

Forecasting Singapore economic growth with mixed-frequency data

A.K. Tsui^a, C.Y. Xu^a and Z.Y. Zhang^b

^a Department of Economics, National University of Singapore, Singapore

^b School of Business, Edith Cowan University, Australia

Email: zhaoyong.zhang@ecu.edu.au

Abstract: In this paper we intend to forecast the economic growth of Singapore by employing mixed frequency data. This study is motivated by the following observations: macroeconomic variables are the important indicators of the economic performance, but they are normally available at low frequencies, e.g. quarterly for GDP and monthly for inflation. In contrast, the financial variables such as stock returns are available at high frequency, and often the asset prices are forward-looking and believed to contain useful information about future economic developments (Stock and Watson 2003). It is therefore an interesting question to raise whether or not one can use the high-frequency financial variables to better estimate and forecast the macroeconomic variables. Using mixed-frequency data in forecasting is clearly against the conventional forecasting models which generally require data with the same frequency. Time-aggregating, such as averaging, of the high frequency data is usually practiced to match the sampling rate of lower frequency data. But time-aggregation always leads to loss of individual timing information that might be important for forecasting. Hence, finding a suitable method to handle the high frequency data is a crucial task for every forecaster dealing with mixed frequency data.

We employ the Mixed Data Sampling (MIDAS) regression model introduced by Ghysels, et al. (2004). MIDAS regressions are essentially tightly parameterized, highly parsimonious regressions that deal with mixed frequency data. It is designed to find a balance between retaining the individual timing information of the high frequency data and reducing the number of parameters that need to be estimated. It is believed to have better estimating and forecasting ability than many other conventional models. A number of studies adopted MIDAS models to forecast quarterly series using monthly or daily financial data, mostly from the US (Anthony 2007; Clements and Galvão 2009).

Singapore is a small open economy, and vulnerable to the global economic conditions. Although its stock market is not comparable with that of the US in term of capitalization, the Singapore stock market performance is believed to be highly correlated with its real macroeconomic variable and contains important information for economic forecasting. In this paper, we forecast one-quarter-ahead Singapore GDP growth rate using Singapore stock market return sampled at various high frequencies. We investigate the forecasting performance from three models: a Mixed Data Sampling (MIDAS) regression model, a direct regression model on high frequency data and a time-averaging regression model. Our results show that MIDAS regression using high frequency stock return data produces better forecast of GDP growth rate than the other two models. Best forecasting performance is achieved using weekly stock return. The forecasting result is further improved by performing intra-period forecasting.

Keywords: *Mixed frequencies; Forecasting; Financial variable; Forecast evaluation*

1. INTRODUCTION

When forecasting macroeconomic variables, researchers often face a dilemma situation that data are not all sampled at the same frequency. Most macroeconomic data are sampled monthly (e.g., inflation, employment) or quarterly (e.g., GDP), while most financial variables (e.g., interest rates and asset prices), on the other hand, are sampled daily or even more frequently. The challenge is how to best use available data in our economic forecast. Forecasting macroeconomic variables is an important task for central banks, financial firms, and any other entity whose outcome depends on business cycle conditions. This study intends to forecast the economic growth of Singapore by employing mixed frequency data. This study is motivated by the following observations: macroeconomic variables are the important indicators of the economic performance, but they are normally available at low frequencies, for instance, quarterly for GDP and monthly for inflation. In contrast, the financial variables such as stock returns are available at high frequency, and often the asset prices are forward-looking and believed to contain useful information about future economic developments (Stock and Watson 2003). It is therefore an interesting question to raise whether or not one can use the high-frequency financial variables to better estimate and forecast the macroeconomic variables. Using mixed-frequency data in forecasting is expected to outperform the conventional forecasting models which generally require data with the same (low) frequency. However, the main challenge is how to effectively conduct the time-aggregating, such as averaging, of the high frequency data to match the sampling rate of lower frequency data, as time-aggregation always leads to loss of individual timing information that might be important for forecasting. Hence, finding a suitable method to handle the high frequency data is a crucial task for every forecaster dealing with mixed frequency data.

The Mixed Data Sampling (MIDAS) regression model was introduced by Ghysels, et al. (2004) and Armesto et al. (2010) provides a comprehensive review of some common methods for dealing with mixed-frequency data. It is essentially tightly parameterized, highly parsimonious regressions that deal with mixed frequency data. It is designed to find a balance between retaining the individual timing information of the high frequency data and reducing the number of parameters that need to be estimated. It is believed that the MIDAS models would have better estimating and forecasting ability than many other conventional models. Most of the existing studies focus on the US economy to forecast quarterly series using monthly or daily financial data (see Anthony 2007; Clements and Galvão 2009), but a few on the developing economies. In this study we employ the MIDAS regression model to forecast the quarterly GDP growth rate of the Singapore economy. Singapore is a small open economy, and an important financial centre in the world. Given its small open characteristic, the economy is very vulnerable to the global economic conditions. Although its stock market is not comparable with that of the US in term of capitalization, the Singapore stock market performance is believed to be highly correlated with its real macroeconomic variable and contains important information for economic forecasting. In this paper, we forecast one-quarter-ahead Singapore GDP growth rate using Singapore stock market return sampled at various high frequencies. We investigate the forecasting performance from three models: a Mixed Data Sampling (MIDAS) regression model, a direct regression model on high frequency data and a time-averaging regression model. Our results show that MIDAS regression using high frequency stock return data produces better forecast of GDP growth rate than the other two models. Best forecasting performance is achieved using weekly stock return. The forecasting result is further improved by performing intra-period forecasting.

The rest of this study is organized as follows. Section 2 discusses the models and the methodology used in this study. Section 3 analyzes the data sets, and discusses the empirical results. Section 4 concludes.

2. METHODOLOGY AND THE MODEL

In this section we discuss the four models used in this study. Following Sinko et al. (2010), let Y_t^Q be the quarterly series of the variable with t representing quarter. We assume that the high frequency series include daily series $X_{t,\tau}^D$, weekly series $X_{t,\tau}^W$, bi-weekly series $X_{t,\tau}^{BW}$, and monthly series $X_{t,\tau}^M$, where i denotes respectively the i^{th} day/week/bi-week/month in quarter t (with $i = 1$ be the first incidence of the quarter).

First, we consider a linear regression model running directly on the high frequency data without time aggregation. Given daily data $X_{t,\tau}^D$, a direct regression model using one quarter lagged data can be specified as:

$$Y_{t+1}^Q = \mu_0 + \mu_1 Y_t^Q + \beta_1 X_{1,t}^D + \dots + \beta_{N_D} X_{N_D,t}^D + u_{t+1} \quad (1)$$

However, this simple model is generally not suitable when the difference between the high and low sample frequencies is very large or when longer lagged data is included. For instance, if we model quarterly GDP growth rate with daily stock return, assuming there are 66 trading days in a quarter, then we have $N_D = 66$, and the number of coefficients to be estimated will be 68. This will lead to the potential problem of parameter proliferation (Ghysels *et al.*, 2004).

Then, we turn to the time-averaging regression model. Taking time-average is another easy way to deal with high frequency data. This approach is to take the average of the samples of the lower frequency variable to get the high frequency data, namely, $X_t^Q = (X_{N_D, t}^D + X_{N_D-1, t}^D + \dots + X_{1, t}^D)/N_D$. We then specify the time-averaging regression model in the ADL(1,1) form as:

$$Y_{t+1}^Q = \mu_0 + \mu_1 Y_t^Q + \beta X_t^Q + u_{t+1} \quad (2)$$

The above regression model implicitly adopted a flat weighting scheme of the high frequency data. Sinko *et al.* (2010) show that, by imposing flat weights on the high frequency, it may yield an omitted variable bias in the regression model, which has many econometric issues, such as pertaining to asymptotic inefficiencies at best or, as typical, asymptotic biases (Andreou *et al.*, 2010).

Further, a typical ADL-MIDAS (p_Y^Q, p_X^Q) regression model can be specified as follows:

$$Y_{t+1}^Q = \mu + \sum_{j=0}^{p_Y^Q-1} \mu_j Y_{t-j}^Q + \beta \sum_{j=0}^{p_X^Q-1} \sum_{i=0}^{N_D-1} \omega_{1+t+j \cdot N_D}(\theta^D) X_{N_D-i, t-j}^D + u_{t+1} \quad (3)$$

where (p_Y^Q, p_X^Q) are the number of quarterly lags of Y and X to be included in the regression, and $\omega_{1+t+j \cdot N_D}(\theta^D)$ is the weighting function depends on parameter vector θ^D . A typical normalized Beta polynomial weighting function is:

$$\omega_i = \omega_i(\theta_1, \theta_2) = \frac{x_i^{\theta_1-1} (1-x_i)^{\theta_2-1}}{\sum_{i=1}^N x_i^{\theta_1-1} (1-x_i)^{\theta_2-1}} \quad (4)$$

where $x_i = (i-1)/(N-1)$. The Beta function can be illustrated in **Error! Reference source not found.** It can be seen from Figure 1 that different values of parameters θ_1, θ_2 determine the weighting shapes, which can be a strictly decreasing or humped-shaped function by changing the parameters (Armesto *et al.*, 2010). It is also noted that the simple time-average can be obtained when $\theta_1 = \theta_2 = 1$. By using the variable weighting function, less restriction is imposed on the weights, and thus results in less estimation bias.

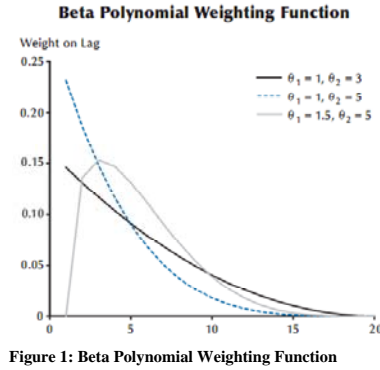


Figure 1: Beta Polynomial Weighting Function

In comparison with the direct regression and time-averaging regression models, the MIDAS model has the advantage of preserving the timing information of the high frequency data while remain parsimonious. In the above MIDAS setting (Equation 3), only 3 coefficients on quarterly lagged daily variables, i.e. $\beta, \theta_1, \theta_2$ need to be estimated, instead of 66 coefficients in the direct regression, and only 2 more compared with the case in the time-averaging model.

Finally, as regards the intra-period forecasting in the MIDAS Regression, the main advantage is to make use of the realized high frequency data before the actual realization of the low frequency data. By setting $p_Y^Q = p_X^Q = 1$, Equation (3) can then be extended to:

$$Y_{t+1}^Q = \mu + \mu_1 Y_t^Q + \beta^{Lead} \sum_{i=N_D-m}^{N_D-1} \omega_{1+i}(\theta^{D,Lead}) X_{N_D-i, t+1}^D + \beta^{Lag} \sum_{i=0}^{N_D-1} \omega_{1+i}(\theta^{D,Lag}) X_{N_D-i, t}^D + u_{t+1} \quad (5)$$

where m stands for the new observations of the high frequency data for the quarter which is to be forecasted. It is noted that, if we impose $\beta^{Lead} = \beta^{Lag}$ and $\theta^{D,Lead} = \theta^{D,Lag}$, the forecast will be no different between the current- and past-quarter data (Armesto *et al.*, 2010).

3. EMPIRICAL ANALYSIS

3.1 Data Description

The data sets used in this study were collected from the Singapore Department of Statistics and the Singapore Exchange. We collected seasonally adjusted quarterly GDP data at current market prices (which are available from 1975Q1) to calculate the quarterly GDP growth rate. The growth rate is not annualized. The Straits Times Index (STI) is regarded as the benchmark index for Singapore stock market, which is available from January 1988. The growth rate and returns of the STI for four different time periods, namely daily, weekly, bi-weekly and monthly, are calculated by the following formula:

$$r_{it} = \ln\left(\frac{P_{it}}{P_{it-1}}\right) * 100 \tag{6}$$

where r_{it} stands for the growth rate or return of the STI, P_{it} stands for GDP in year t or the close price of the STI at time t . Due to data availability issue, our sample period can only cover the period starting from 1988Q1 to 2010Q4. We have in total 92 observations for quarterly GDP growth rate. Figure 2 plots the quarterly GDP growth rate and weekly stock return ranging from 1988Q1 to 2010Q4. It can be seen from Figure 2 that the STI returns are no exception but highly volatile over time. Moreover, it is interesting to observe the strong trend of co-movements between the growth rate and the returns, especially during the economic recession periods such as in 1997, 2001 and 2008. It can also be observed that the STI returns led changes in the economic growth. This lends support to our hypothesis that stock returns are correlated with the growth and contains useful information for economic forecasts.

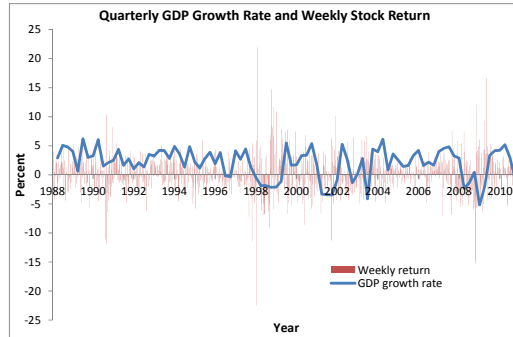


Figure 2: Quarterly GDP Growth Rate and Weekly Stock Return

3.2 Empirical Results

We have assessed the property of the macroeconomic variable, and the results confirm that the quarterly GDP growth rate is a first order autoregressive (AR(1)) process. Based on this information, we are more interested in comparing the forecasting performance between AR(1) model, direct regression model, time-averaging ADL(1,1) model and ADL-MIDAS(1,1) models. We perform the rolling window pseudo out-of-sample forecast. The first in-sample estimation is carried out by using a window from 1988Q1 to 2000Q4, and the first out-of-sample forecast starts from 2001Q1 until 2010Q4. A total of 40 quarterly forecasts over the past 10 years are used to calculate for the root mean square error (RMSE). There are two types of forecasting scheme adopted, namely, the end-of-period forecasting and intra-period forecasting. For the former, we use only available information up to the end of quarter $t-1$ to forecast Y_t^Q . For the latter, as time moves into the quarter to be forecasted and also new information about X_t^P will arrive, we include the newly available information to carry out intra-period forecasting. Due to space limitation, our discussions below will be focused on the forecasting ability of the different models.

Table 1 reports the RMSEs of the end-of-period forecasting using different models. RMSE for daily return cannot be calculated as the number of regressors (66) exceeds the number of available data set (52) within the window. AR(1) model is included as an benchmark model. As it can be seen from Table 1, the MIDAS model using weekly returns shows the best forecasting performance with the lowest RMSE of 2.3502, while the direct regression model performs the best with monthly return data. Given the fact that the quarterly GDP growth

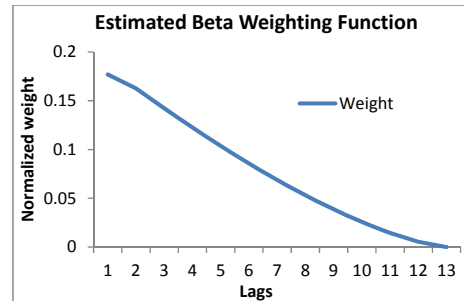


Figure 3: Normalized Estimated Beta Weighting Function

rate over the forecasted period has a mean of 1.539 and a standard deviation of 2.957, the forecasting error is still considered large. But there is a considerable decrease in the forecasting error by adopting MIDAS framework in comparison with other models. In addition, the estimation results of the MIDAS model using the whole sample period with weekly returns show that all the estimated coefficients are statistically significant at the 5% level, and the normalized beta weighting function in terms of weekly lags turns out to be a strictly decreasing function (see Figure 3). Unlike the case in the simple time-averaging regression where each lag is assigned with equal weight, Figure 3 shows that more weight is assigned to the nearer lags in the MIDAS model as information contained in the stock returns predicting GDP growth will die out within 13 weeks. Stock returns nearer to the end of the lagged quarter contain more relevant information relating to the GDP growth, and hence more weight should be assigned. The other models fail to capture this feature in the estimations. Finally, it is also noted that the MIDAS regression allows high frequency data to be aggregated while adopting a weighting function that can be specified using a few parameters. Its parsimonious formulation avoids the problem of parameter proliferation during estimation while imposing a weighting function on high frequency data with higher degree of freedom than a flat one. It balances the tradeoffs between making use of timing information and estimating less number of parameters. In this case, weekly return data seems to have the best noise attenuation while retaining substantial timing information to be captured by MIDAS regression.

Table 1: RMSEs for End-of-Period Forecasts, Horizon = 1 quarter

Models	Data sampling frequency				
	Daily	Weekly	Bi-weekly	Monthly	Quarterly
AR(1)	-	-	-	-	2.7910
Direct Regression	-	3.0163	2.6665	2.4360	2.5273
Time Averaging	2.4966	2.5349	2.5781	2.5139	-
MIDAS	2.4767	2.3502	2.5233	2.4403	-

Based on the RMSEs, we choose the best fit models with different frequency data and conduct the forecasting exercises for the Singapore economy. Figure 4 presents the actual and forecasted GDP growth rate over the period from 2001Q1 to 2010Q4 by using the better performing models with different data frequencies. As it can be seen from Figure 4, in general the better performing models forecast well out-of-sample. The forecasted quarterly GDP growth close follow the up and down of the actual growth rate, especially during the period when the fluctuation is very large. Comparatively one may note that the MIDAS model predicts well the turning points over time.

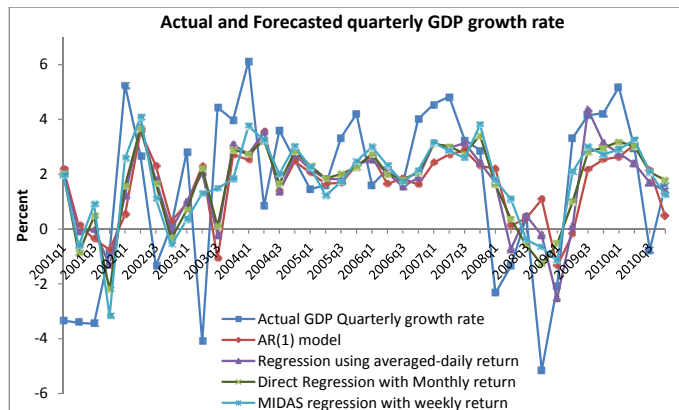


Figure 4: Actual and Forecasted quarterly GDP growth rate

In contrast, as regards the intra-period forecasting, we include high frequency stock returns data realized before the end of current quarter. As the progression of forecasting is carried out in a weekly manner, at certain time during the quarter, regression with bi-weekly and monthly data is not updated since by then new data is still not available. This is also reflected in the RMSEs associated with the number of weeks rolling into the quarter for forecast. Due to space limitation, we report in Figure 5 the RMSE results of MIDAS intra-period forecasting only. It is interesting to note that the RMSEs decrease first as intra-period data is included in the regression, and then increase when approaching to the end of the quarter. The results seem suggest that the intra-period MIDAS model improve the forecasting performance in the first instance as more information becomes available, but the forecasting performance deteriorate when more information arrives in the end of the period. This seems to be against intuitive observation. This is possible due to the restriction we imposed in Equation (5), namely $\beta^{Lead} = \beta^{Lag}$ and $\theta^{D,Lead} = \theta^{D,Lag}$. With this restriction, the current data is actually grouped together with the past quarter data during the estimations, and both share the same weighting function. As we know, the stock market is very sensitive to the economic business cycles, but the real economic indicators like GDP will experience a delay in the actual adjustment. When one time

lagged data is grouped with one quarter lagged data, the former actually adds “noise” to the regressors, which could not be corrected by the estimation of beta function. As a consequence, when near-end-of-quarter stock return data are included for intra-period forecasting, the results deteriorate with high RMSEs. Our results show that the intra-period MIDAS forecasting model performs well with the inclusion of weekly intra-period data up to the end of 5th week with the lowest RMSE.

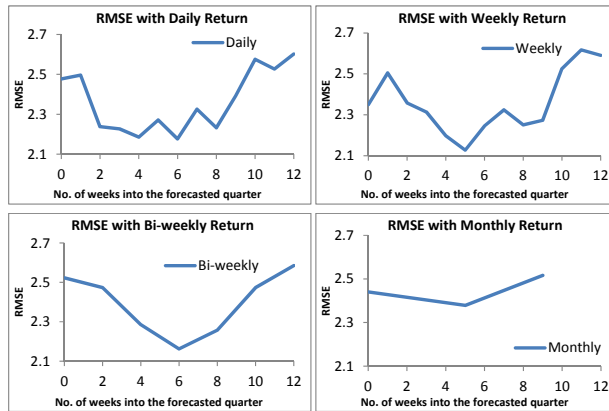


Figure 5: RMSEs for Intra-Period Forecasts

Figure 6 presents the actual and forecasted GDP growth rates using the intra-period MIDAS forecasting model with monthly and up to the 5th week returns. For comparison purpose, we also include in Figure 6 the forecasted GDP using AR(1) model and MIDAS end-of-period forecasting model with weekly returns. As it can be seen from Figure 6, with the inclusion of intra-period data the model shows an improved prediction power, especially the economic crises during the 2001-2002 and 2008-2009 periods. It is also noted that the intra-period MIDAS regression model is the only one that almost fully captured the impact of the global financial crisis (GFC) on the Singapore economy. The results indicate that the intra-period MIDAS model outperforms other forecasting models, as a good forecasting model needs to capture all these important up-and-downs in the economy but a bad forecasting model would just predict minor fluctuations over the entire period (Anthony 2007).

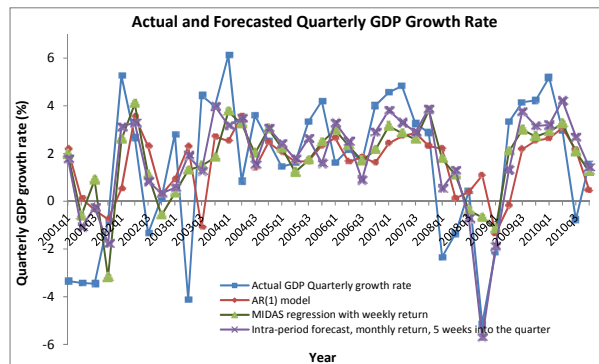


Figure 6: Actual and Forecasted Quarterly GDP Growth Rate with Intra-Period Forecasting

Furthermore, we have conducted the Diebold-Mariano-West test to justify the forecasting performance of these models, and Table 2 reports the DMW test statistics. As the DMW test statistic has asymptotic standard normal distribution, it can be interpreted as standard t-statistic. The models in the column are tested against the models in the row. The negative sign of the test statistics indicates smaller squared forecasting errors of the model in column, i.e. better forecasting performance. As it can be seen in Table 2, although the end-of-period MIDAS model performs better than other models in terms of low RMSE, the null hypothesis could not be rejected based on the DMW test. In other words, the reduction in squared errors for the MIDAS model in comparison with TA-D and DR-M is not statistically significant at the 5% level. In contrast, the intra-period forecasting model does outperform the direct regression and time-averaging model at a 5% significance level. It is also considered better than the MIDAS model if a 10% significance level (-1.28) is adopted.

Table 2: DMW Test Statistics

Models	TA-D	DR-M	MIDAS-W
MIDAS-W vs.	-1.2484	-0.77313	-
MIDAS-W5 vs.	-2.3133	-2.3966	-1.4589

Note: TA-D stands for the time averaging regression with averaged-daily stock returns; DR-M is the direct regression with monthly stock return; MIDAS-W refers to the MIDAS regression with weekly stock returns; and MIDAS-W5 stands for the MIDAS model including weekly stock returns up to the 5th weeks. In this test, we compare the test statistic with one-sided 5% critical value of -1.645, namely, H_0 is rejected for $S < -1.645$.

4. CONCLUDING REMARKS

When forecasting macroeconomic variables, researchers often face a dilemma situation that data are not all sampled at the same frequency, namely some are sampled monthly such as inflation and employment) or quarterly such as GDP, while most financial variables are sampled daily or even more frequently. The challenge is how to best use available data in our economic forecast. In this paper, we have comparatively investigated the forecasting performance of the three models, i.e., the MIDAS regression model, the direct regression model on high frequency data and the time-averaging regression model, by using data from the Singapore economy. Our results show that MIDAS regression using high frequency stock returns data produces better forecast of GDP growth rate than the other models, and the best forecasting performance is achieved using weekly stock returns. It is found that the intra-period MIDAS model outperforms other forecasting models, as it can capture well all the important up-and-downs of the economic performance in Singapore, especially during the economic crises in 2001-2002 and 2008-2009 with the lowest RMRE value. The DMW test statistics further confirm that the intra-period forecasting model does outperform the direct regression and time-averaging model at a 5% significance level, and is also considered better than the MIDAS model at a 10% significance level.

REFERENCES

- Alper, C.E., S. Fendoglu and B. Saltoglu (2012), "MIDAS volatility forecast performance under market stress: Evidence from emerging stock markets", *Economics Letters*, 117 (2): 528-532
- Andreou E., E. Ghysels and A. Kourtellis (2010), "Regression models with mixed sampling frequencies." *Journal of Econometrics* 158(2): 246-261.
- Anthony, S. T. (2007). Financial Variables as Predictors of Real Output Growth, East Asian Bureau of Economic Research.
- Armesto, M. T., K. M. Engemann, and M.T. Owyang (2010). "Forecasting with Mixed Frequencies." *Federal Reserve Bank of St Louis Review* 92(6): 521-536.
- 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(7): 1187-1206.
- Diebold, F. X. and R. S. Mariano (1995). "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13(3): 253-263.
- Ghysels, E., P. Santa-Clara, et al. (2004). The MIDAS Touch: Mixed Data Sampling Regression Models, Anderson Graduate School of Management, UCLA.
- Ghysels, E., A. Sinko, et al. (2007). "MIDAS regressions: Further results and new directions." *Econometric Reviews* 26(1): 53-90.
- Sinko, S.A., M. Sockin and E. Ghysels (2010). "Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regression Models." Unpublished manuscript.
- Stock, J. H. and M. W. Watson (2003). "Forecasting Output and Inflation: The Role of Asset Prices." *Journal of Economic Literature* 41(3): 788-829.
- West, K. D. (1996). "Asymptotic Inference about Predictive Ability." *Econometrica* 64(5): 1067-1084.