MODELING SOUTH KOREA'S ECONOMY:
Model Comparison of Simple ARMA (1,1) vs. High Frequency Principal Component Analysis in Forecasting South Korea's GDP

Macroeconomics

You Chan Kevin Chung
University of Pennsylvania
Honors Thesis

Thesis Advisors:
*Professor Lawrence R. Klein**
Professor John Knowles

May 2006

* Sincere thanks to Dr. Klein for giving me the opportunity to learn Econometrics. I was daily touched by his generosity and his kindness for teaching me whom he had no obligation to. Through him, his passion to teach and his love for the field of economics, I have come to a decision to pursue the field of economics and it is my desire to follow his footsteps.

ACKNOWLEDGEMENTS

It would not have been possible for me to have carried out this thesis if it were not for Professor Lawrence R. Klein's generosity. I still vividly remember walking into his office for the first time, in the summer of 2005, simply wanting to have a conversation with an econometrician who has revolutionize the field of macroeconomics through econometric model building and forecasting. Simple conversation led to an eventual weekly meeting, where I was introduced to the field of economics that Professor Klein has been doing all his life. This unique experience of learning economics outside of the classroom setting has been an unforgettable experience and it will surely remain as one of the highlights of my undergraduate career. I sincerely thank Professor Lawrence R. Klein for giving me the opportunity to recognize my passion for the field and further allowing me to excel in this field of economics.

I would also like to thank Ms. Wendy Mak who helped me in learning the forecasting techniques. Due to her help and her kindness in sharing her experience with forecasting the high frequency China model, I was able to learn and use the techniques in this paper.

I also thank Kwanjeong Educational Foundation of South Korea who supported me financially during my four years of studies at the University of Pennsylvania. It would not have been possible for me to study at the University of Pennsylvania without their support. The fellowship signified their trust in my potential and my ability which enabled me to discover my passion and further thrive in the field. I sincerely thank their support.

Lastly, my thank goes out to my parents for all they have done for me in my life.

ABSTRACT

This paper assesses the High Frequency Forecasting Model using the specific example of South Korea's Economy. 26 monthly indicators were selected based on the Cob-Web model to represent South Korea and was used to forecast its quarterly GDP (1 step ahead, 4 steps ahead forecasts). The Principal Component analysis was employed to avoid the multicollinearity problem and fine tuning of model for each 10 samples were done accordingly. As a benchmark model, extrapolation was replicated on the ARMA(1,1) model for comparison sake. For the squared point forecast error, the result show that it is unable to reject the null hypothesis of the two models having equal predictive accuracy. For interval forecast, while ARMA(1,1) process is successful with its ability to encompass the actual values within the confidence interval, it also indicates that ARMA(1,1) is much more of a conservative model than the High Frequency Model. The test showed that High Frequency model has a smaller standard error band width, where three out of four tests showed the band width less than 40% of the bandwidth of the ARMA model and one out of four test with bandwidth less than 60% of the benchmark model. While only one out of four High Frequency model was successful in the confidence interval test in terms of its coverage, the test showed that High Frequency Model has the potential to become even more powerful by becoming more precise and accurate in nature. Especially when the modeling of the High Frequency model required much judgment and experience, and provided that High Frequency modeling was done by a novice (myself), one could only expect an improvement in this specific High Frequency model.

1. INTRODUCTION

With the advent of the information age, the abundant flow of datasets has been made possible that enables the econometricians of this age to exploit a new avenue of economic forecasting, forecasting in high frequency interval. With the wealth of information flow available, in a frequency that goes as frequent as monthly and weekly, produced and publicly published by the statistical office of respective countries or organizations, repeated modification and adaptation is made possible for modeling and forecasting. While improvement and expansion of information flow signify more resources available for the econometricians, by no mean does this suggest that econometricians' task in forecasting more accurately is mitigated. According to Klein(1993)[1] even with the expansion of information flow, there are problems that must be faced in attempt to take advantage of the provision. For example, higher frequency data flows (monthly, weekly, daily, etc) are prone to high serial correlation. While it is beneficial in that it provides the basis for extrapolation, it produces inefficiencies in parameter estimates. Also, at a frequency level that is in the monthly, weekly or daily form econometricians must cope with very short run shocks such as severe winter, drought in summer that is often not encountered in frequency of annual data analysis. [2]

This paper takes advantage of the current availability of datasets by modeling and forecasting South Korea's GDP using high frequency interval methods. Monthly datasets that are available in the National Statistical Office of Korea[3] and Bank of Korea[4] was used once indicators that well represent South Korea's GDP were selected. Selection of indicators was based on the familiar cob-web.[5] Compromise needed to be made with the indicators since not all the indicators that were desired were available. Furthermore, for comparisons sake, in order to test the adeptness of high frequency forecasting method (monthly) done through the Principal Component analysis, simple

[1] Economic Forecasting at High Frequency Intervals, Lawrence R. Klein & J.Y. Park. Journal of Forecasting, Volume12, (1993), pp301-319
[2] Economic Forecasting at High Frequency Intervals, Lawrence R. Klein & J.Y. Park. Journal of Forecasting, Volume12, (1993), pp301-319
[3] National Statistical Office of Korea, http://www.nso.go.kr/eng/index.html
[4] Bank Of Korea, http://www.bok.or.kr/index.jsp
[5] The Treatment of Expectation in Econometrics, C.F Carter and J.L Ford, Uncertainty and Expectations in Economics, eds. Pp175-190.

Autoregressive moving-average process forecast was replicated in the same interval using the quarterly data of GDP of South Korea.

It must be mentioned that high frequency modeling approach while done under the supervision of Dr. Lawrence R. Klein, it does not signify an exact reproduction of Dr. Klein's brilliant work such as the Wharton Model. Not only does this model of South Korea lack the size in terms of equations, but also due to the lack of the experience of the modeler (myself), I doubt that the full potential of the high frequency modeling approach developed by Dr. Klein transpired in this project. As Paul Samuelson[6] pointed out,

"Of a dozen such models that I know, moreover, eleven out of twelve, including Klein's, also put in judgment at the last stage. In other words, there are add-ons to the intercept coefficients of the regressions. Of course, the goal of science would be something that is reproducible, so that the assistant to Michelangelo could be almost as good as Michelangelo but that isn't the case with the models."

I have learned over the year that modeling require not only technical skills in the statistical methods but also experience and know-how. Therefore, if there is a case of any deficiencies in this paper regarding the high frequency model, I am accountable for all of the omissions.

[6] Research News, The 1980 Nobel Memorial Prize in Economics, Science Vol. 210. 14 November 1980

2. BACKGROUND INFORMATION FOR MODELING

Autoregressive moving-average (ARMA) process

For autoregressive process (AR), current observation depends on the lagged observations which is also known as a *stochastic difference equation*[7], while moving average process (MA) observes random variable dependent on the lagged unobservable shocks. The two processes are often combined and is called the ARMA process. Autoregressive moving average is often denoted as ARMA(p,q) where p signifies the order of AR while q signifies the order of MA. Thus, when either one of the values of p or q is 0, the ARMA reverts to either an AR process or a MA process. Mathematically, ARMA(p, q) is represented as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$
$$\varepsilon_t \approx WN(0, \sigma^2)$$

where $|\theta|$<1 and $|\varphi|$<1 for invertibility and stationarity respectively. ARMA(1,1) which will be employed for the analysis corresponds to p=q=1,

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$
$$\varepsilon_t \approx WN(0, \sigma^2)$$

ARMA models, by taking into account both the Autoregressive component and the Moving Average component is known to be highly accurate and highly parsimonious.[8]

Principal Components of Multivariate Observations

[7] Element of Forecasting, Third Edition, Francis X Diebold (2004)
[8] Element of Forecasting, Third Edition, Francis X Diebold(2004)

Principal Component analysis also known as Hoteling (1933)[9] Transform is a statistical method used to simplify a dataset. It is a linear transformation that selects the maximum amount of variance in components. Each of these components is ranked in order so that the first principal component accounts for the greatest variance. The second component accounts for maximum variance that is not accounted in the first component. Therefore, Nth component accounts for the maximum variance that is not accounted in all the previous components. These ranked weights are found in EVIEWS by looking at the eigenvalues which represents the variation in each component. The Principal Component analysis has a property of each components being uncorrelated with each other allowing the avoidance of multicollinearity problem that frequently arises in modeling a country's economy

For mathematical understanding of the Principal Component Analysis, Morrison (1980)[10] gives a concise definition. Suppose there are random variables X_1 ….., X_p with multivariate distribution with mean vector μ and covariance matrix ∑ with the assumption that the elements of each are finite. The rank of ∑ is r ≤ p, and q largest characteristic roots of ∑ are all distinct

$$\lambda_1 > \ldots . > \lambda_q$$

From N independent observation vectors, it can be written as a matrix form that is N x p

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots \\ x_{N1} & \cdots & x_{Np} \end{bmatrix}$$

With the ordered ranks of ∑ and X the first component, which represents the largest variance is a linear compound

[9] Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24,* 417-441.
[10] Multivariate Statistical Methods, Third Edition, Donald F Morrison (1980)

$$Y_1 = a_{11} X_1 + \ldots + a_{p1} X_p = a_1^{'} x$$

whose coefficient a_{i1} are the elements of the characteristics vector with the greatest characteristics roots ι_1 of the covariance matrix from the sample. If it is the case that $a_1^{'} a_1 = 1$, the characteristics root is interpreted as the variance of the same Y_1. Similarly, the second principal component is the linear compound

$$Y_2 = a_{12} X_1 + \ldots + a_{p2} X_p = a_2^{'} x$$

where

$$a_1^{'} a_2 = 0$$

which suggests orthogonal property ie) avoidance of multicollinearity of each components. This orthogonal property also allows the variances of successive components sum to the total variance of the responses.

Finally, the jth principal component of the sample of p-variate observation is

$$Y_j = a_{1j} X_1 + \ldots + a_{pj} X_p = a_j^{'} x$$

3. CONSTRUCTION OF MODEL

Data Selection on South Korea's GDP

In order to carry out the construction of High-Frequency Forecasting Model of South Korea, indicators that had "higher" frequency needed to be selected. So for GDP, since it is announced quarterly, monthly data or even weekly data would need to be selected to build the High Frequency Model. For South Korea's case, Statistical Office of South Korea[11] and Bank of Korea[12] had wealth of monthly statistics available. Concerning the actual selection of indicators that would represent GDP of South Korea, Cob-Web model

[11] National Statistical Office of Korea, http://www.nso.go.kr/eng/index.html
[12] Bank Of Korea, http://www.bok.or.kr/index.jsp

was taken into account. Named by the Economist Nicolas Kaldor, Cob-Web involves the market adjustment according to the prices and the outputs which occurs during time lags in production and prices. Familiar in the agricultural markets for perishable commodities, the model could be represented as[13]

$$q_t^s = \alpha_0 + \alpha_1 p_{t-1} + e_t \qquad \text{supply}$$

$$q_t^d = \beta_0 + \beta_1 p_t + u_t \qquad \text{demand}$$

$$q_t^s = q_t^d \qquad \text{market clearing}$$

$$e_t, u_t = error$$

Known for the description of the agricultural market, the model states that the producer produces according to the expected price and will supply to the market and will fetch whatever price it will bear. As can be seen above the supply equation, expected price is the lag price. Carter (1972) states that while this model is a typical model in the agricultural market, the dynamic system of Cob-Web does fit data for many different markets in a reasonable good manner as well.[14]

Selection of monthly indicators were thus carried out with the guidance of the structure of the Cob-Web model presented above, with a balance of Supply indicators, Demand Indicators as well as Market clearing indicators. The approach that was made for the actual selection of indicators was by studying each “potential” indicators and comparing with the historic trend of the GDP of South Korea. Many indicators followed the fluctuation of the GDP of South Korea which enabled the inference of those indicators being a strong component in the movement of South Korea’s GDP. Total of 26 indicators were selected where the breakdown were 17 supply, 4 demand and 5 market clearing.[15] Furthermore, in terms of frequency, because the employment of the Principal Component Analysis was going to be made, the indicators had to be shortened to match the indicator that had the latest starting date in the frequency. Therefore, while many

[13] The Treatment of Expectation in Econometrics, C.F Carter and J.L Ford, Uncertainty and Expectations in Economics, eds. Pp175-190.

[14] The Treatment of Expectation in Econometrics, C.F Carter and J.L Ford, Uncertainty and Expectations in Economics, eds. Pp175-190.

[15] Refer to the appendix for the GDP indicator summary sheet

datasets ranged from 1970's and 1980's to 2005, due to the restriction of assembling the datasets that starts at the same year and month, the sample size was shorten to 191 observations that ranged from January 1990 to November 2005. For the ARMA(1,1) model, as described in part 2, just the simple quarterly GDP was needed for construction. All of the indicators were multiplied by logarithmic for smoothing purposes.

ARMA(1,1) model

ARMA(1,1) modeling simply required the quarterly data of the GDP of South Korea for construction. With the autoregressive being the first lag, GDP of South Korea was forecasted by performing the regression of the equation shown below.

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\varepsilon_t \approx WN(0, \sigma^2)$$

High Frequency Model (Principal Component Analysis)

Once 26 indicators were decided, implementation of the Principal Component Analysis was made to construct a model that represents South Korea's GDP. As presented in part 2, the advantage of the Principal Component Analysis is that it does the job in avoiding the multicollinearity problem that might arise in the indicators. Especially in the case of constructing a model for South Korea, without solving this multicollinearity problem, substantial compromise must be made to compensate for the correlation problem. In the selection process of each indicator, it was not difficult to see many indicators moving in a very similar fashion and without the principal component indicators like these could not have survived in the selection process. For example, looking at the 26 indicators chosen for South Korea's GDP, there is a Motor Vehicle, Trailers and Semi trailers (MV) in the supply side and Sales of Motor Vehicles and Automotive Fuel (SALEMV) in the demand side which represents production of motor vehicles and sales of motor vehicles respectively. It is not difficult to imagine (and indeed it is) that these two indicators are highly correlated which could pose a problem in the regression analysis in the independent variable. As it could be seen from the scatter plot below in figure 1, without the employment of Principal Component Analysis two of the indicators would not have made it to the list with such high correlation.[16]

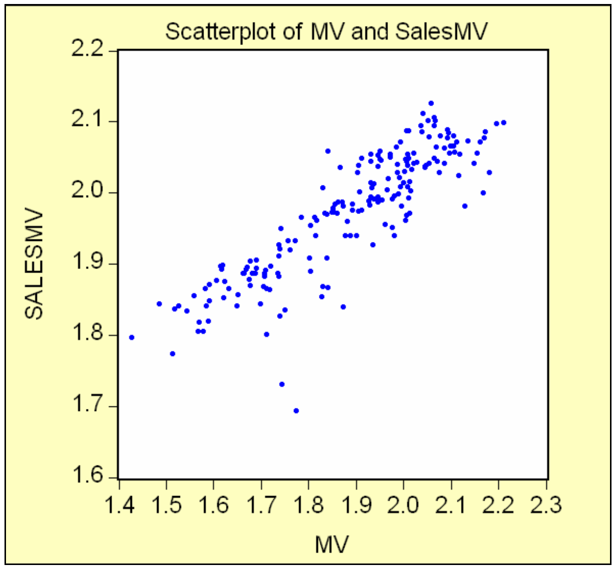


Figure 1

Once the indicators were chosen, principal component analysis was done. Most of the principal component analysis showed that the first 6 components accounts for more than 90.0% of the variance where it is ranked according to the greatest eigenvalues to the smallest.[17] As described in part 2,

$$Y_1 = a_{11}X_1 + \ldots. + a_{p1}X_p = a_1^{'}x$$

$Y_1, a_1^{'}, x$ represent the linear compound of the 1st component, coefficient of each indicators (26 in this case) and raw monthly indicator values respectively. Therefore, as could be seen from the appendix, the principal component has a matrix that is 26x 26,

$$PC_{GDP} = \begin{bmatrix} x_{1,1} & \cdots & \cdots & x_{1,26} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ x_{26,1} & \cdots & \cdots & x_{26,26} \end{bmatrix}$$

[16] Refer to the Appendix for the graph of all 26 indicators
[17] Refer to the Appendix for the principal component analysis output

where the first column space represents the component 1 with the greatest eigenvalue followed by component 2, component 3 and etc. The row space represents each indicator.[18]

Taking the first 6 component from the analysis, each coefficient was multiplied by its respective raw values for each indicator. Once $Y_1 \ldots\ldots Y_6$ were calculated, which spans the period of January 1990 to November 2005, that is represented as a 191 x 6 matrix, it was averaged out to a quarterly form to match the dependent variable frequency of the GDP. (63 x 6 matrix)

Dependent variable being the Logarithm of GDP each $Y_1 \ldots\ldots Y_6$ was included as the independent variable, where the equation was constructed to determine the Coefficients $C_1 \ldots . C_6$.

$$LogGDP = C + C_1 Y_1 + C_2 Y_2 + C_3 Y_3 + C_4 Y_4 + C_5 Y_5 + C_6 Y_6 + resid$$

Before making a "pseudo out of sample forecast", the model underwent fine tuning by performing the "ex-post" estimation. Also known as "unconditional forecasts", although strictly it is not considered as a "forecast", the process enables the identification of the model structure in terms of the significant coefficients as well as the dummy variables. In order to perform the "Ex-post" process, the sample size was shortened by 2 quarters where the Principal Component Analysis was replicated. With $Y_1 \ldots\ldots Y_6$ that is 2 quarters short, components $C_1 \ldots . C_6$ were found. Determination of the coefficients along with the inclusion of dummy variables as well as autoregressive or moving average made this model a final model. Keeping the significant coefficients and throwing out insignificant ones, the best model was selected based on the criteria such as SIC, AIC and Durban Watson statistics.[19] In order to test this model, new sets of variables (now including 2 quarters of raw data that were taken out) underwent the principal component analysis where a new set of components $Z_1 \ldots .. Z_6$ was calculated. Taking the last 2 quarters and multiplying each principal component coefficients with

[18] Refer to the Appendix for the principal component analysis output
[19] For More detail on the criteria refer to Element of Forecasting, Third Edition, Francis X Diebold(2004)

the raw values of indicators, the last 2 quarter sets were included in the original equation that consist of $C_1 \ldots C_6$ (which were determined by estimation) and $Y_1 \ldots\ldots Y_6$ (Coefficient of Principal Component analysis of 2 quarters short multiplied by the raw indicator values) With the LogGDP value available for verification, forecast was made using the new components $Q_1 \ldots\ldots Q_6$,

$$LogGDP = C_1Q_1 +. C_2Q_2 + C_3Q_3 + C_4Q_4 + C_5Q_5 + C_6Q_6 + resid$$

$$Q_n = Y_n + Z_n \qquad \text{where} \quad Y_n = \sum_{i=1}^{k-2} a_n^{'} x \;\; , \; Z_n = \sum_{i=k-1}^{k} a_n^{'} x$$

Once "ex-post" was done "pseudo out of sample forecast" was carried out. While "Ex-ante" is an out of sample forecast, where forecast is done outside of the existing sample space, due to the reason of testing the adeptness of the High Frequency Model which can only be done if the actual values are present for comparison, "simulated" ex-ante, which is labeled as the "pseudo out of sample forecast" in this paper, was performed. Pseudo out of sample forecast process basically implements exactly what is performed in the Ex-ante process with the exception that it was done within the sample size by shrinking the sample size with the pretension that the independent variables were not present. Therefore, by shrinking the sample size, independent variables were forecasted through ARIMA process to forecast the ultimate dependent variable of GDP.

To briefly describe the process of ARIMA, the process was developed by Box and Jenkins (1970) in the context of forecasting.[20] Relying only on the past behaviors of the variable, in this case each indicator *i,* Box and Jenkins analysis makes sure that the variable Y_t is stationary, where the mean of Y_t, its variance, do not depend on t. Usually, this is inspected through visual inspection of the correlogram of the estimated kth order autocorrelation coefficient where autocorrelation graph should show a die out fairly quickly as k becomes large. [21] ARIMA(p,d,q) constitutes p, d, q which signifies autogressive dimension, difference in model Y and dimension of moving average respectively. As stated in the introduction, identification and model selection of values

[20] A Guide to Econometrics, Fifth Edition, Peter Kennedy (2003)
[21] A Guide to Econometrics, Fifth Edition, Peter Kennedy (2003)

p,d and q while done under the criteria of such tests as SIC, AIC and Durban Watson Statistics, as a modeler, it required the most personal judgment to interpret some selected statistics where it was apparent that extensive experience in the field was preferred.

In the case of South Korea GDP, with 26 indicators, this meant forecasting each indicators through ARIMA process and obtaining the extrapolation values which subsequently was multiplied by the principal component coefficient of the shorten sample size. This was done for 10 different sample sizes that were simulated by shrinking the existing sample. For each sample size simulation of "ex-ante", 1 step ahead forecast as well as 4 steps ahead forecast were done. For comparison, same was done for ARMA(1,1) process and both simulated "ex-ante" were compared to the actual value for the assessment of the adeptness of each model. Figure 2 shows the diagram of the ex-ante forecast implemented for both ARMA and High Frequency Model.

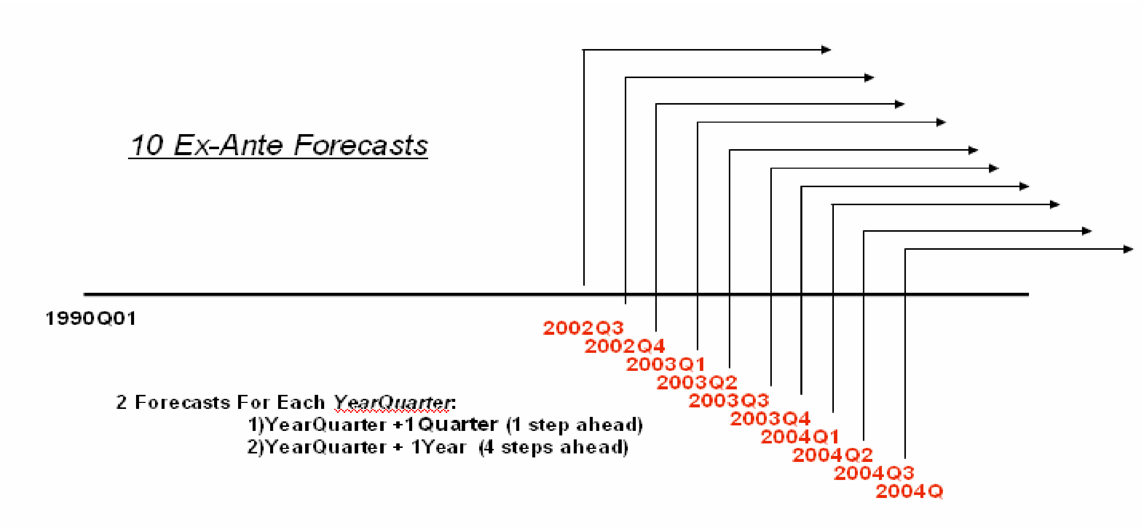


Figure 2

4. INTERPRETATION OF ESTIMATATION IN FINAL MODEL

In order to illustrate the model selection process for both ex-post and ex-ante forecast, one example from each tests were selected out of 10 samples.[22] As it can be seen in the

[22] For ARMA(1,1) , all 10 samples have same exact constitute in the model with AR(1) and MA(1). For High Frequency Forecast Model, each samples vary in their independent variables and this was selected according to the criteria that best makes the model white-noise.

EVIEWS output in figure 3 and figure 4, both has a sample size ranging from 1990Q2 to 2003Q2. Therefore, this particular model was used to forecast 2003Q3 GDP (1 step ahead forecast) as well as 2004Q2 GDP (4 steps ahead forecast).

Looking at table 3, the simple ARMA(1,1) shows that the model is white noise, indicative from the Durban Watson stat of 1.919. The R^2 is high with the value of 0.993 AR(1) is significant with the t-stat well above 2 while MA(1) stat has about 87% confidence level. Looking at the graph itself, as it should be in a simple ARMA(1,1) model, the fitted value has a lag of 1 period. To mention the shock that is apparent in the period of 1997-1998, the dip in the residual as well as the actual value of GDP represents the economic crisis that Korea went through during the time of the Asian currency crisis. Historically, South Korea announced in November 21st 1997 that it would seek about $20 billion in aid from International Monetary fund.[23] Following this movement by the Korean government, the Korean conglomerates made announcements to slim down their investment strategies in year 1998. For example, Samsung Group, one of the strongest Chaebols (conglomerates) stated that they will invest 30% less in year 1998.[24] This movement, with the reconstructing and reformation of the economy in the year 1998 caused the dip in the GDP of South Korea.

Simple ARMA(1,1) Model

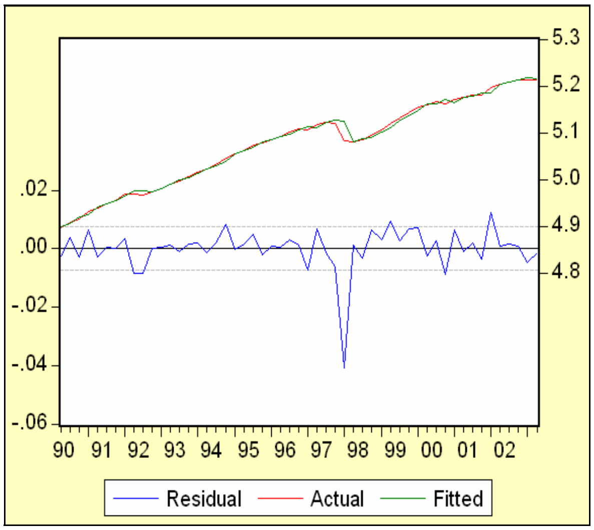


Figure 3

[23] Asian Wall street Journal. "Seoul Prepares to Wince as IMF Considers Pulling Strings" Aid to Bail out Korea from Deb Crisis May Depend on Changing Rules for Loans and Spending, December 1st 1997
[24] Asian Wall street Journal. "Chaebols Plan to Slim down by Trimming Spending, Debt" Conglomerates feel the pain as funds dry up, but will Korea inc Keep its promise to reform? December 1st 1997

Notes: This graph is one out of ten sample forecast of simple ARMA(1,1) model that was used as a benchmark for comparison purposes.

ARMA(1,1)				
Dependent Variable: LOGGDP				
Method: Least Squares				
Sample(adjusted): 1990Q2-2003Q2				
Variable	Coefficient	Std. Error	T-Statistics	Prob
C	5.37	0.21	25.56	0
AR(1)	0.98	0.013	73.28	0
MA(1)	0.21	0.139	1.537	0.1305
R squared	0.994			
DurbanWatson	1.920			
Akaike info Crit	-6.914			
Schwarz Crit	-6.802			
F-Statistics	3911.968			
Prob(F statistics)	0.000			
Inverted AR Roots	0.98			
Inverted MA Roots	-0.21			

Table 1

For High Frequency Model, that was employed by using the monthly data of the 26 indicators, has a much more sophisticated composition.[25] With the first 6 components representing the total variance of 90%-95% in the principal component analysis, as described in section 3, fine tuning of the model was made by including independent variables that would make the model white-noise. As it can be seen below in figure 4, out of the 6 components, represented as Cn where n is the nth component, only 4th component was dropped for its insignificance. Other component as it could be seen, are all significant with the t-statistics value well above 2. Also, in order to take into account the seasonality component, dummy variables were introduced. Quarterly dummies were introduced and were tested for its significance. Out of 4 quarterly dummies, 1st and 3rd dummies showed significance. In the EVIEWS output, DUMMY98 represents the dummy variable for the whole year of 1998 which was the shock year when South Korea went through the economic crisis. Because this event in 1998 was something that was 'unusual', dummy variable was introduced. Along with the dummy variable, autoregressive as well as moving average processes were included and those with significant values were kept. The final model represents a white noise property where Durban Watson stat is close to 2 while R^2 is 0.999. Graphically, it is not difficult to see that the model fits well with the actual values and the residual being stochastic.

[25] It took significantly more time in constructing the model with the principal component analysis in comparison to the simple ARMA(1,1) model.

While for the simple ARMA(1,1) model, no real selection were needed since it was a benchmark model test that was done for a comparison sake, for the High Frequency Model, basic criteria that was used in assessing the model and selecting the model was done through first removing the insignificant independent variables and including variables that made the model white noise property. In cases there were equally qualified model, selection was made primarily looking at the SIC and AIC and selecting the lowest value of SIC of the model.

High Frequency Model

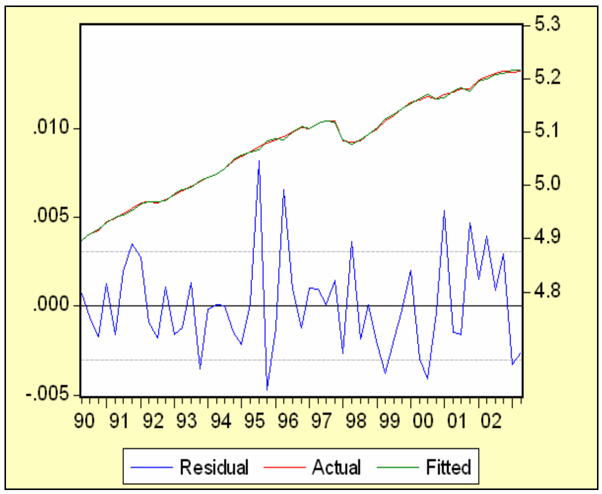


Figure 4

High Frequency Model				
Dependent Variable: LOGGDP				
Method: Least Squares				
Sample(adjusted): 1990Q2-2003Q2				
Variable	Coefficient	Std. Error	T-Statistics	Prob
C	4.806	0.161	29.822	0.000
C1	0.042	0.010	4.390	0.000
C2	-0.040	0.004	-9.077	0.000
C3	0.020	0.004	4.745	0.000
C5	0.013	0.003	4.542	0.000
C6	0.021	0.005	4.465	0.000
DUMMY1	0.006	0.001	5.144	0.000
DUMMY3	0.007	0.001	6.003	0.000
DUMMY98	-0.010	0.002	-4.087	0.000
AR(1)	0.988	0.003	334.721	0.000
MA(1)	-0.253	0.093	-2.714	0.010
MA(3)	-0.715	0.092	-7.805	0.000
R squared	0.999			
DurbanWatson	2.113			
Akaike info Crit	-8.560			
Schwarz Crit	-8.114			
F-Statistics	4274.110			
Prob(F statistics)	0			
Inverted AR Roots	0.99			
Inverted MA Roots	0.99			
Cn= Nth Principal Component				
DummyS= Sth quarter seasonal dummy				
Dummy98= Asian Crisis dummy for year 1998				

Table 2

Notes: The above output of both the table and figure represents one out of ten samples of the high frequency forecast. Because each forecasts differ in terms of the principal components used, each of ten output table differ. The above was selected for discussion purposes.

5. RESULT

In order to assess the adeptness of the two models that has been constructed and performed for Ex-ante forecasting attention to mere point forecast values is not sufficient. A point forecast, in this case the forecast of respective 1 step ahead and 4 steps ahead ex-ante forecast done by both the ARMA(1,1) model and the High Frequency Model does not provide much information on its own about the precision and the reliability of the forecast of the model. Therefore, it is necessary to take other approaches to test the model's adeptness. One of the approach is the confidence interval. Information about the precision of an interval estimate is revealed by the width of the interval.[26] The standard error that the EVIEWS program produces is from the normal distribution of the population. Therefore, 1 SE signifies about 66% chance of the true result being present in the interval. Similarly, 2 SE signifies about 95% chance of the true result being present in the interval. So for the case of the forecast done for both ARMA(1,1) and High Frequency Model, smaller the value of SE, it suggests more precision in the ability to forecast. This result section compares both the point forecast and the interval forecast by employing methods to test for the adeptness of the two models.

Average of Forecast Error Comparison

As it could seen from Table 3, for the Average of Forecast Error (Actual-Forecast)

$$\frac{1}{n}\sum(y_t - \hat{y}_t)$$

for 10 sample sizes, ARMA(1,1) performs better for 1 step ahead forecast with the value of 0. However, for 4 steps ahead forecast, High Frequency model produces lower

[26] Probability and Statistics, For engineering and the Sciences, Sixth Edition, Jay L. Devore (2004) Chapter 7 pg282.

average values of differences in the sample size of 10. It must be noted that, looking at the Average of Difference $\frac{1}{n}\sum(y_t - \hat{y}_t)$, it shows that for 10 samples forecasts we performed, on average, High Frequency Model "over-forecasts" while ARMA(1,1) "under-forecasts" by its respective values.

Average of Forecast Error		
	HF	ARMA(1,1)
1step ahead	-0.003	0.000
4 steps ahead	-0.002	0.003
Average of SE		
	HF	ARMA(1,1)
1 step ahead	0.005	0.012
4 steps ahead	0.007	0.019

Table 3

Absolute Point Forecast Error Comparison

Figure 5 and Figure 6 shows each difference (actual – forecast) for 10 samples. (absolute value of difference for comparison)

$$e_t = \left| y_t - \hat{y}_t \right|$$

$$y_t = actual$$

$$\hat{y}_t = forecast$$

As could be seen, for 1 step-ahead forecast, 5 of the ARMA(1,1) forecast lie below High Frequency Model. This suggests that 5 out of 10 ARMA(1,1) forecast performed better in the sense that it was closer to the actual value. High Frequency in 1 step-ahead forecast managed to be lower in only 2 samples where 3 were tied. In contrast to figure 5, Figure 6 shows the 4 step-ahead forecast difference, where in this case shows High Frequency Model out performing ARMA(1,1) by 4 to 3 in terms of its closeness to the actual value. 4 out of 10 ARMA(1,1) model forecast resides above the High Frequency forecast which suggests that ARMA(1,1) forecast is further away from the actual value. This suggests that in 4 steps-ahead forecasts, High Frequency Model performs better in terms of its closeness to the actual value. It must be mentioned that the table and graph shown in table 4, table 5 and figure 5 and figure 6 respectively is the absolute value of the difference which was done in order to compare the distance from the actual value.

Difference*- 1 step ahead		
	ARMA(1,1)	HF
1	0.003	0.003
2	0.002	0.007
3	0.005	0.018
4	0.007	0.002
5	0.002	0.003
6	0.005	0.006
7	0.006	0.002
8	0.003	0.003
9	0	0.005
10	0.003	0.003
*Absolute value		

Table 4

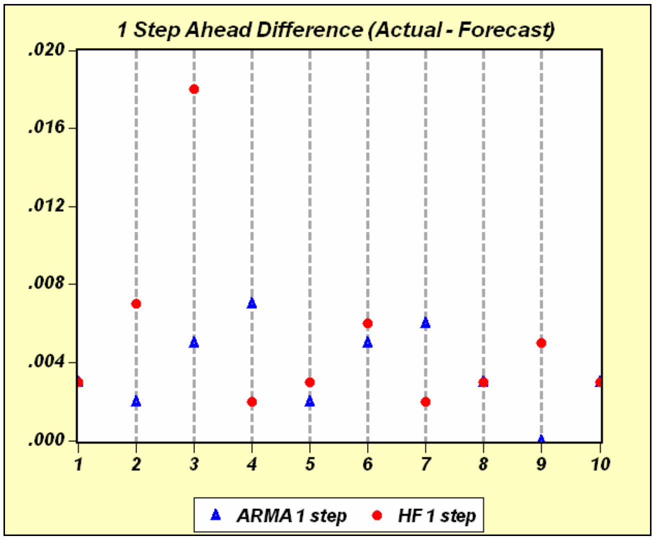


Figure 5

Difference*- 4 steps ahead		
	ARMA(1,1)	HF
1	0.005	0.003
2	0.003	0.007
3	0.004	0.008
4	0.004	0.003
5	0.01	0.004
6	0.011	0.001
7	0.005	0.005
8	0.004	0.011
9	0.001	0.001
10	0.003	0.003
*Absolute value		

Table 5

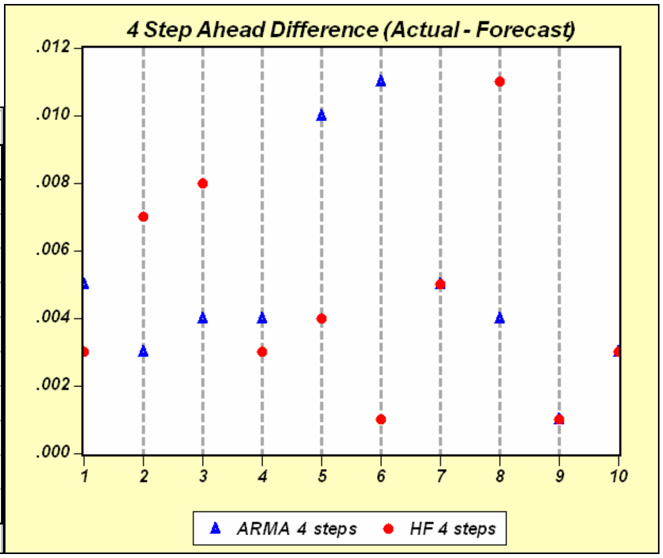


Figure 6

Squared Point Forecast Error Comparison

Squared forecast error, similar to the absolute error comparison allows the identification of the distance between the actual value and the forecast value. By combing the two errors that are squared, it gives a direct comparison by looking at the magnitude of the squared forecast error.

$$SFE = \{e^2{}_{HF,t} - e^2{}_{ARMA,t}\}$$

$$e^2 = (Actual - Forecast)^2$$

As shown above, the equation SFE, Squared forecast error which is the subtraction of the ARMA from High Frequency is tabulated and plotted in table 6 and figure 7 respectively. The positive sign of the values in table 6 signifies that high Frequency has a larger squared forecast error where the distance from the actual value is further away from the forecast value. As it could be seen from the 1 step ahead forecast, only 3 out of 10 forecasts show that high frequency had a better performance in terms of the forecast value being closer to the actual value than the simple ARMA forecast. However, for 4 steps ahead forecast, half of the forecasts of High Frequency was superior to the ARMA model as the square forecast error (High Frequency – ARMA) is negative. Although by inspection of 10 sample sets it is possible to discern which test has done better, it is clear that the values of all the squared forecast errors are very close to 0 where in the case the values were rounded off to the 3rd decimal places all of the values would be 0. Therefore, it is imperative to do a significant test for these values for verification of whether or not these squared forecast errors are actually different statistically.

Square Forecast Error (High Frequency - ARMA)	
1 step ahead	4 steps ahead
0.00000690	0.00006812
0.00004092	0.00003682
0.00029986	0.00004449
-0.00004880	-0.00001212
0.00000409	-0.00008298
0.00001938	-0.00011429
-0.00003737	-0.00000517
-0.00000089	0.00011536
0.00002060	0.00000016
0.00000783	-0.00003129

Table 6

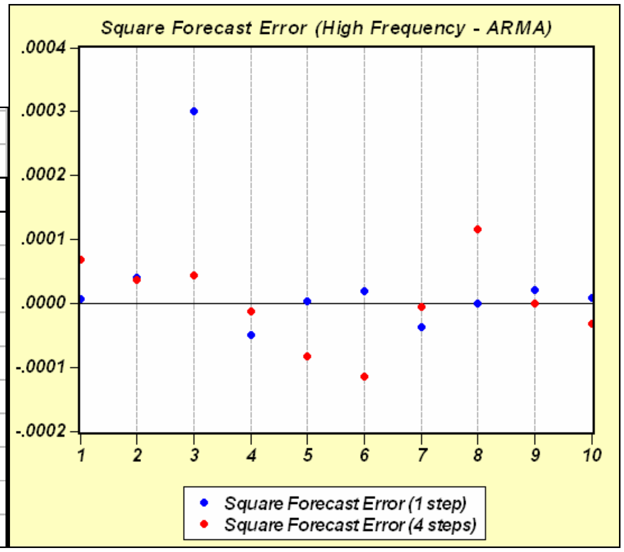


Figure 7

Diebold-Mariano Test[27]

Diebold-Mariano test takes the loss function to determine whether one model predicts better than another. With the null hypothesis of the two models having the same loss function, Diebold-Mariano statistics is used to comparing predictive accuracy.

$$H_0 : \{e^2{}_{HF,t} - e^2{}_{ARMA,t}\} = 0$$
$$H_1 : \{e^2{}_{HF,t} - e^2{}_{ARMA,t}\} \neq 0$$

The Diebold-Mariano test statistics is[28]

$$S = \frac{\bar{d}}{(LRV_{\bar{d}} / T)^{1/2}} \quad \text{where}$$
$$\bar{d} = \frac{1}{T_0} \sum_{t=t_0}^{T} (d_t)$$

[27] Diebold, Francis X, Mariano, S Roberto, "Comparing Predictive Accuracy", *Journal of Business & Economic Statistics,* July 1995, Vol. 13, No.3

[28] Zivot, Eric, The Diebold-Mariano Statistics for Comparing Predictive Accuracy, pg 6-7, April 8, 2004 http://faculty.washington.edu/ezivot/econ584/notes/forecast.pdf

$$d_t = (e^i_{t+h|t})^2 - (e^i_{t+h|t})^2 \qquad \text{i=1,2}$$

$$LRV_{\bar{d}} = \mathrm{cov}(d_t, d_{t-j})$$

Diebold and Mariano show that under the null of equal accuracy,

$$S \overset{A}{\sim} N(0,1)$$

Where we reject the null hypothesis at 5% level if

$$|S| > 1.96$$

Specifically for the 10 sample tests that were done for this paper,

$$\bar{d} = \frac{1}{10}\sum_{t=1}^{10}\{e^2{}_{HF,t} - e^2{}_{ARMA,t}\}$$

$$e^2 = (Actual - Forecast)^2$$

$$LRV = \frac{SE_d}{\sqrt{10}}$$

The results of the Diebold-Mariano test show that it is unable to reject the null hypothesis of equal predictive accuracy. Therefore, while it was shown in the previous section that the ARMA model outperformed High Frequency by 7 to 3 in the 1 step ahead squared forecast error, statistically, this test show that the two does not differ in terms of their squared forecast errors. Table 7 is the summary of the Diebold-Mariano Test.

Diebold Mariano Test		
5 % level		
	1 step ahead	4 steps ahead
S	1.00744	0.08804
5% level	unable to reject	unable to reject

Table 7

Interval Forecast Comparison

While point forecast error is an important measure of the success of a model and its ability to forecast, as it measures how close it is from the actual value, it is the interval that is more important in assessing the adeptness of a model. Comparison of Standard Error is much more meaningful in the regard of assessing the model since it signifies the precision of the model's ability to forecast. The width of the interval specifies the precision or accuracy[29]

$$IF = PF \pm z \cdot SE$$

$$SE = s\sqrt{1 + x_t'(X'X)^{-1}x_t}$$ [30]

z = standard deviation, IF= interval forecast, PF = point forecast

Most well known interval forecast comparison is the confidence interval of 95% which represents 2 standard deviations. Plotting the interval forecast of 2 standard error bands, and as it could be seen on figure 10, the width of the error band of High Frequency model is narrower than the simple ARMA(1,1) process. However, while the high frequency model has a narrow bar which could signify more precision in the forecast model, in the confidence interval test, it fails the 95% confidence interval test as only 9 out of 10 samples (90%) encompasses the actual value that is represented as a black dot in figure 8. While the model should be 95% accurate, the High frequency model is only 90% accurate. On the other hand, the simple ARMA(1,1) model encompasses all 10 forecasts within the 95% interval as shown in figure 10.

[29] Probability and Statistics, For engineering and the Sciences, Sixth Edition, Jay L. Devore (2004) Chapter 7 pg287.
[30] EViews 5.1 Help "Forecast Basics"

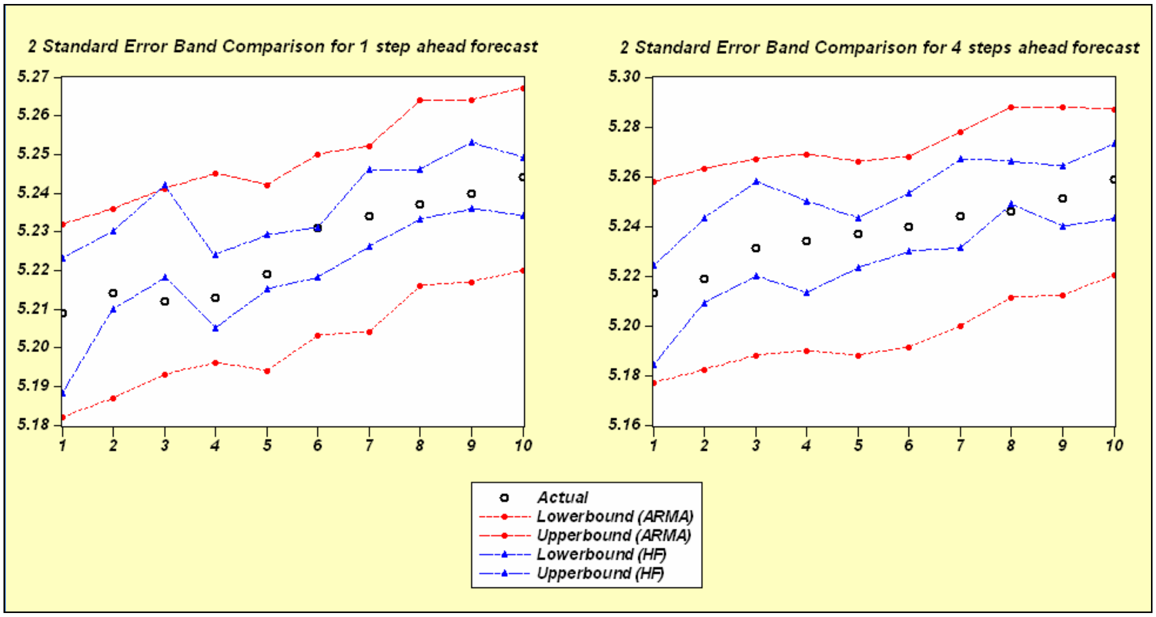


Figure 10

Figure 11 shows the 50% confidence interval test which signifies 50% of the probability of the actual value being under the confidence interval of 50%. As it was seen in figure 10, the 50% confidence interval (0.68xSE) also shows that High Frequency model has a narrower width which could signify more precision or accuracy in the model .However, similar to the case of the 2 standard error bar, when put under test for the confidence interval the 1 step ahead forecast for High Frequency only shows 4 out of 10 samples encompassing the actual value. While the model should have 50% of the chance of having the actual value in the confidence interval, for the 1 step ahead forecast, High Frequency fails this test. On the other hand, the 4 steps ahead forecast interval encompasses 70% of the actual values which is above the 50% confidence interval.

For simple ARMA(1,1) test, both 1 step ahead and 4 steps ahead forecast intervals picks up all 10 samples showing 100% rate in the 50% confidence interval test.

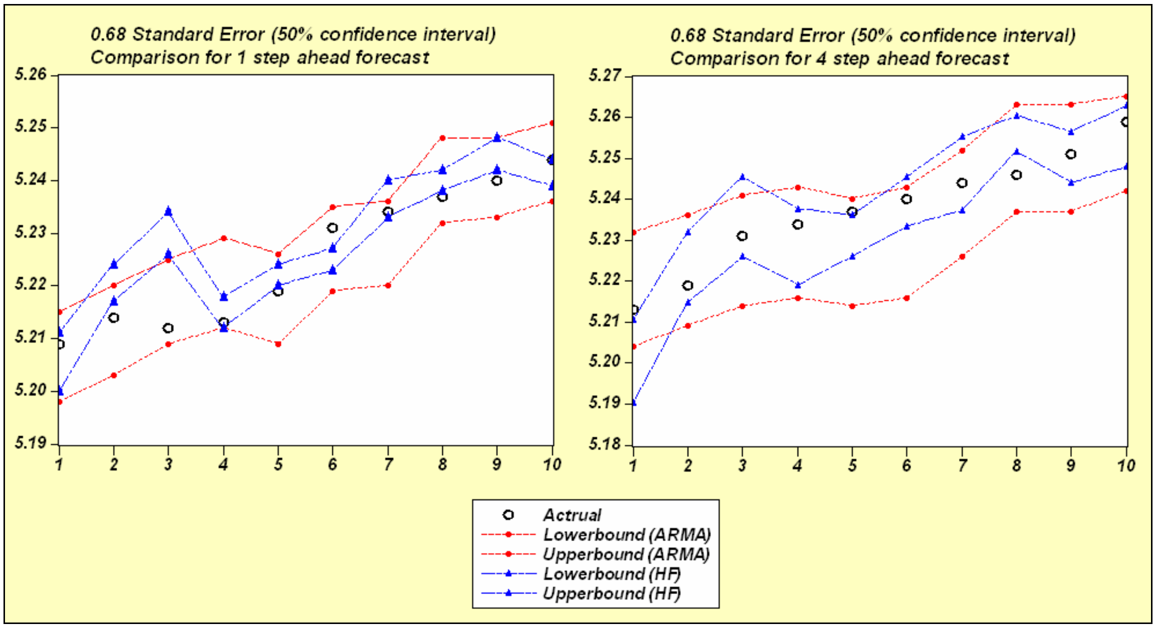


Figure 11

Table 8 and table 9 summarize the interval forecast for both the 95% confidence interval and the 50% confidence interval. As table 8 shows, for all of the confidence interval tests, High Frequency model outperforms in terms of the average length of the interval. The lengths for all of the high frequency are shorter where this is easily seen from the ratio of High Frequency/ARMA where the value is less than 1. In fact, three out of four average lengths of the intervals are less than 40%[31] of the length of its direct comparison model the ARMA model. The model that is above the 40% mark is still less than the average length of the ARMA model in that the ratio of High Frequency to ARMA is 0.579.

Average Length of Interval			
95% Confidence Interval			
	ARMA(1,1)	HF	HF/ARMA
1 Step AHead	**0.048**	**0.019**	0.390
4 Steps Ahead	**0.077**	**0.030**	0.386
50% Confidence Interval			
	ARMA(1,1)	HF	HF/ARMA
1 Step AHead	**0.016**	**0.006**	0.390
4 Steps Ahead	**0.02624**	**0.01518**	0.579

Table 8

[31] Values of HF/ARMA are 0.39, 0.386 and 0.39 for 95% confidence interval of 1 step ahead, 4 steps ahead and 50% confidence interval of 1 step ahead respectively.

As stated above concerning the actual performance of the confidence interval testing, ARMA(1,1) model picks up all the actual values by encompassing all of its actual values in their interval. However, for High Frequency model, only 90% are picked up by the 95% confidence interval test for both 1 step ahead and 4 steps ahead forecast while for the 50% interval test, 1 step ahead forecast produces 40% of the success rate in picking up the actual value in its interval. For 4 steps ahead, it is well above the 50% confidence interval by picking up 70% of the actual values in the interval width.

Coverage Frequency		
95% Confidence Interval		
	ARMA(1,1)	HF
1 Step AHead	**100%**	**90%**
4 Steps Ahead	**100%**	**90%**
50% Confidence Interval		
	ARMA(1,1)	HF
1 Step AHead	**100%**	**40%**
4 Steps Ahead	**100%**	**70%**

Table 9

6. CONCLUSION

This paper, using a specific example of South Korea's GDP, assessed the strength and weaknesses of the High Frequency Model. In order to assess the adeptness of a monthly forecasting model that was constructed using the Principal Component analysis, simple ARMA(1,1) model was used as a benchmark model for comparison. While simple ARMA(1,1) takes into account the past value of GDP to forecast the GDP of South Korea, thus forecasting quarterly GDP using quarterly reported GDP, High Frequency Model differs with the simple ARMA(1,1) in that it first employs the monthly indicators of South Korea to forecast the quarterly GDP of South Korea. In terms of modeling, in essence, High Frequency Model constructs a model that could best represent South Korea's Economy and uses these "high frequency" indicators to first forecast its indicators (6 months ahead typically) in order to forecast the GDP of South Korea. 26 indicators were chosen based on the structure of the Cob Web Model, and the Principal Component analysis was done to avoid the multicollinearity problem between the indicators. 6 months extrapolation was done using the ARIMA for each indicator which was multiplied by the principal component found in the principal component analysis (typically the first 6 component) and these values were divided in quarterly form to determine the independent variable. Fine tuning of model was done by including dummy variables as well as autoregressive and moving average component to extrapolate the GDP for its forecast.

The result show that for the point forecast error, the two are not significant as Diebold-Mariano test failed to reject the null hypothesis of the squared forecast error being of equal predictive accuracy. In terms of interval forecast, while ARMA(1,1) was able to coverage all the actual values with 100% rate for both 95% and 50% confidence interval, High Frequency out performed the ARMA(1,1) in terms of the error band width where all 4 of the tests (2 for 1step ahead and 2 for 4 steps ahead) , had the bandwidth size of 60% of its benchmark model's error width. This suggested that while ARMA(1,1) is highly successful in coverage, it also indicated its conservative nature of the model. For High Frequency, while only one out of four tests showed success in terms of confidence interval, it was not significantly off from the actual value in terms of its coverage.

As mentioned in the introduction section, unlike the ARMA(1,1) model, modeling the high frequency model for South Korea required many important decisions from the modeler. Not only did the High Frequency model require deep understanding of the dynamic macroeconomic behavior of the South Korea in the selection process of the monthly indicators, which is a pivotal part of modeling the economy, but also in the statistical, econometric sense, through the principal component analysis and the ARIMA extrapolation process, experience and know-how was preferred as it required important decision in model selection. And as a novice who has undertook this model building task for South Korea for the first time, it is the personal opinion that the result does not reflect the true nature of the high frequency model's potential. One thing for sure is that the high frequency model, while failed 3 out of 4 confidence interval test, it did show great precision in terms of the error band. Therefore, while ARMA(1,1) was statistically successful, it could be said that it is also a very conservative method of forecasting with the standard error being very large. However, for the case of High Frequency, this paper showed that High Frequency model is not as conservative as the ARMA(1,1) model. The error band showed that it is substantially smaller which signifies that when done right, the forecast could result in a much more precise and accurate way. The High Frequency model that has been constructed on this paper has sufficient room for improvement as the modeler gains more experience and expertise in the field. Not only has this method been proven with the larger model employed by Professor Lawrence R. Klein throughout the decades but as a modeler, having experienced this process once, it is the belief that when done again, vast improvement could be expected.

7. REFERENCES

Bodkin, Ronald G., Lawrence R. Klein, Kanta Marwah (1991): "A history of Macroeconometrics Model Building"

Carter, C.F, J.L Ford (1972): The Treatment of Expectation in Econometrics, *Uncertainty and Expectations in Economics*, eds. Pg175-190

Cho, Namju (1997): "Seoul Prepares to Wince as IMF Considers Pulling Strings" Aid to Bail out Korea from Deb Crisis May Depend on Changing Rules for Loans and Spending, *Asian Wall Street Journal*, December 1st 1997

Devore Jay L. (2004): Probability and Statistics for Engineering and the Sciences 6th edition. Thomson Brooks/Cole

Dhrymes, Phoebus J. (1978): Introductory Econometrics, Springer-Verlag

Diebold, Francis X, Mariano, Roberto S.(1995): Comparing Predictive Accuracy, *Journal of Business & Economic Statistics*, Vol. 13 No3

Diebold, Francis X. (2004): Element of Forecasting 3rd Edition, Thomson southwestern

Marwah, Kanta., (1997): Selected Papers of Lawrence R Klein, Theoretical Reflections and Econometrics Applications.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24,* 417-441.

Jeffrey M. Wooldridge (2003): Introductory Econometrics: A modern Approach 2nd edition. Thomson Southwestern

Johnston J (1972): Econometric Methods 2nd Edition, McGraw Hill book

Kennedy, Peter (2003): A guide to Econometrics 5th Edition, The MIT press

Klein, Lawrence R., Richard M. Young. (1980): Introduction to Econometrics Forecasting and Forecasting Models, Lexington Books

Klein, Lawrence R., J.Y. Park, (1993): Economic Forecasting at High Frequency Intervals, *The Journal of Forecasting,* pp. 301-319

Klein, Lawrence R., Shinichi Ichimura (2000): Econometric Modeling of China, World Scientific

Morrison, Donald F. (1990): Multivariate Statistical Methods 3rd Edition, McGraw Hill Publishing Company

Quantitative Micro software (2004): Eviews 5 User's Guide

Research News, The 1980 Nobel Memorial Prize in Economics, *Science* Vol. 210. 14 November 1980

Schuman, Michael (1997): "Chaebols Plan to Slim down by Trimming Spending, Debt" Conglomerates feel the pain as funds dry up, but will Korea inc Keep its promise to reform?, *Asian Wall Street Journal*, December 1st 1997

Zivot, Eric, The Diebold-Mariano Statistics for Comparing Predictive Accuracy, pg 6-7, April 8, 2004 http://faculty.washington.edu/ezivot/econ584/notes/forecast.pdf

Data Sets

Most of the datasets were collected from the central bank, Bank of Korea (http://www.bok.or.kr/index.jsp) and the National Statistical office of South Korea (http://www.nso.go.kr/eng/index.html). For each indicator, refer to the appendix where there is a legend of the GDP indicators that were used with the unit of output.

8. APPENDIX

GDP				
South Korea				
MONTHLY INDICATORS (26) - 1990M1-2005M11				
AGR	SUPPLY:	Agriculture	VOLUME INDEX	Y2000=100
COMMERCIALBUILDING	SUPPLY:	Commercial	FLOOR AREA	Thousand m2
COMMUNICATION	SUPPLY:	Communication	DOLLAR	Y2000=100
COMPANDOFFICE	SUPPLY:	Computers and Office Machinery	SEASONALLY ADJUSTED INDEX	Y2000=100
CONST	SUPPLY:	Construction	PRODUCTION INDEX	Y2000=100
DWELLINGBUILDING	SUPPLY:	Dwellings	FLOOR AREA	Thousand m2
ELEC	SUPPLY:	Production: electricity seasonally adjusted	PRODUCTION INDEX	Y2000=100
ELECTRICMACHINE	SUPPLY:	Electrical Machinery and Apparatus n.e.c.	SEASONALLY ADJUSTED INDEX	Y2000=100
EXPORTTOTAL	SUPPLY:	Total exports	DOLLAR	Thousand dollar
FACTORYBUILDING	SUPPLY:	Factory	FLOOR AREA	Thousand m2
FISH	SUPPLY:	Fisheries	VOLUME PER WEIGHT	M/T (metric/ton)
INTERMEDGOOD	SUPPLY:	Intermediate Goods	PRODUCTION INDEX	Y2000=100
MANUFACEQUIP	SUPPLY:	Manufacturing Equipment	PRODUCTION INDEX	Y2000=100
MANUFACT	SUPPLY:	Manufacturing	SEASONALLY ADJUSTED INDEX	Y2000=100
MINING	SUPPLY:	Mining	SEASONALLY ADJUSTED INDEX	Y2000=100
MV	SUPPLY:	Motor Vehicles, Trailers and Semitrailers	SEASONALLY ADJUSTED INDEX	Y2000=100
TEXT	SUPPLY:	Textiles(Except Sewn Wearing Apparel)	SEASONALLY ADJUSTED INDEX	Y2000=100
PETOIL	DEMAND:	Crude Oil and Petroleum products import	VOLUME	1000 barrel
RETAILTRADE	DEMAND:	Retail Trade	VOLUME INDEX	Y2000=100
SALESMV	DEMAND:	Sales of Motor Vehicles and Automotive Fuel	VOLUME INDEX	Y2000=100
WHOLESALETRADE	DEMAND:	Wholesale Trade	VOLUME INDEX	Y2000=100
DEPOSITORCORP	MARKET CLEARING:	Deposits at CBs and SBs by depositors:CORPORATIONS	WON	Billion Won
DEPOSITORHOUSEHOLD	MARKET CLEARING:	Deposits at CBs and SBs by depositors:HOUSEHOLD	WON	Billion Won
EXCHWONVSUS	MARKET CLEARING:	Exchange rate	WON/US DOLLAR	Closing Rate
EXCHWONVSYEN	MARKET CLEARING:	Exchange rate	WON/JAPANYen(100Yen)	Closing Rate
M2ENDOF	MARKET CLEARING:	M2(End Of)	WON	Billion Won

Table 10

Notes: Table 10 is the legend of the 26 indicators used for High Frequency Monthly Forecast. The very left column represents the label that was used in Principal Component Analysis for abbreviation.(3rd Column is explanation of the abbreviation) 2nd column indicates which component of the Cob Web it belongs to. 4th and 5th column indicates the measure of each indicators.

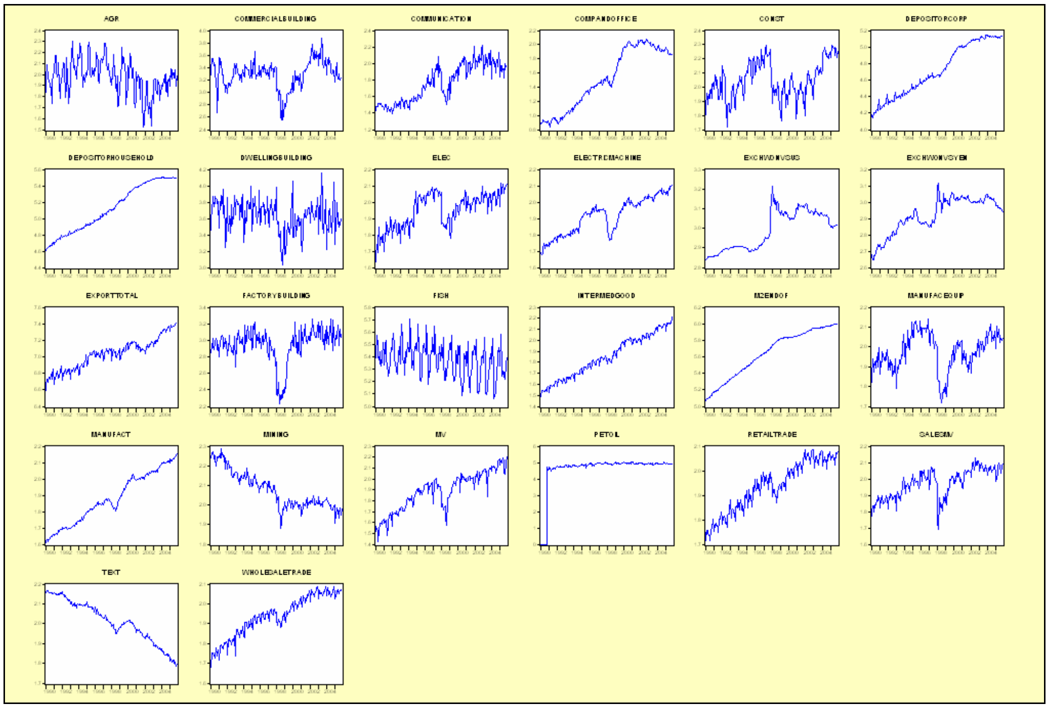


Figure 8

Notes: Figure 9 represents the graph of 26 indicators plotted. As shown in the figure, some show a very similar pattern of movement where it would not have been possible for to if it were not for the principal component analysis that was employed.

Principal Components

Included observations: 191

Correlation of AGR COMMERCIALBUILDING COMMUNICATION COMPANDOFFICE CONST DEPOSITORCORP DEPOSITORHOUSEHOLD DWELLINGBUILDING ELEC ELECTRICMACHINE EXCHWONVSUS E

A	B	C	D	E	F	G	H	I	J	K	L	M
	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Comp 8	Comp 9	Comp 10	Comp 11	Comp 12
Eigenvalue	16.45208	3.815140	1.627689	0.890448	0.721888	0.548819	0.525998	0.356739	0.278191	0.178581	0.124085	0.115216
Variance Prop.	0.632772	0.146736	0.062603	0.034248	0.027765	0.021108	0.020231	0.013721	0.010700	0.006868	0.004772	0.004431
Cumulative Prop.	0.632772	0.779509	0.842112	0.876360	0.904125	0.925233	0.945464	0.959185	0.969884	0.976753	0.981525	0.985957

Eigenvectors:

Variable	Vector 1	Vector 2	Vector 3	Vector 4	Vector 5	Vector 6	Vector 7	Vector 8	Vector 9	Vector 10	Vector 11	Vector 12
AGR	-0.070245	0.149407	-0.550054	0.066219	-0.031936	0.744829	-0.167102	0.186462	-0.034544	0.094814	-0.081855	-0.078429
COMMERCIALBUIL...	0.055227	0.366277	0.338219	-0.140202	0.035611	0.332889	-0.109548	-0.579396	-0.265908	-0.332418	-0.197921	-0.003206
COMMUNICATION	0.227474	0.033725	0.128694	-0.093980	-0.067195	0.069392	0.077737	0.084308	-0.337339	0.375007	0.176772	-0.413682
COMPANDOFFICE	0.236364	-0.073441	0.103911	-0.052175	-0.015328	0.094071	-0.096387	0.138189	-0.123535	0.049625	0.154279	-0.144414
CONST	0.111611	0.299343	-0.324468	0.288258	-0.304856	-0.040322	0.273039	-0.396718	0.142227	0.024427	0.482744	0.141741
DEPOSITORCORP	0.241261	-0.057428	0.088817	-0.004518	-0.032915	0.031803	-0.086633	-0.013509	0.112101	0.041788	0.002726	-0.135324
DEPOSITORHOUS...	0.240423	-0.080529	0.076155	-0.005126	-0.048010	0.051793	-0.063060	-0.052803	0.112080	0.047342	0.013127	-0.090520
DWELLINGBUILDING	-0.039559	0.346224	0.241506	-0.347121	0.202677	0.235287	0.613210	0.215186	0.344811	0.164789	-0.056877	0.035532
ELEC	0.200058	0.204539	-0.166899	-0.027942	0.067758	-0.201822	0.232710	0.118475	-0.319866	-0.050790	0.110241	0.093779
ELECTRICMACHINE	0.236064	0.108506	-0.020604	0.063883	-0.011166	-0.081164	-0.077916	0.065595	-0.046143	-0.035638	-0.117435	-0.164664
EXCHWONVSUS	0.188817	-0.278415	0.082527	-0.160114	0.046776	0.224205	0.099234	-0.052282	-0.090288	-0.128598	0.476789	0.041985
EXCHWONVSYEN	0.212516	-0.168054	-0.068343	-0.137755	0.203166	0.122216	0.016465	0.171693	-0.030356	-0.516933	0.137941	0.187981
EXPORTTOTAL	0.236090	0.016565	-0.087938	0.066770	-0.144349	-0.049970	0.137512	0.075464	0.213596	-0.115818	-0.120971	-0.103128
FACTORYBUILDING	0.058226	0.393926	0.239334	0.070468	0.136359	-0.008164	-0.536100	0.231290	0.406178	-0.029022	0.369447	0.123537
FISH	-0.082522	0.076176	-0.286621	-0.807546	-0.373389	-0.197801	-0.218335	-0.065349	0.087444	-0.003400	0.041567	-0.004419
INTERMEDGOOD	0.242696	-0.030265	0.024323	0.014724	-0.093455	0.018707	-0.015429	0.007293	0.179609	0.056026	-0.061838	-0.114935
M2ENDOF	0.240019	-0.101960	-0.035299	-0.017641	-0.020886	0.070440	0.037500	-0.009340	0.002282	-0.013679	-0.024249	0.059044
MANUFACEQUIP	0.090832	0.425951	-0.154362	0.083012	0.007392	-0.249081	0.071110	0.279296	-0.152848	-0.409136	-0.114247	-0.252203
MANUFACT	0.243047	-0.016452	0.040682	0.046858	-0.087003	0.012170	-0.073773	0.005504	0.142685	0.047355	-0.088743	-0.170625
MINING	-0.215576	0.196151	0.091052	0.025403	0.020829	-0.046653	-0.083943	-0.128552	-0.035267	0.086584	0.189733	-0.452400
MV	0.233722	0.083944	-0.089338	-0.000724	-0.023722	-0.039711	-0.113730	0.072236	-0.042797	0.064785	-0.284445	0.077136
RETAILTRADE	0.241186	0.008915	0.015938	-0.110153	-0.013740	-0.023507	-0.010485	-0.073400	-0.073329	0.086131	0.024289	0.013363
SALESMV	0.203247	0.228036	0.025558	0.006437	0.075070	-0.071604	-0.132490	-0.005088	-0.247936	0.440582	-0.113132	0.533662
PETOIL	0.126737	-0.031886	-0.382601	-0.115048	0.749380	-0.163584	-0.055104	-0.351568	0.170641	0.124389	-0.010515	-0.190566
TEXT	-0.228918	0.069785	-0.021043	-0.071184	0.206129	-0.038314	-0.070867	0.230919	-0.357304	0.030610	0.283514	0.006545
WHOLESALETRADE	0.242881	0.024516	-0.011979	-0.094884	0.029218	0.026374	-0.022602	0.019955	-0.018059	-0.034739	0.036614	0.117062

Table 11

Notes: Table 11 is the output of the principal component analysis done through the EVIEWs program. 26 indicators were analyzed and as it could be seen, the first 6 components a cumulative variance 0.925 which represents 92.5% of explanation of the variances. For each sample size, as explained in part 2 and part 3, principal component analysis was done. While the actual matrix above is 26 x 26, only 12 components were taken to show that the first 6 components are meaningful in describing the variance and its eigenvalue.

1 Step Ahead (High Frequency)								
Sample size	**Forecast value**	**Actual value**	**Difference* (actual -forecast**	**SE**	**2SE**	**2SE - Difference**	**Plus 2SE**	**Minus 2SE**
1990Q01- 2002Q02	5.206	5.209	0.003	0.009	0.017	0.014	5.223	5.188
1990Q01- 2002Q03	5.220	5.214	0.007	0.005	0.010	0.003	5.230	5.210
1990Q01- 2002Q04	5.230	5.212	0.018	0.006	0.012	-0.006	5.242	5.218
1990Q01- 2003Q01	5.215	5.213	0.002	0.005	0.009	0.008	5.224	5.205
1990Q01- 2003Q02	5.222	5.219	0.003	0.003	0.007	0.004	5.229	5.215
1990Q01- 2003Q03	5.225	5.231	0.006	0.003	0.006	0.000	5.231	5.218
1990Q01- 2003Q04	5.236	5.234	0.002	0.005	0.010	0.008	5.246	5.226
1990Q01- 2004Q01	5.240	5.237	0.003	0.003	0.007	0.004	5.246	5.233
1990Q01- 2004Q02	5.245	5.240	0.005	0.004	0.008	0.004	5.253	5.236
1990Q01- 2004Q03	5.241	5.244	0.003	0.004	0.007	0.004	5.249	5.234
**absolute value*								

1 Step Ahead (ARMA(1,1))								
Sample size	**Forecast value**	**Actual value**	**Difference* (actual -forecast**	**SE**	**2SE**	**2SE - Difference**	**Plus 2SE**	**Minus 2SE**
1990Q01- 2002Q02	5.207	5.209	0.002	0.013	0.025	0.023	5.232	5.182
1990Q01- 2002Q03	5.212	5.214	0.002	0.012	0.025	0.023	5.236	5.187
1990Q01- 2002Q04	5.217	5.212	0.005	0.012	0.024	0.020	5.241	5.193
1990Q01- 2003Q01	5.220	5.213	0.007	0.012	0.024	0.017	5.245	5.196
1990Q01- 2003Q02	5.218	5.219	0.002	0.012	0.024	0.022	5.242	5.194
1990Q01- 2003Q03	5.227	5.231	0.005	0.012	0.024	0.019	5.250	5.203
1990Q01- 2003Q04	5.228	5.234	0.006	0.012	0.024	0.017	5.252	5.204
1990Q01- 2004Q01	5.240	5.237	0.003	0.012	0.024	0.021	5.264	5.216
1990Q01- 2004Q02	5.241	5.240	0.000	0.012	0.023	0.023	5.264	5.217
1990Q01- 2004Q03	5.243	5.244	0.001	0.012	0.023	0.022	5.267	5.220
**absolute value*								

4 Step Ahead (High Frequency)								
Sample size	**Forecast value**	**Actual value**	**Difference* (actual -forecast**	**SE**	**2SE**	**2SE - Difference**	**Plus 2SE**	**Minus 2SE**
1990Q01- 2002Q02	5.204	5.213	0.009	0.010	0.020	0.011	5.224	5.184
1990Q01- 2002Q03	5.226	5.219	0.007	0.008	0.017	0.010	5.243	5.209
1990Q01- 2002Q04	5.239	5.231	0.008	0.009	0.019	0.011	5.258	5.220
1990Q01- 2003Q01	5.231	5.234	0.003	0.009	0.018	0.015	5.250	5.213
1990Q01- 2003Q02	5.233	5.237	0.004	0.005	0.010	0.006	5.243	5.223
1990Q01- 2003Q03	5.241	5.240	0.001	0.006	0.012	0.011	5.253	5.230
1990Q01- 2003Q04	5.249	5.244	0.005	0.009	0.018	0.013	5.267	5.231
1990Q01- 2004Q01	5.257	5.246	0.011	0.004	0.008	-0.003	5.266	5.249
1990Q01- 2004Q02	5.252	5.251	0.001	0.006	0.012	0.011	5.264	5.240
1990Q01- 2004Q03	5.258	5.259	0.001	0.007	0.015	0.014	5.273	5.243
**absolute value*								

*absolute value								
4 Step Ahead (ARMA(1,1))								
Sample size	**Forecast value**	**Actual value**	**Difference* (actual -forecast**	**SE**	**2SE**	**2SE - Difference**	**Plus 2SE**	**Minus 2SE**
1990Q01- 2002Q02	5.218	5.213	0.005	0.020	0.041	0.036	5.258	5.177
1990Q01- 2002Q03	5.223	5.219	0.003	0.020	0.040	0.037	5.263	5.182
1990Q01- 2002Q04	5.228	5.231	0.004	0.020	0.040	0.036	5.267	5.188
1990Q01- 2003Q01	5.230	5.234	0.004	0.020	0.039	0.035	5.269	5.190
1990Q01- 2003Q02	5.227	5.237	0.010	0.019	0.039	0.029	5.266	5.188
1990Q01- 2003Q03	5.229	5.240	0.011	0.019	0.038	0.028	5.268	5.191
1990Q01- 2003Q04	5.239	5.244	0.005	0.019	0.039	0.033	5.278	5.200
1990Q01- 2004Q01	5.250	5.246	0.004	0.019	0.038	0.035	5.288	5.211
1990Q01- 2004Q02	5.250	5.251	0.001	0.019	0.038	0.037	5.288	5.212
1990Q01- 2004Q03	5.253	5.259	0.006	0.017	0.034	0.028	5.287	5.220
*absolute value								

Table 12

Notes: Table 12 represents the result of the forecast done for both 1 step ahead forecast as well as 4 steps ahead forecast. For both ARMA(1,1) and HF model, the table above summarizes the forecast values with the standard error and its difference with the actual values.

1 step ahead							
95% confidence interval (2SE)							
	ARMA(1,1)		**HF**			ARMA(1,1)	HF
	Plus 2SE	Minus 2SE	Plus 2SE	Minus 2SE	Actual	Within the band?	Within the band?
1	5.23202	5.18165	5.22285	5.18841	5.20911	yes	yes
2	5.23629	5.18685	5.23025	5.21043	5.21362	yes	yes
3	5.24140	5.19250	5.24232	5.21806	5.21225	yes	no
4	5.24460	5.19607	5.22408	5.20546	5.21317	yes	yes
5	5.24167	5.19362	5.22875	5.21548	5.21942	yes	yes
6	5.25048	5.20287	5.23138	5.21847	5.23134	yes	yes
7	5.25161	5.20388	5.24620	5.22595	5.23415	yes	yes
8	5.26362	5.21622	5.24634	5.23321	5.23675	yes	yes
9	5.26391	5.21710	5.25307	5.23649	5.24024	yes	yes
10	5.26660	5.22030	5.24862	5.23411	5.24429	yes	yes
						100%	90%
4 step ahead							
95% confidence interval (2SE)							
	ARMA(1,1)		**HF**			ARMA(1,1)	HF
	Plus 2SE	Minus 2SE	Plus 2SE	Minus 2SE	Actual	Within the band?	Within the band?
1	5.25835	5.17707	5.22369	5.18380	5.21317	yes	yes
2	5.26269	5.18249	5.24321	5.20933	5.21942	yes	yes
3	5.26748	5.18814	5.25783	5.21993	5.23134	yes	yes
4	5.26893	5.19042	5.24962	5.21305	5.23415	yes	yes
5	5.26560	5.18802	5.24278	5.22276	5.23675	yes	yes
6	5.26795	5.19102	5.25315	5.22966	5.24024	yes	yes
7	5.27773	5.20004	5.26692	5.23146	5.24429	yes	yes
8	5.28834	5.21144	5.26578	5.24901	5.24595	yes	no
9	5.28836	5.21236	5.26426	5.24004	5.25121	yes	yes
10	5.28701	5.21969	5.27267	5.24315	5.25906	yes	yes
						100%	90%

Table 13

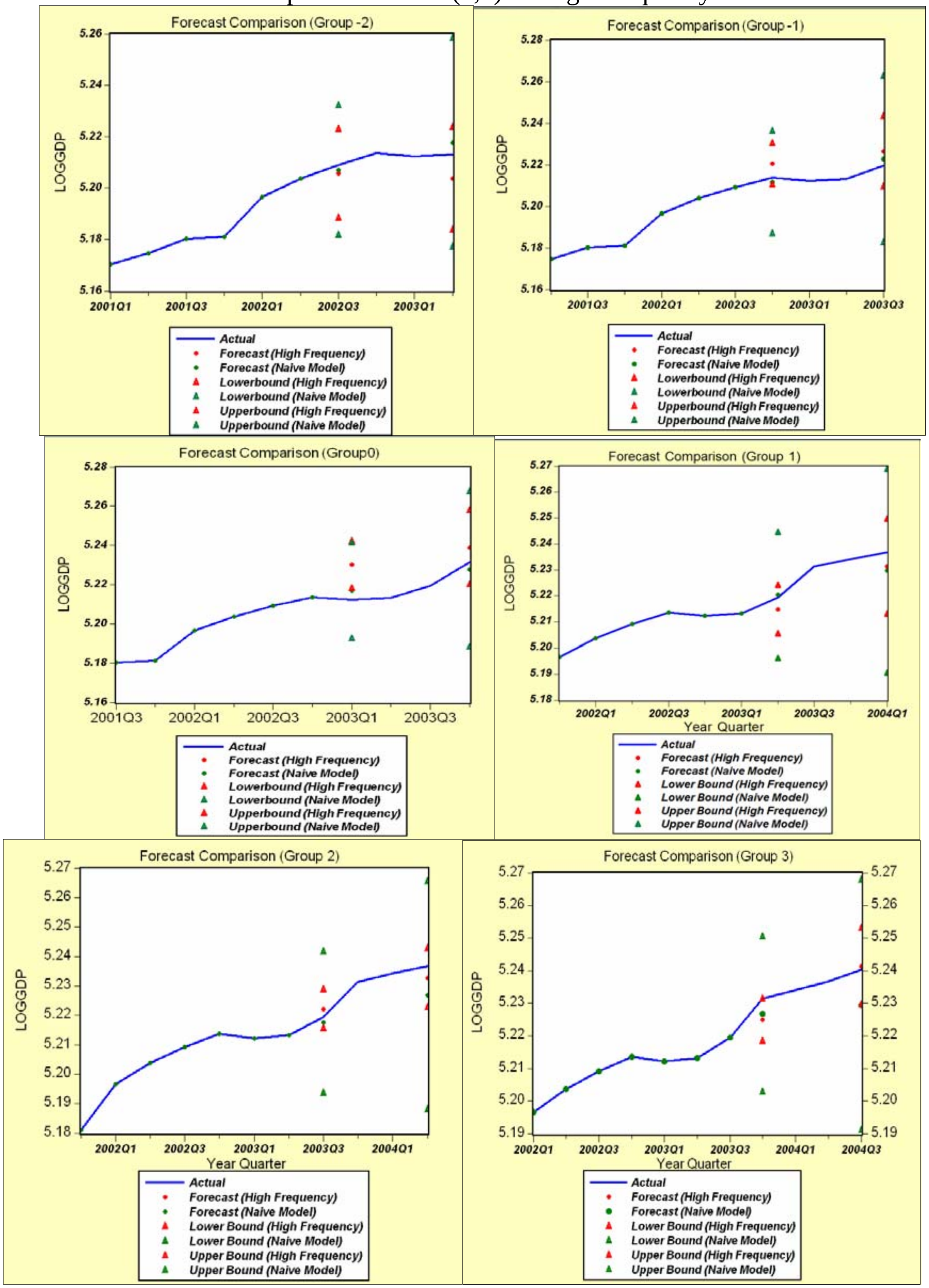
Notes: Table 13 represents the Confidence Interval test that was done for 95% interval. As summarized in the table, for 95% confidence interval it was stated in the result section that simple ARMA(1,1) model picks up every points while only 90% of the actual values are picked up by High Frequency Model.

1 step ahead							
50% confidence interval (0.68 * SE)							
	ARMA(1,1)		**HF**			ARMA(1,1)	HF
	Plus 0.68SE	Minus 0.68SE	Plus 0.68SE	Minus 0.68SE	Actual	Within the band?	Within the band?
1	5.21540	5.19827	5.21149	5.19978	5.20911	yes	yes
2	5.21998	5.20317	5.22371	5.21697	5.21362	yes	no
3	5.22526	5.20864	5.23432	5.22607	5.21225	yes	no
4	5.22859	5.21208	5.21794	5.21160	5.21317	yes	yes
5	5.22581	5.20948	5.22437	5.21986	5.21942	yes	no
6	5.23477	5.21858	5.22712	5.22273	5.23134	yes	no
7	5.23586	5.21963	5.23952	5.23263	5.23415	yes	yes
8	5.24798	5.23186	5.24201	5.23754	5.23675	yes	no
9	5.24847	5.23255	5.24760	5.24196	5.24024	yes	no
10	5.25132	5.23558	5.24383	5.23890	5.24429	yes	yes
						100%	40%
4 step ahead							
50% confidence interval (0.68 * SE)							
	ARMA(1,1)		**HF**			ARMA(1,1)	HF
	Plus 0.68SE	Minus 0.68SE	Plus 0.68SE	Minus 0.68SE	Actual	Within the band?	Within the band?
1	5.23153	5.20389	5.21053	5.19018	5.21317	yes	no
2	5.23623	5.20896	5.23203	5.21475	5.21942	yes	yes
3	5.24130	5.21432	5.24532	5.22600	5.23134	yes	yes
4	5.24302	5.21633	5.23756	5.21891	5.23415	yes	yes
5	5.24000	5.21362	5.23617	5.22596	5.23675	yes	no
6	5.24256	5.21640	5.24540	5.23342	5.24024	yes	yes
7	5.25209	5.22568	5.25521	5.23713	5.24429	yes	yes
8	5.26296	5.23682	5.26024	5.25169	5.24595	yes	no
9	5.26328	5.23744	5.25627	5.24391	5.25121	yes	yes
10	5.26480	5.24191	5.26293	5.24787	5.25906	yes	yes
						100%	70%

Table 14

Notes: Table 14 represents the 50% confidence interval test. As compared in the result section and represented as a graph, it shows that while ARMA(1,1) picks up 100% of the actual value, High Frequency picks up 40% and 70% respectively for 1 step ahead and 4 steps ahead forecast.

Comparison ARMA(1,1) vs High Frequency[32]



[32] For 10 sample size, group number starts from -2 and ends at 7. This was due to the fact that the sample size was originally 7 but expanded to 10 by shortening the sample size even more.

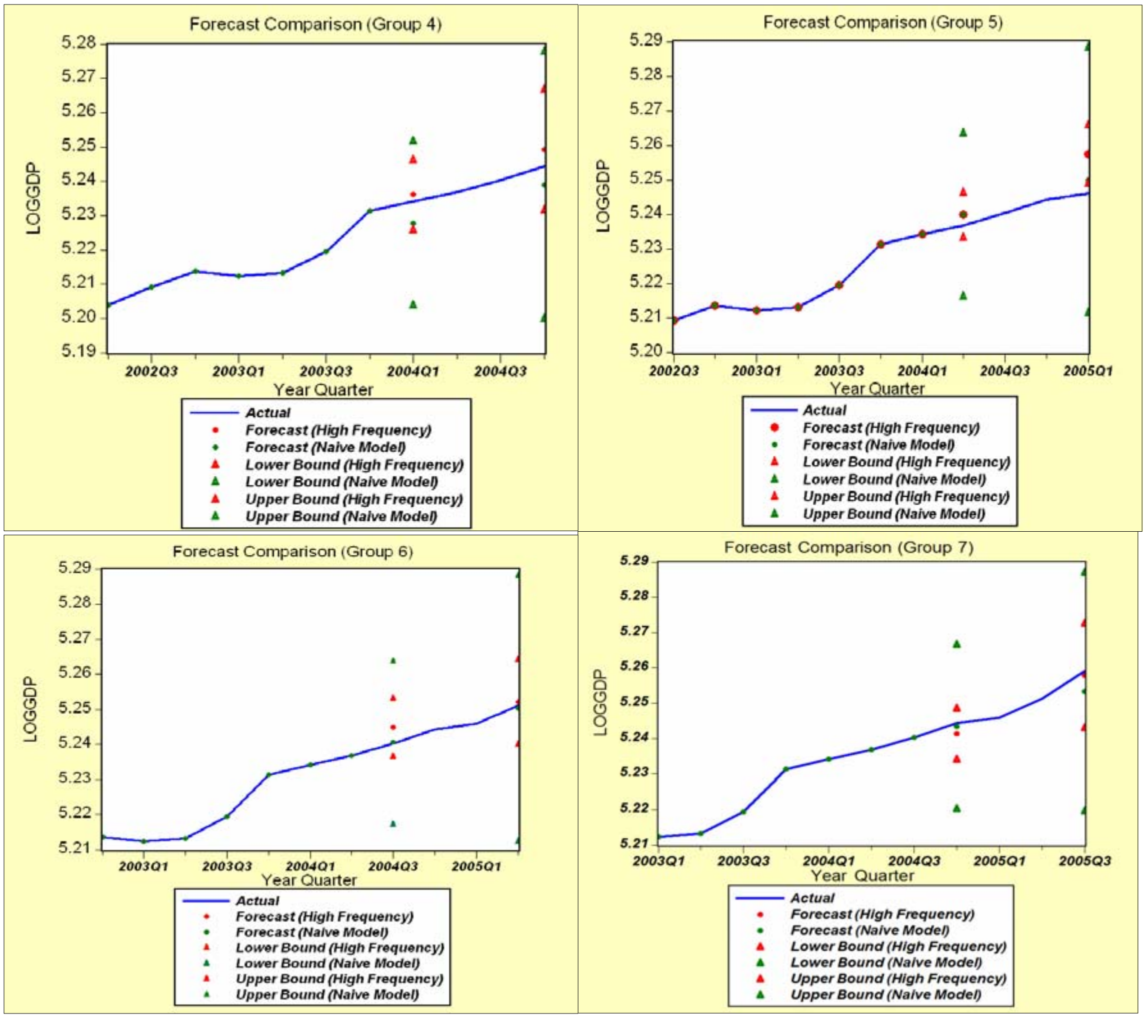


Figure 9

Notes: Figure 10 shows the graph plot of forecast comparison for the ARMA(1,1) model vs High Frequency Model for each sample space. There were 10 sample spaces and the line represents the actual value while the dots represent the value of each forecasts while the triangles represent their respective confidence interval (95% in this case)