A Study on Estimation and Prediction of Vector Time Series Model Using Financial Big Data (Interest Rates)

Jae-HyunKim[1], Chang-HoAn[*2]
[1]Department of Computer Engineering, Seokyeong Univ., SEOUL, Republic of Korea
[*2]Department of Financial Information Engineering, Seokyeong Univ., SEOUL, Republic of Korea
statsr@skuniv.ac.kr[1], choan@skuniv.ac.kr[*2]

Abstract: Due to the global economic downturn, the Korean economy continues to slump. Hereupon the Bank of Korea implemented a monetary policy of cutting the base rate to actively respond to the economic slowdown and low prices. Economists have been trying to predict and analyze interest rate hikes and cuts. Therefore, in this study, a prediction model was estimated and evaluated using vector autoregressive model with time series data of long- and short-term interest rates. The data used for this purpose were call rate (1 day), loan interest rate, and Treasury rate (3 years) between January 2002 and December 2019, which were extracted monthly from the Bank of Korea database and used as variables, and a vector autoregressive (VAR) model was used as a research model. The stationarity test of variables was confirmed by the ADF-unit root test. Bidirectional linear dependency relationship between variables was confirmed by the Granger causality test. For the model identification, AICC, SBC, and HQC statistics, which were the minimum information criteria, were used. The significance of the parameters was confirmed through t-tests, and the fitness of the estimated prediction model was confirmed by the significance test of the cross-correlation matrix and the multivariate Portmanteau test. As a result of predicting call rate, loan interest rate, and Treasury rate using the prediction model presented in this study, it is predicted that interest rates will continue to drop.

Keywords:database, VAR, granger causality test, ADF-unit root test, multivariate Portmanteau test.

1. Introduction

The global economy is on the rise with uncertainty due to the continued trade conflict between the United States and China and the increased possibility of no-deal Brexit in the UK. As COVID-19 spreads around the world and global economic recession is expected, central banks in major countries are turning to monetary policy to cut rates. The Fed cut interest rates in anticipation of slowing economic growth due to trade conflicts with China. Recently, the Bank of Korea also drastically cut the base rate as a way to alleviate the volatility of the financial market and reduce the impact on growth and prices. The base rate cut decided by the Bank of Korea is initially reflected in the very short-term interest rate (call rate), and subsequently affects the interest rates that we experience directly through credit line or mortgage loans in the market. On the positive side, it can be expected to boost the economy by reducing the debt and interest burdens of households and businesses and increasing consumption and investment capacity. However, it is likely to be a "double-edged sword", because it is likely to cause housing prices to rise and household debt to expand as the biggest risk on the negative side. Although interest rates have recently been lowered, market interest rates have been rising. Financial experts predict that bond interest rates will continue to rise. In the midst of the debate on interest rate cuts at home and abroad, the study of interest rate prediction is very necessary considering the current global and the domestic economy, as one of the monetary policies.

Some of the previous studies on interest rate prediction are as follows. Wright (2006) suggested that predictability improved when the nominal federal funds rate was added as an explanatory variable when estimating economic predictability using term spread [1]. Andrew, Monika, and Min (2006) suggested that short-term interest rate was better at predicting GDP growth rate than term spread [2-13]. Dewachter and Maes (2001) modeled the correlation between bonds by countries using a multifactor linear term structure model [3]. Kaminska, Meldrum, and Smith (2013) estimated the zero-coupon forward rates curve for the US, the UK, and Germany [4]. In-soo Kim and Dong-cheol Park (2012) analyzed the correlation between mortgage loans and macroeconomics (including interest rates) and asserted that mortgage loans could affect interest rates, leading to increase of financial costs and overdue [5]. Song-bae Kim and Jong-jin Kim (2015) analyzed that decrease in interest rates led to increase in mortgage loans [6-14]. In this study, we build a vector time series prediction model using time-series data of call rate (1 day), loan interest rates, and Treasury rate (3 years).

2. Research Model

2.1 Vector autoregressive model

In this study, the vector autoregressive model, which is a vector time series model, was used to predict call rate (1 day), loan interest rate, and Treasury rate (3 years)). As the vector autoregressive (VAR) model does not

need to distinguish dependent from independent variables, and uses a small number of variables, it has been widely used for analysis and prediction of dynamic relationships between economic time series variables [7-8].

The VAR model is a model that analyzes the real economy by maximizing the information provided by economic time series actually observed, not by setting hypotheses based on any economic theory. The VAR(p) model, which is the VAR model of order p on $Z_t = (Z_{1t}, Z_{2t}, \cdots, Z_{lt})'$ that is consisted of l multivariate time series, is as follows.

$$Z_t = \delta_0 + \Phi_1 Z_{t-1} + \cdots + \Phi_p Z_{t-p} + \varepsilon_t$$
$$= \delta_0 + \sum_{i=1}^{p} \Phi_i Z_{t-i} + \varepsilon_t \qquad (1)$$

where δ_0 is an $(l \times 1)$ constant vector, Φ_i is an $(l \times l)$ parameter matrix, and ε_t is a white noise vector that satisfies $E(\varepsilon_t) = 0, E(\varepsilon_t \varepsilon_t') = \sum$, and $E(\varepsilon_t \varepsilon_s') = 0, t \neq s$.

2.2 Unit root test

The VAR model is basically modeling the dynamic interrelationship between stationary time series variables. A stationary time series means the case where the average value of the time series is constant with the passage of time, and a nonstationary time series means the case where the average value is not constant with the passage of time, and in this case, the unit root exists. When a nonstationary time series with unit roots is analyzed, there is a possibility of spurious regression that appears to be correlated between variables even though there is no correlation. In order to avoid such spurious regression, it is necessary to perform the analysis after converting it into a stationary time series by performing differencing. ADF (Augmented Dickey-Fuller) unit root test, one of the methods to confirm the existence of the unit root, is as follows [9].

$$\nabla Z_t = c_t + \phi_0 Z_{t-1} + \sum_{i=1}^{p-1} \phi_i \nabla Z_{t-i} + \varepsilon_t \qquad (2)$$

Where, if $\phi_0 = \phi - 1$, the test for null hypothesis $H_0: \phi = 1$ is same as test for $H_0: \phi = 0$.

2.3 Granger causality test

The Granger causality test is a test to determine whether one variable can be used as a predictor (independent variable) in predicting another variable. An autoregressive model for the stationary time series model is as follows [10].

$$Z_t = \sum_{i=1}^{m} \alpha_i Z_{t-i} + \sum_{i=1}^{m} \beta_i X_{t-i} + \varepsilon_{1t}$$

$$X_t = \sum_{i=1}^{m} \gamma_i X_{t-i} + \sum_{i=1}^{m} \delta_i Z_{t-i} + \varepsilon_{2t} \qquad (3)$$

Where, the error terms $\varepsilon_{1t}, \varepsilon_{2t}$ are mutually independent and equal variances are assumed.

The hypotheses for (Equation 3) are

$$H_{10}: \beta_i = 0 (i = 1,2, \cdots, m) H_{11}: \beta_i \neq 0 (1,2, \cdots, m)$$

$$H_{20}: \delta_i = 0 (i = 1,2, \cdots, m) H_{21}: \delta_i \neq 0 (1,2, \cdots, m) \qquad (4)$$

If both hypotheses H_{10} and H_{20} are rejected, causal relationships exist in both directions, so the VAR model can be used as a predictive model.

2.4 Model identification

There are several statistics indicating the relative goodness of fit when selecting one model that fits best to the given data among several vector time series models. The information criteria used in this study are corrected Akaike information criterion (AICC), Schwarz Bayessian criterion (SBC) and Hannan-Quinn criterion (HQC) as follows [11-12].

$$AICC = \log(|\widehat{\textstyle\sum_\varepsilon}|) + 2m/(n - m/l)$$
$$SBC = \log(|\widehat{\textstyle\sum_\varepsilon}|) + m \log(n)/n$$
$$HQC = \log(|\widehat{\textstyle\sum_\varepsilon}|) + 2m \log(\log(n))/n \qquad (5)$$

Where, m is the number of parameters in the model, l is the number of univariate time series that make up the multivariate time series, n is the size of the time series data used to estimate the parameters, and $\widehat{\sum_\varepsilon}$ is the maximum likielihood estimate of $\sum_\varepsilon$, the covariance matrix of the multivariate white noise process.

3. Results

3.1 Data conversion and stationary test

Since three time series data used in this study, the trends of call rate (1 day), loan interest rate, and Treasury rate (3 years), appear to have characteristics of nonstationary time series as shown in Figure 1, it is necessary to convert them to the stationary time series.

Figure 1. Time series trends of raw data

In addition, as confirmed through the ADF-unit root test as shown in Table 1, p-values of Tau statistics of all variables are greater than the significance level $\alpha = 0.05$, and therefore the null hypotheses $H_0: \phi = 1$ are adopted, so that they are nonstationary time series with the unit roots.

Table 1. Unit root test of raw variable

Augmented Dickey-Fuller Unit Root Tests			
Variable	**TYPE**	**Tau**	**Pr< Tau**
CR	Zero Mean	-1.25	0.1938
	Single Mean	-1.65	0.4539
	Trend	-3.09	0.1124
LR	Zero Mean	-1.70	0.0847
	Single Mean	-1.43	0.5668
	Trend	-2.94	0.1521
TR	Zero Mean	-1.55	0.1129
	Single Mean	-1.25	0.6550
	Trend	-3.00	0.1339

As the result of the ADF-unit root test after the first order differencing, it was confirmed that p-values of the Tau statistics of all variables were smaller than the significance level $\alpha = 0.05$, so that the null hypothesis $H_0: \phi = 1$ was rejected, which mean that it was a stationary time series that no longer had a unit root (Table 2).

Table 2. Unit root test after the first order differencing

Augmented Dickey-Fuller Unit Root Tests			
Variable	**TYPE**	**Tau**	**Pr< Tau**
∇CR	Zero Mean	-5.68	<.0001
	Single Mean	-5.71	<.0001
	Trend	-5.70	<.0001
∇LR	Zero Mean	-9.11	<.0001
	Single Mean	-9.26	<.0001
	Trend	-9.25	<.0001
∇TR	Zero Mean	-8.90	<.0001
	Single Mean	-9.01	<.0001
	Trend	-9.00	<.0001

3.2 Granger causality test

In Table 3, the null hypothesis of Test 1 is $H_{10}: \nabla CR \nleftarrow \nabla LR\ \nabla TR$, the null hypothesis of Test 2 is $H_{20}: \nabla LR \nleftarrow \nabla CR\ \nabla TR$, and the null hypothesis of Test 3 is $H_{30}: \nabla TR \nleftarrow \nabla LR\ \nabla CR$. They are all rejected in the

bidirectional linear dependence test for the call rate, loan interest rate, and Treasury rate variables that are converted to a stationary time series (after first-order differencing). Therefore, there are two-way dependencies that are influenced by the past values of call rate, loan interest rate, and Treasury rate, respectively, and the past values of the other two variables.

Table 3. Granger causality test

Granger-Causality Wald Test			
Test	DF	Chi-Square	Pr<ChiSq
1	4	43.68	<.0001
hypothesis	$H_{10}: \nabla CR \nleftarrow \nabla LR \ \nabla TR$ $H_{11}: \nabla CR \leftarrow \nabla LR \ \nabla TR$		
2	4	16.27	0.0027
hypothesis	$H_{20}: \nabla LR \nleftarrow \nabla CR \ \nabla TR$ $H_{21}: \nabla LR \leftarrow \nabla CR \ \nabla TR$		
3	4	84.61	<.0001
hypothesis	$H_{30}: \nabla TR \nleftarrow \nabla LR \ \nabla CR$ $H_{31}: \nabla TR \leftarrow \nabla LR \ \nabla CR$		

3.3 Prediction model estimation and diagnosis

As the time just before truncation of the sample partial autocorrelation matrix (SPAM) of the stationary time series toward 0 was r=3, and that of the sample cross-correlation matrix (SCCM) toward 0 was s=2, the model $0 \leq p \leq r = 3, 0 \leq q \leq s = 2$ was applied. The statistics of AICC, SBC, and HQC, which are the minimum information criteria including these two models, are shown in Table 4. The smallest value of AICC statistic is -12.95672 in AR(2), the smallest value of SBC statistic is -12.62122 in AR(2), and the smallest value of HQC statistic is -12.81868 in AR(2). Therefore, the vector autoregressive model was selected as the VAR(2) model.

Table 4. Statistics of minimum information criteria for model identification

Minimum Information Criterion Based on AICC			
Lag	MA0	MA1	MA2
AR0	-11.77328	-12.32436	-12.53714
AR1	-12.80569	-12.81352	-12.83819
AR2	-12.95672	-12.91722	-12.93704
AR3	-12.94591	-12.92716	-12.93259
Minimum Information Criterion Based on SBC			
Lag	MA0	MA1	MA2
AR0	-11.72638	-12.13654	-12.20626
AR1	-12.61908	-12.61152	-12.48393
AR2	-12.62122	-12.49737	-12.31111
AR3	-12.49575	-12.40773	-12.22688
Minimum Information Criterion Based on HQC			
Lag	MA0	MA1	MA2
AR0	-11.75441	-12.26358	-12.40566
AR1	-12.73155	-12.80092	-12.76878

AR2	-12.81868	-12.78223	-12.68143
AR3	-12.77876	-12.77804	-12.68265

And since, in the t-test for significance of the constant term, p-value of ∇CR was 0.1587, p-value of ∇LR is 0.3911, and p-value of ∇TR is 0.2984, which were not significant at the significance level $\alpha = 0.05$, it was identified as a VAR(2) model without a constant term. The estimating equation of the prediction model is as follows.

$$
\begin{pmatrix} \nabla CR_{1t} \\ \nabla LR_{2t} \\ \nabla TR_{3t} \end{pmatrix}
= \begin{pmatrix} 0.66132 & 0.28955 & -0.60533 \\ 0.33083 & 0.52172 & -0.38410 \\ 0.02937 & -0.05807 & 0.47521 \end{pmatrix} \begin{pmatrix} \nabla CR_{1,t-1} \\ \nabla LR_{2,t-1} \\ \nabla TR_{3t,t-1} \end{pmatrix} \quad (6)
$$
$$
+ \begin{pmatrix} 0.49024 & 0.36864 & -0.05428 \\ -0.18403 & 0.40321 & -0.22430 \\ 0.27442 & 0.09125 & -0.31676 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix}
$$

The results of testing the goodness of fit of the model are as follows. First, after testing the significance of the cross-correlation matrix of the residuals as shown in Table 5, it was found that the autocorrelation and cross-correlation did not exist for each variable at time lag 1 or higher.

Table 5. Significance test of cross correlation matrix

Schematic Representation of Cross Correlations of Residuals							
Variable /Lag	0	1	2	…	10	11	12
∇CR_{1t}	++	…	…		…	…	…
∇LR_{2t}	+	…	…	⋮	…	…	…
∇TR_{3t}	++	…	…		…	…	…
	+						
	++						
	+						

As shown in Table 6, the multivariate Portmanteau test result is $Q(12) = 86.48$ when $K = 12$, and p-value is 0.5106. Since the null hypothesis cannot be rejected at the significance level of 5%, the residual vector appears not to have autocorrelation and cross-correlation.

Table 6. Multivariate Portmanteau test

Portmanteau Test for Cross Correlations of Residuals		
Up To Lag	Chi-Square	Pr>ChiSq
3	18.11	0.0847
4	20.72	0.1255
5	34.17	0.2183
6	49.04	0.2927
7	51.43	0.2075
8	57.65	0.3021
9	61.67	0.4033
10	69.24	0.4352
11	81.42	0.4534
12	86.48	0.5106

The results of the goodness of fit test for each univariate time series model are shown in (Table 7), In the results, the p-values are 0.0001, and the univariate models are all significant at the significance level of 5%. However, in the case of loan interest rate (∇LR_{2t}),the explanatory power of the model is low as $R^2 = 0.1413$.

Table 7. Goodness of fit of the univariate model

Univariate Model ANOVA Diagnostics				
Variable	R-Square	Standard Deviation	F-Value	Pr> F
∇CR_{1t}	0.5227	0.09165	37.59	<.0001
∇LR_{2t}	0.1413	0.17434	5.65	<.0001
∇TR_{3t}	0.4976	0.10379	5.65	<.0001

3.4 Prediction results

The results predicted by the prediction model (Equation 6) are shown in (Figure 2), (Figure 3), and (Figure 4). (Figure 2) is the prediction result of call rate, (Figure 3) is the prediction result of loan interest rate, and (Figure 4) is the prediction result of Treasury rate. They all are predicted to continue to drop over time.
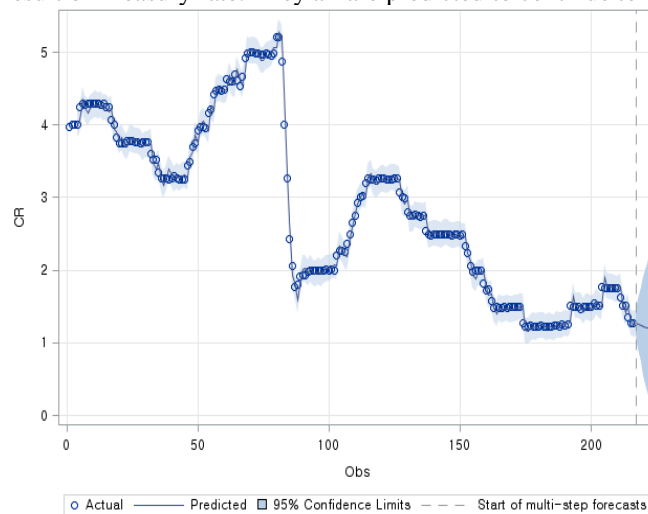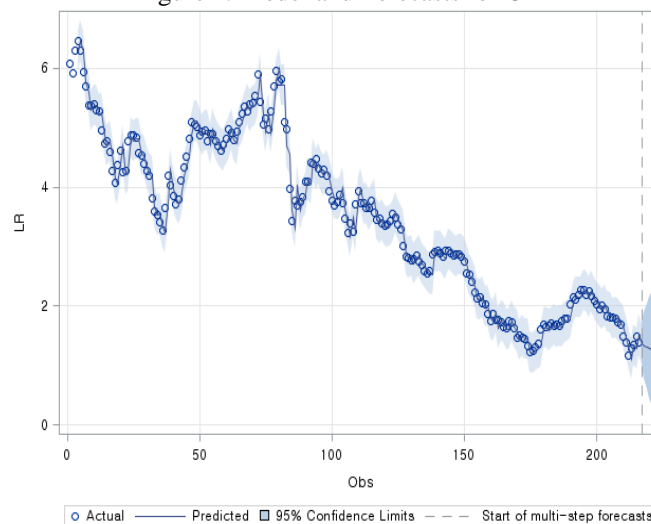

Figure 2. Model and Forecasts for CR

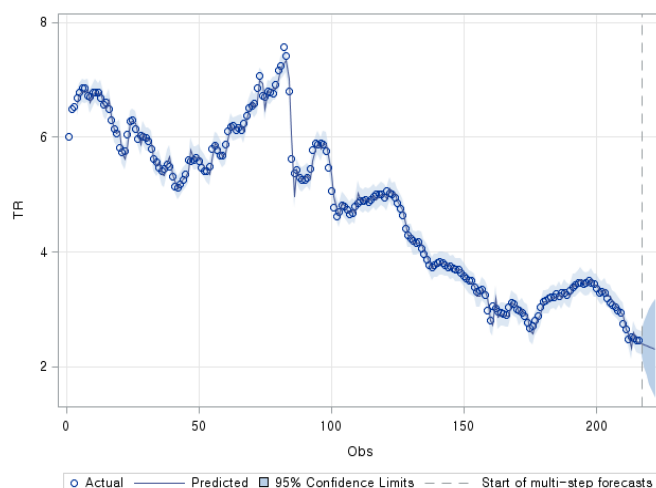
Figure 3. Model and Forecasts for LR

Jae-HyunKim[1], Chang-HoAn[*2]

Figure 4. Model and Forecasts for TR

4. Conclusions

The global economy has recorded the lowest growth rate since the global financial crisis due to sluggish global trade and manufacturing, and slowing investment after prolonged US-China trade disputes. As a result, central banks in major countries have turned to financial easing to respond to the global economic downturn, and the easing is expected to continue next year. The Bank of Korea is also expected to cut the base rate additionally in order to respond to the impact of low prices and low growth. In this study, empirical analysis was conducted using vector time series models for various prospects. The main analysis results are summarized as follows.

Three nonstationary time series variables used in the study were stabilized through time differencing, and as the result of the ADF-unit root test, p-value of the Tau statistic was found to be significant at the significance level of 5%. As the result of the Grandeur causality test, there was a causal relationship between variables at the significance level of 5%, and the model based on the minimum information criteria was identified as VAR(2) model. As the result of the significance test of cross-correlation matrix and the multivariate Portmanteau test to test goodness of fit of the VAR(2) prediction model, autocorrelation and cross-correlation of the residuals did not exist between time lags 1 and up to 12. Using this model, it was predicted that call rate, loan interest rate, and Treasury rate would continue to fall.

Considering the results of this study and the trend of spreading the base rate cut globally, the downturn of the domestic economy is expected to be faster than that of the global economy. To overcome the economic slowdown, the monetary policy to lower interest rates may have economic stimulus effects such as reducing households and corporate debt interest burdens, expanding consumption or investment, and rising consumer prices, but it is highly likely to cause housing prices to rise and household debts to expand. Therefore, the government and financial institutions should use macro-prudential policies to prepare for the possibility of increasing household debt and increasing demand for funds.

5. Acknowledgment

References

1. Wright, J. H. (2006). The Yield Cueve and Predicting Recessions. Board of Governors of the Federal Reserve System.
2. Andrew, A., Monika P., and Min, W.(2006). What Does the Yield Curve Tell us About GDP Growth?.*Journal of Econometrics,* 131, 359-403.
3. Dewachter, H., and Maes, H.(2001). An Affine Model for International Bond Market.*CES Discusion Paper,* No. 01.06.
4. Kaminska, I., Meldrum, A., and Smith, J.(2013). A Global Model of International Yield Curves: No-Arbitrage Term Structure Approach.*International Journal of Finance & Economics,* 18-4, 309-338.

5. In-Su Kim, Dong-Chul Park. (2012). A Study on Relationship between Housing Finance and Macroeconomic Variables using the VAR Model.*Journal of the Korean Regional Economics,* 22, 3-18.

6. Song-Bae Kim, Jong-Jin Kim. (2015). A Study on the Influence of Macroeconomy Variable upon the Mortgage Loan - Focused on Kyeongnam Province. *Journal of The Residential Environment Institute of Korea,* 13(04), 77-88.

7. Sims, C.A. (1980).acroeconomics and reality.*Econometrica,* 1-48.

8. Stock, J.H. and M.W. Watson. (2001). Vector autoregressions.*Journal of Economic Perspectives,* 15, 101-115.

9. Dickey, D.A., and Fuller, W.A.(1979). Distribution of the estimation for autoregressive time series with with a unit root.*Journal of the American Statistical Association,* 74, 424-431.

10. C.W.J. Granger. (1980). Testing for Causality, a Personal Viewpoint.*Journal of Economic Dynamics and Control,* 2, 239-352.

11. Akaike, H.(1974). A new look at the statistical model identification.*IEEE Trans Automatic Control,* AC-19, 716-723.

12. Schwartz, G.(1978). Estimating the dimension of a model.*Annal of Statistics,* 6, 461-464.

13. Maluleke, W., Mokwena, R. J., &Olofinbiyi, S. A. (2019). An Evaluative Study On Criminalistics: Stock Theft Scenes. International Journal of Business and Management Studies, 11(1), 101-138.

14. Manamela, M. G., &Molapo, (2019). K. K. The Green Revolution And Organic Farming Context: Effects And Public Health Disparities In Africa. International Journal of Business and Management Studies, 11(2), 16-31.