H,t-j)}$
- Matrix of weights $(\omega(\theta_1,\theta_2))$</td></tr>
<tr><td>mfrvobj_adl.m;
Jacob.m</td><td>- These functions are the objective functions which govern the non-linear least square (lsqnonlin.m);
- mfrvobj_adl.m: generates the estimation residuals for given pairs of (β,θ).
- Jacob.m: constructs the Jacobian matrix, $\left[\frac{\partial \epsilon_t}{\theta}\right]$</td><td>- Residuals (ϵ_t)
- Jacobian matrix
- Matrix of optimized weights $(\omega(\widehat{\theta_1},\widehat{\theta_2}))$</td></tr>
</table>

※ The original program developed by Ghysels (2012) provides different options of estimation methods, polynomial functions, and the inclusion of exogenous variables, but these options are not available in this package.

II. Estimation algorithm

Step (1) Data preparation

1. Explanatory variable vector

$$Y^{L}_{(T,1)} = \begin{bmatrix} y^{L}_{t=1} \\ \vdots \\ y^{L}_{t=T} \end{bmatrix}$$

2. Matrix of high frequency regressors

$$X^{H}_{(N,p,j)} = \begin{bmatrix} \begin{bmatrix} x^{H}_{i=1(h-j,t)} & \cdots & x^{H}_{i=1(h,t)} \\ \vdots & \ddots & \vdots \\ x^{H}_{i=1(h-j,t-p)} & \cdots & x^{H}_{i=1(h,t-p)} \end{bmatrix} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \begin{bmatrix} x^{H}_{i=N(h-j,t-1)} & \cdots & x^{H}_{i=N(h,t-1)} \\ \vdots & \ddots & \vdots \\ x^{H}_{i=N(h-j,t-p)} & \cdots & x^{H}_{i=N(h,t-p)} \end{bmatrix} \end{bmatrix}$$

Step (2) Optimization with Newton method

1. Initiate the estimation with a pre-specified vector of hyperparameters

As suggested in Ghysels (2012), it is reasonable to specify the initial value of the hyperparameters $(\widehat{\theta_1}, \widehat{\theta_2}) = (-1,0)$.

2. Mapping high-frequency regressors into low-frequency

Each high-frequency regressor can be converted into a low-frequency vector in the following way.

$$X^{(L)}_{t}(\hat{\theta})_{(p,N)} = X^{(H)}_{(N,(p,j))} \begin{bmatrix} \begin{bmatrix} \omega_{h-j}(-1,0) \\ \vdots \\ \omega_{h}(-1,0) \end{bmatrix}_{i=1} & \cdots & \begin{bmatrix} \omega_{h-j}(-1,0) \\ \vdots \\ \omega_{h}(-1,0) \end{bmatrix}_{i=N} \end{bmatrix}'_{((j,1),N)}$$

3. Given the pre-specified θ, compute the slopes (β_i) of high-frequency variables using OLS

$$\widehat{\mathrm{B}}|_{\hat{\theta}} = \left(X^{(L)^T} X^{(L)}\right)^{-1} (X^{(L)^T} Y^{(L)})$$

4. Optimizing hyperparameters using NLS

Given the $\widehat{\mathrm{B}}|_{\hat{\theta}}$, the optimization algorithm (newton method) examines the gradient of hyperparameters $(\hat{\theta})$ with respect to the estimation residuals, which is represented as the following Jacobian matrix.

$$\mathrm{J}(\hat{\theta})|_{\hat{\beta}} = -\begin{bmatrix} \dfrac{\partial \epsilon_t}{\partial \widehat{\beta_0}}_{(p,1)} & \dfrac{\partial \epsilon_t}{\partial \widehat{\beta_\iota}}_{(p,N)} & \begin{bmatrix} \widehat{\beta_\iota} \dfrac{\partial X^{(L)}_{i}(\theta)}{\partial \theta} \end{bmatrix}_{(p,n\times 2)} & \begin{bmatrix} Y^{(L)}_{t-1} & \cdots & Y^{(L)}_{t-p} \end{bmatrix}_{(p,t-p)} \end{bmatrix}$$

, where $\frac{\partial X_i(\theta)}{\partial \theta} = \lim_{\Delta\theta \to 0} X_i(\theta + \Delta\theta) - X_i(\theta - \Delta\theta)$ for each of the two hyperparameters.

Reference

Asimakopoulos, P., Asimakopoulos, S., Kourogenis, N. and Tsiritakis, E., 2017. Time-Disaggregated Dividend–Price Ratio and Dividend Growth Predictability in Large Equity Markets. *Journal of Financial and Quantitative Analysis*, *52*(5), pp.2305-2326.

Bacchiocchi, E., Bastianin, A., Missale, A. and Rossi, E., 2018. Structural analysis with mixed-frequency data: A MIDAS-SVAR model of US capital flows. *arXiv preprint arXiv:1802.00793*.

Baele, L., Bekaert, G. and Inghelbrecht, K., 2010. The determinants of stock and bond return comovements. *The Review of Financial Studies*, *23*(6), pp.2374-2428.

Carriero, A., Clark, T.E. and Marcellino, M., 2015. Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *178*(4), pp.837-862.

Colacito, R., Engle, R.F. and Ghysels, E., 2011. A component model for dynamic correlations. *Journal of Econometrics*, *164*(1), pp.45-59.

Davidson, R. and MacKinnon, J.G., 2004. *Econometric theory and methods* (Vol. 5). New York: Oxford University Press.

Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *74*(1), pp.33-43.

Engle, R.F., Ghysels, E. and Sohn, B., 2013. Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, *95*(3), pp.776-797.

Ghysels, E. and Marcellino, M., 2018. *Applied economic forecasting using time series methods*. Oxford University Press.

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

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

Online materials

Lee, Brandon, (2010), Nonlinear Least Squares - Applications to MIDAS and Probit Models, [Online], Available at https://ocw.mit.edu/courses/sloan-school-of-management/15-450-analytics-of-finance-fall-2010/, Accessed 10 August 2019.

Hansen, C. , Pereya, V., and Scherer, G. (2012), 02610 Optimization and Data Fitting { Nonlinear Least-Squares Problems [Online], Available at http://www.imm.dtu.dk/~pcha/LSDF/, Accessed 10 August 2019.