SENIOR RESEARCH

Forecasting private consumption and investment using financial data and mixed-data sampling (MIDAS) models

NAME: Julawat Paoin

ID: 5645823629

Advisor: Supachoke Thawornkaiwong,Ph.D.

Date: 4th May 2017

Senior Research Submitted in Partial Fulfillment of the Requirements

for the Bachelor of Arts degree in Economics (International Program)

The Bachelor of Arts Program in Economics

Faculty of Economics

Chulalongkorn University

Academic year 2016

Approve

____

(Assoc.Prof.Sothithorn Mallikamas,Ph.D.)

Chairman

Data of Approval____________________________

Abstract

This paper analyzed forecasting performances of the Autoregressive Mixed-Data Sampling model (AR-MIDAS) with the use of financial real time data on Consumption and Investment expenditures components of Thailand gross domestic product. An AR-MIDAS model is a regression model that allows dependent variables and independent variables to be in different frequencies. The model makes great use of real time data with high frequencies despite the facts that macroeconomic variables are collected in low frequencies. By comparing it to the alternative time series models which are Autoregressive model and Autoregressive Distributed Lags model, the results showed that AR-MIDAS model outperforms other models almost all of the time for all independent variables that were selected. It is clear that financial data actually can be used to help forecast macroeconomic variables. In addition, choosing independent variables are very important to the performance of the models.

Table of contents **Page**

1.) Introduction

Forecasting macroeconomic variables are very important for policy makers, firms and institution. Under the assumption that financial data are forward looking and any information available at the time are already incorporated to the values of those assets, using financial indicators in the regression models should help forecasting macroeconomic indicators. However, macroeconomic variables are sampled in low frequency, while financial data are sampled in higher frequency. Different frequencies of data makes forecasting complicated because regression models typically use variables that have the same frequency only. Therefore, within-period financial data cannot be used to help forecast until the data are available at the end of the dependent variable period.

Ghysels, Santa-Clara, and Valkanov (2004) introduced the mixed-data sampling (MIDAS) model to deal with this problem. MIDAS models allow variables to be sampled at different frequencies. Unlike other models, MIDAS models make use of within-period of higher frequency data.

Many previous studies confirmed that MIDAS model outperforms other alternative models. For example, Andreou, Ghysels, and Kourtellos (2013) showed that MIDAS model is able to incorporate information from forward-looking nature of financial daily data and used it for forecasting the quarterly economic growth. MIDAS model combined daily information in an effective way and outperformed traditional models. Similarly, Clements, and Galvão (2008) found that MIDAS model makes use of monthly data better than other models, and performed even better with within quarter data. However, Armesto, Engemann, and Owyang (2010) had different results. MIDAS model do not have big advantage over other models. Each model varied.

In this paper, MIDAS models were being evaluated and compared with alternative models based on the information of previous studies. This paper focused on forecasting consumption and investment components of Thailand gross domestic product (GDP) separated by types of expenditure between the first quarter of 1996 to the fourth quarter of 2016 only while financial data are monthly data between the first month of 1996 and the last month of 2016.

2.) Regression Models

There are three regression models used in this paper which are Autoregressive (AR) model, Autoregressive Distributed Lag Model (ARDL) model, and Autoregressive Mixed Data Sampling (AR-MIDAS) model.

Autoregressive (AR) model

Autoregressive (AR) models were being used as a benchmark model for this paper. An AR model is a time series model that includes lagged values of independent variable (y) in the regression model. The general form of an AR model is:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon_t \quad \ldots (1)$$

where μ is the intercept coefficient, p is the number of lags of y, γ is the estimated coefficient, t indexes the time unit, and ε is an error term.

Autoregressive distributed lag (ARDL) model

Distributed lag models are time series models that include lagged values of independent variables (x). An Autoregressive Distributed Lag (ARDL) model is one of the distributed lag models. The ARDL models are used in this paper because the model not only include the lagged values of independent variable (x) but also include lagged values of dependent variables (y) just like the AR models. The ARDL model can be specified as:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{j=0}^{r} \beta_j x_{t-j} + \varepsilon_t \quad \ldots (2)$$

This equation is achieved by include independent variables part $\sum_{j=0}^{r} \beta_j x_{t-j}$ in the equation (1) where r is the number of lags of independent variable. Therefore, the ARDL models use other variables to help explained dependent variables rather than using only its past values.

Time aggregation scheme

As financial data in this paper are all monthly frequency while dependent variables which are consumption and investment are quarterly data. It is necessary to convert monthly financial data to quarterly data in order to estimate ARDL models. The time aggregation scheme that are used in this paper is simple average scheme. By averaging three months data, quarterly data will be achieved as in the equation below.

$$x_t^Q = (x_{1,t}^M + x_{2,t}^M + x_{3,t}^M)/3$$

where x_t^Q is the quarterly data, while $x_{1,t}^M$, $x_{2,t}^M$, $x_{3,t}^M$ are the first month, the second month and the third month data of the quarter at time t respectively.

Autoregressive Mixed data sampling (AR-MIDAS) model

The MIDAS models of Ghysels, Santa-Clara, and Valkanov (2004) are closely related to distributed lag models. However, the model allows independent variables (x) to be sampled in different frequency from the dependent variable (y). In this paper, An AR-MIDAS model is used instead of a MIDAS model because it includes autoregressive term. The AR-MIDAS model with one independent variable is specified as;

$$y_t = \beta_0 + \sum_{i=1}^{p} \gamma_i y_{t-i} + \beta_1 \sum_{i=1}^{K} w(i;\theta)\, x_{t-(i-1)/3}^{(m)} + \varepsilon_t \quad \dots (3)$$

where K is the number of lags of independent variable, t indexes the basic time unit (lower frequency), $w(i;\theta)$ is the weight function, and m is the number of higher sampling frequency in the lower sampling frequency. As the higher frequency in this paper is monthly while the lower one is quarterly, so m = 3. Therefore, equation (3) can be written as:

$$y_t = \beta_0 + \sum_{i=0}^{p} \gamma_i y_{t-i} + \beta_1 [w(1;\theta) x_t^{(3)} + w(2;\theta) x_{t-\frac{1}{3}}^{(3)}$$

$$+ w(3;\theta) x_{t-\frac{2}{3}}^{(3)} + \cdots + w(K;\theta) x_{t-\frac{K-1}{3}}^{(3)}] + \varepsilon_t$$

The differences between equation (2) and (3) are that independent variables are in higher frequency than dependent variables and that coefficient are in the weight function form.

Weighting scheme

The weight function $w(i;\theta)$ has to be chosen in order to estimate AR-MIDAS models. The reason for using weight function for independent variables is because it reduces the number of parameters that are needed to be estimated. The main weight function that will be used in this paper is Exponential Almon polynomial function.

$$w(i;\theta) = \frac{\exp(\theta_1 i + \theta_2 i^2)}{\sum_{i=1}^{r} \exp(\theta_1 i + \theta_2 i^2)}$$

This function reduces the number of estimated parameters. The function also has high flexibility, time series samples can have a decay memory form which recent pasts have higher weight than more distant pasts.

Forecast horizon

Forecast horizon is the number of periods ahead from current period that will be forecasted in a certain model. All of the information available before the current period is being used. In this paper, models are selected for each forecast horizon. As AR models and ARDL models have the same frequency for each variable, forecast horizon will be in a whole unit. For instance, h=1 represents one quarter ahead forecast horizon, h=2 represents two quarter ahead forecast horizon.

However, in this paper, dependent variables in AR-MIDAS models are collected in quarterly period while independent variables are collected in monthly period. If there are independent variables available on the first two months (t=2/3) of the current quarter are available, forecasting horizon for the next quarter (t = 1) and the next two quarters (t = 2) will be h = 1/3 and h = 4/3 respectively. If the information are available only on the first month of the quarter (t=1/3), forecasting horizon for the next quarter and next two quarters will be h = 2/3 and h = 5/3 respectively.

Forecast horizon h=0 is possible for ARDL and AR-MIDAS models because independent variables which are financial data are available in real time.

3.) Data

Variables Selection

Consumption

There are some existing works related to variables that have correlation with private consumption. Some papers have found correlation between consumer spending and consumer confidence index. Ludvigson (2004) found that consumer confidence can be used to forecast quarterly growth of consumption expenditure because it has high predicting ability. Croushore (2005) found that there is a correlation between consumer confidence and consumer spending but he did not prove that consumer confidence has any predictive role. The correlation can be drove by other macroeconomic variables. On the other hand, Batchelor, and Dua (1998) found that consumer confidence improved forecast only in the recession period, it should not be considered in other period of time.

There are also many previous studies that found positive correlation between stock prices and consumer spending. Poterba (2000) suggested that the correlation is an effect of increase in stock market wealth. However, he also stated that stock prices affect household consumption because of consumers' perception and consumer confidence even though they do not own any stock. Similarly, Dynan and Maki (2001) stated that there are two reasons why changes in stock prices affect consumption which they found that changes in wealth affect consumption more than changes in future income. Ludwig and Siek (2004) also found positive relationship between stock prices and consumption. However, they found that changes in stock prices affect consumption for market-based financial system countries more than countries with a bank-based financial system.

Therefore, consumer confidence index and stock market indexes were used as independent variables for consumption regression model.

Investment

In term of investment component, there are several studies relating stock markets and investment. Barro (1990) stated that stock market price has higher forecasting ability for investment. It is much better than Tobin's q. In contrast, Harvey (1989) found that stock market index explained only five percent of economic growth while bond markets could explain more than thirty percent.

Independent variables for investment regression that were used in this paper are stock market indexes

Table1: Data: variable names, data frequencies and sources of data

Name	**Variable**	**Measurement**	**Source**
Dependent variables			
Consumption	Private Consumption Expenditures	**Quarterly**	National Economic and Social Development Board
Investment	Gross Fixed Capital Formation, Change in inventories		
(Chained Volume Measures, reference year = 2002)			
Independent variables			
CCI	Thailand Consumer Confidence Index	**Monthly**	University of the Thai Chamber of Commerce
SETcons	SET Fashion, SET Home, SET Person Indexes		Stock Exchange of Thailand
SETindex	SET Index		
SET50	SET50 Index		

Unit root test

In order to use time series models, time series samples should be stationary, any non-stationary time series should be differenced until it becomes. In this paper, Augmented Dicky-Fuller (ADF) tests, Phillips-Perron (PP) Unit Root tests and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests are used to see whether a certain time series is needed to take difference or not.

ADF tests and PP tests test the null hypothesis that a unit root is present in a certain time series sample. The alternative hypothesis is that the time series sample is stationary. In contrast, KPSS tests test the null hypothesis of stationary process, while the alternative is a presence of a unit root.

Table2: p-values of ADF test, PP test and KPSS test for each variable

Variables	**Augmented Dickey-Fuller test (Ha: Stationary)**	**Phillips-Perron Unit Root test (Ha: Stationary)**	**KPSS test (Ha: presence of a unit root)**
log(Consumption)	p-value = 0.1466	p-value = 0.5637	p-value = 0.01
d/dt(log(Consumption))	**p-value = 0.01609**	**p-value = 0.01**	**p-value = 0.1**
log(Investment)	p-value = 0.3107	p-value = 0.8268	p-value = 0.01
d/dt(log(Investment))	p-value = 0.1192	**p-value = 0.01**	**p-value = 0.1**
log(CCIq*)	p-value = 0.5406	**p-value = 0.03612**	p-value = 0.04949
d/dt(log(CCIq))	**p-value = 0.01**	**p-value = 0.01**	**p-value = 0.09312**
log(SETindexq*)	p-value = 0.5397	p-value = 0.8212	p-value = 0.01
d/dt(log(SETindexq))	**p-value = 0.01**	**p-value = 0.01**	**p-value = 0.1**
log(SETconsq*)	**p-value = 0.01756**	p-value = 0.5582	p-value = 0.01
d/dt(log(SETconsq))	p-value = 0.0719	**p-value = 0.01**	**p-value = 0.08231**
log(SET50q*)	**p-value = 0.01371**	p-value = 0.4953	p-value = 0.01
d/dt(log(SET50q))	**p-value = 0.02959**	**p-value = 0.01**	**p-value = 0.1**
log(CCI)	**p-value = 0.02272**	p-value = 0.05658	p-value = 0.04039
d/dt(log(CCI))	**p-value = 0.01**	**p-value = 0.01**	p-value = 0.04491
log(SETindex)	p-value = 0.7156	p-value = 0.8121	p-value = 0.01
d/dt(log(SETindex))	**p-value = 0.01**	**p-value = 0.01**	**p-value = 0.1**
log(SETcons)	p-value = 0.07642	p-value = 0.6542	p-value = 0.01
d/dt(log(SETcons))	**p-value = 0.01**	**p-value = 0.01**	**p-value = 0.09317**
log(SET50)	**p-value = 0.02382**	p-value = 0.5096	p-value = 0.01
d/dt(log(SET50))	**p-value = 0.01**	**p-value = 0.01**	**p-value = 0.1**

*CCIq, SETindexq, SETconsq and SET50q are quarterly data that have been converted from monthly data to quarterly data by using simple average scheme mentioned before.

It is clear that every variable are stationary after the first difference. Therefore, the first difference of each variable are used to estimate regression models.

4.) Methodology

Regressions

For both consumption and investment, regression models for each of the three time series models were choose. AR models are selected using Akaike Information Criterion (AIC). ARDL models for each independent variable with minimum AIC are selected in three forecast horizons which are current period (h=0), one period ahead (h=1) and two periods ahead (h=2). Lastly, AR-MIDAS models for each independent variable with minimum AIC are selected in seven forecast horizons which are present period (h=0), one month ahead (h=1/3), two months ahead (h=2/3), one quarter ahead (h=1), four months ahead (h=4/3), five months ahead (h=5/3) and two quarters ahead (h=2)

<u>Consumption</u>

Data that are used to estimate each model are from Jan, 1999 to Dec, 2012. For ARDL and AR-MIDAS models, Consumer Confidence index, SET index and SET Consumption index are used as independent variables.

Table3: Consumption regression models: variables used for each independent variable

Models	Independent variables		
	CCI	**SET index**	**SET consumption**
AR	y = d/dt(log(Consumption))		
ARDL	y = d/dt(log(Consumption))		
	x = d/dt(log(CCIq))	x = d/dt(log(SETindexq))	x = d/dt(log(SETconsq))
AR-MIDAS	y = d/dt(log(Consumption))		
	x = d/dt(log(CCI))	x = d/dt(log(SETindex))	x = d/dt(log(SETcons))

Investment

Data that are used to estimate each model are from Jan, 1996 to Dec, 2012. For ARDL and AR-MIDAS models, SET index and SET 50 index are used as independent variables.

Table4: Investment regression models: variables used for each independent variable

Models	Independent variables: SET index	Independent variables: SET50 index
AR model	y = d/dt(log(Investment))	
ARDL	y = d/dt(log(Investment))	
	x = d/dt(log(SETindexq))	x = d/dt(log(SET50q))
AR-MIDAS	y = d/dt(log(Investment))	
	x = d/dt(log(SETindex))	x = d/dt(log(SET50))

Diagnostic check

Before using any model to forecast, it is necessary to do a diagnostic check. The residuals of the model should not be correlated. The tests that are used in this paper for diagnostic check are Box-Pierce tests which the null hypothesis is that the data are purely random, while the alternative hypothesis is that the data have serial correlation. Therefore, the tests are used on the residuals of each selected models to see whether the residuals are purely random or not. The Table5 and Table6 show the p-values of the tests of each selected model. Each p-value should be bigger than 0.05 in order to use the model to forecast.

Out-of-sample forecasts

The data will be divided into two parts; in-sample and out-sample. In-sample data are used to estimate regression models, for Consumption, in-sample periods are from Jan, 1999 to Dec, 2012, while for Investment, in-sample periods are from Jan, 1996 to Dec, 2012. Out-sample data are the data used as actual values to be compared with forecast values of each estimated model from in-sample periods. The differences between actual values and forecast values are forecast errors. Out-sample periods for both Consumption and Investment are from the first quarter of 2013 to the last quarter of 2016 which are 16 periods. Within out-sample periods, root mean squared forecast errors (RMSFEs) are calculated in order to compare between the three models, forecast horizons and independent variables.

As the AR models are used as benchmark models, in order to compare the ARDL models forecasting performance with the AR models, the ratios between RMSFEs of the ARDL and the AR models are computed. The RMSFEs of the ARDL models with forecast horizons equal to 0 and 1 are divided by the RMSFEs of the AR models with forecast horizon equal to 1, while the RMSFEs of the ARDL models with h=2 are divided by the RMSFEs of the AR models with h=2. In order to compare AR-MIDAS with AR models, the RMSFEs of the AR-MIDAS models with forecast horizons equal to 0, 1/3, 2/3 and 1 are divided by the RMSFEs of the AR models with forecast horizon equal to 1. The RMSFEs of the AR-MIDAS with h=4/3, 5/3 and 2 are divided by the RMSFEs of the AR models with h=2.

Table5: p-values of the Box-Pierce tests for the residuals
of each selected Consumption regression models

Forecast Horizon	**AR**	**CCI**		**SET index**		**SET consumption**	
		ARDL	**MIDAS**	**ARDL**	**MIDAS**	**ARDL**	**MIDAS**
h=0		p-value = 0.66	p-value = 0.452	p-value = 0.1658	p-value = 0.3597	p-value = 0.2027	p-value = 0.5145
h=1/3			p-value = 0.4711		p-value = 0.291		p-value = 0.4309
h=2/3			p-value = 0.6069		p-value = 0.3803		p-value = 0.1268
h=1	p-value = 0.7094	p-value = 0.726	p-value = 0.7827	p-value = 0.1948	p-value = 0.2316	p-value = 0.1918	p-value = 0.13
h=4/3			p-value = 0.739		p-value = 0.3239		p-value = 0.3123
h=5/3			p-value = 0.6577		p-value = 0.1527		p-value = 0.1348
h=2	p-value = 0.4844	p-value = 0.5145	p-value = 0.3528	p-value = 0.1953	p-value = 0.1602	p-value = 0.3667	p-value = 0.0944

Table6: p-values of the Box-Pierce tests for the residuals

of each selected Investment regression models

Forecast Horizon	**AR**	**SET index**		**SET50 index**	
		ARDL	**MIDAS**	**ARDL**	**MIDAS**
h=0		p-value = 0.5563	p-value = 0.5822	p-value = 0.5062	p-value = 0.5389
h=1/3			p-value = 0.5145		p-value = 0.4294
h=2/3			p-value = 0.4747		p-value = 0.3692
h=1	p-value = 0.7703	p-value = 0.4095	p-value = 0.4668	p-value = 0.38	p-value = 0.3792
h=4/3			p-value = 0.2327		p-value = 0.212
h=5/3			p-value = 0.3364		p-value = 0.3024
h=2	p-value = 0.7065	p-value = 0.6287	p-value = 0.8827	p-value = 0.591	p-value = 0.8589

5.) Results

Table7 and Table8 below show the ratio of RMSFEs of each model to the RMSFEs of benchmark AR model. Therefore, if the ratio is less than 1, it means that the model outperforms its alternative AR model.

Consumption

Table7: The forecast performance of ARDL and AR-MIDAS models compared to AR models for Consumption

Forecast Horizon	CCI		SET index		SET consumption	
	ARDL	**AR-MIDAS**	**ARDL**	**AR-MIDAS**	**ARDL**	**AR-MIDAS**
	Ratio to AR					
h=0	1.035302	1.000581	1.050737	1.026999	1.039864	1.017324
h=1/3		1.012685		**0.998781**		**0.99109**
h=2/3		**0.980667**		**0.982854**		**0.98063**
h=1	**0.994901**	**0.980505**	1.064003	1.040143	1.060734	**0.980443**
h=4/3		**0.895857**		**0.888637**		**0.874545**
h=5/3		**0.89965**		**0.939741**		**0.95846**
h=2	**0.936462**	**0.9259**	**0.941669**	**0.933714**	**0.947237**	**0.977243**

For the first independent variable; CCI, ARDL models slightly outperform AR models when the forecast horizons equal to 1 and 2. Similarly, AR-MIDAS models outperform AR models when the forecast horizons are higher than 2 months. In addition, AR-MIDAS also outperform ARDL in all forecast horizons. Therefore, CCI has high predictive ability for forecast horizon more than 2 months as ARDL and AR-MIDAS outperform AR models which use only Consumption itself to explain. It can be viewed that Consumer Confidence Index is the consumers' perception about future and it is more related to their consumption behavior in the future period.

For SET index, ARDL models outperform AR models only when h=2, while AR-MIDAS still perform very well, only when h=0, and h=1 that the model perform worse than AR. Within-period SET index data actually help improve forecast performance. However, if compare SET index with CCI, CCI still perform better most of the time.

For SET consumption index, ARDL models have the same results as for SET index, only when h=2 that ARDL outperform AR models. AR-MIDAS models also have similar results, the model only lose to AR when h=0. In addition, If compare SET consumption index to the other variables, the index is slightly better than the other two for more than half of the forecast horizons, h=1/3, 2/3, 1 and 4/3. Therefore, it is clear that SET consumption is more related to Consumption components of GDP than other variables.

Investment

Table8: The forecast performance of ARDL and AR-MIDAS models compared to AR models for Investment

Forecast Horizon	SET index		SET50 index	
	ARDL	AR-MIDAS	ARDL	AR-MIDAS
	Ratio to AR			
h=0	1.079558	**0.9785**	1.079558	**0.978759**
h=1/3		**0.982022**		**0.986078**
h=2/3		**0.997162**		1.004187
h=1	1.09498	1.008651	1.086476	1.005359
h=4/3		**0.861111**		**0.857338**
h=5/3		**0.862125**		**0.849108**
h=2	**0.99079**	**0.903861**	**0.967765**	**0.892103**

For SET index, ARDL models only slightly outperform AR models when h=2. On the other hand, AR-MIDAS outperform AR models in almost all forecast horizons except for h=1. It can be seen that SET index can be used to help improve forecast Investment.

For SET50 index, ARDL model also outperform AR models only when h=2 just like for SET index. However, AR-MIDAS models are a little bit worse than those with SET index. Surprisingly, SET index improve forecast performance more than SET50 index. SET index can used to represent Investment activities more than SET50 index which concentrate and represent only 50 companies listed in the stock exchange.

6.) Seasonal Adjusted Data

As Consumption and Investment variables which were used in the previous sections both exhibit seasonal pattern throughout the samples and so far the data that have been used are non-seasonal adjusted data only, this part of the paper is focus on whether non-seasonal adjusted (NSA) data or seasonal adjusted (SA) data are better in term of forecasting performance.

The evaluation method is the same which is by comparison between root mean squared forecast errors (RMSFEs). The RMSFEs of SA data are divided by the RMSFEs of NSA data for all three models, AR, ARDL and AR-MIDAS models. However, the independent variables for ARDL and AR-MIDAS models are only the best performing one of each dependent variable on the previous results, so for Consumption, the independent variable is SET consumption index, and for Investment, the independent variable is SET index.

Table8: The ratios between seasonal adjusted and non-seasonal adjusted root mean square forecast errors.

	Consumption SA			**Investment SA**		
		SET consumption			**SET index**	
Forecast	**AR**	**ARDL**	**AR-MIDAS**	**AR**	**ARDL**	**AR-MIDAS**
Horizon	**Ratio to Consumption NSA**			**Ratio to Investment NSA**		
h=0		**0.569995**	**0.542519**		1.018514	**0.999285**
h=1/3			**0.593997**			1.102713
h=2/3			**0.616645**			1.093548
h=1	**0.606078**	**0.573997**	**0.619834**	1.045229	1.025826	1.082534
h=4/3			**0.613668**			1.022859
h=5/3			**0.562745**			1.044322
h=2	**0.573492**	**0.581284**	**0.530906**	**0.914399**	**0.994563**	1.030302

For Consumption, seasonal adjusted data are much more easier to forecast as the root mean square forecast errors ratio are about 0.6 for every models and horizons. All of AR, ARDL and AR-MIDAS models have much higher predictive power.

For Investment, in contrast to the results for Consumption, seasonal adjusted data do not have advantage over non-seasonal adjusted data. Almost all of the models that used the NSA data outperform the models that used the SA data. Only the AR model when h=2, the ARDL model when h=2 and the AR-MIDAS model when h=0 outperform their alternative models using the NSA data.

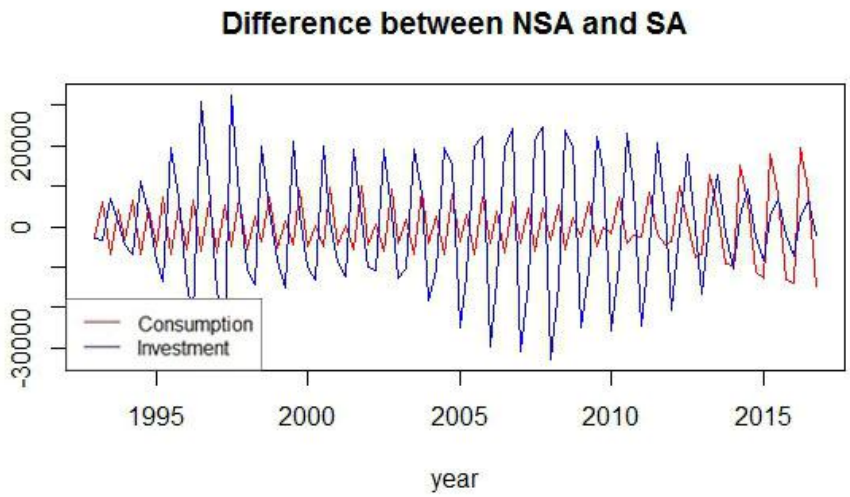

Figure1: Difference between non-seasonal adjusted data and seasonal adjusted data

The figure above shows the difference between the non-seasonal adjusted data and the seasonal adjusted data of both Consumption and Investment. It can be seen that the difference between the NSA and the SA of Investment data is much higher than that of Consumption. It means that the Investment data got seasonal adjust in higher manner than the Consumption data. However, the results show that the SA data of Investment did not improve forecast performance. Therefore, it is inconclusive whether seasonal adjusted data should be used to forecast or not.

7.) Conclusion

AR-MIDAS models outperform both alternative ARDL models and benchmark AR models in almost all forecast horizons for all independent variables. By outperforming ARDL models, it can be seen that AR-MIDAS models make great use of real time data. On the other hand, time aggregation schemes of ARDL models are not effective enough.

Financial data can be used to help forecast macroeconomic variables as it helps improve forecast performance. However, the choices of variables are important. These models will perform even better if the independent variables are more closely related to dependent variables. The best explanatory for Consumption in this paper is the SET consumption index, while the best one for Investment is the SET index.

Lastly, it is not possible to conclude whether seasonal adjusted data or non-seasonal adjusted data are more suitable for forecasting. Seasonal adjusted data of Consumption have a big advantage over non-seasonal adjusted data. On the other hand, models with non-seasonal adjusted data of Investment perform better.

References

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.

Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, *92*(6), 521-36.

Batchelor, R., & Dua, P. (1998). Improving macro-economic forecasts: The role of consumer confidence. *International Journal of Forecasting*, *14*(1), 71-81.

Barro, R. J. (1990). The stock market and investment. *Review of Financial Studies*, *3*(1), 115-131.

Clements, M. P., & Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics*, *26*(4), 546-554.

Croushore, D. (2005). Do consumer-confidence indexes help forecast consumer spending in real time?. *The North American Journal of Economics and Finance*, *16*(3), 435-450.

Dynan, K. E., & Maki, D. M. (2001). Does stock market wealth matter for consumption?.

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

Harvey, C. R. (1989). Forecasts of economic growth from the bond and stock markets. *Financial Analysts Journal*, 38-45.

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *The Journal of Economic Perspectives*, *18*(2), 29-50.

Ludwig, A., & Siek, T. (2004). The Relationship between Stock Prices, House Prices and Consumption in OECD Countries. *MEA*.

Poterba, J. M. (2000). Stock Market Wealth and Consumption. *The Journal of Economic Perspectives*, *14*(2), 99-118.