Volume E-3, No.7 (11)

Comparison of ARIMA and ARIMA/GARCH Models in EVN Traffic Prediction

Tran Quang Thanh, Trinh Quang Khai
Sichuan Provincial Key Laboratory of Information Coding and Transmission,
Southwest Jiaotong University, Chengdu 610031, China
Telecommunication Department, Faculty of Electrical and Electronic Engineering,
University of Transport and Communications, Hanoi, Vietnam
Email: khaitqvn@utc.edu.vn

***Abstract* - This paper focuses on building statistical models to capture and forecast the traffic of mobile communication network in Vietnam. Following Box-Jenkins method, a multiplicative seasonal ARIMA model is constructed to represent the mean component using the past values of traffic, a GARCH model is then incorporated to represent its volatility. The traffic is collected from EVN Telecom mobile communication network. The numerical result comparisons show that the multiplicative seasonal ARIMA/GARCH model built in this paper gives a better estimate when dealing with volatility clustering in the data series. However, in short-term prediction where the volatility has an insignificant influence, the achieved ARIMA model also can be considered as a good model to capture well the characteristics of EVN traffic series and gives reasonable forecasting results.**

Keywords - Traffic prediction, ARIMA, GARCH, Multiplicative Seasonal ARIMA/GARCH, EViews.

I. INTRODUCTION

Wireless traffic prediction has been an interest for years and become more and more important in the current context of international telecommunications. In Vietnam, the explosive development of mobile communications and internet has made a violent competition between service providers. Thus, there exists an increasing demand on a better network management. Traffic prediction is thus a key factor for that.

With the idea of considering traffic data stream as a predictable time series, statistical procedure can be applied to describe the characteristics and forecast that traffic data. Analysis on wireless network traffic shows that the data series normally contain seasonal component and meet volatility problem [1],[2]. Seasonal component can be processed by a multiplicative seasonal ARIMA model. While a GARCH model can be used to deal with the problem of volatility in data series.

Statistical procedure has been used in developing forecasting models that have been applied to many different areas such as seasonal ARIMA in wireless traffic modeling and prediction [1][2][3], tourism forecasting [4], air pollution analysis [5], or ARCH in load forecasting [6] and seasonal GARCH in internet traffic [7]. Those analyses present many successful applications of ARIMA in forecasting time series data. However, ARIMA can only help presenting the conditional mean of the series. With the implicit assumption of homoskedasticity, GARCH is absolutely efficient in investigating the volatility characteristics of time series. Therefore, the combination of ARIMA and GARCH is a good choice to give a better result in capturing and forecasting time series such as wireless traffic data [8 - 10], wind speed data [11], crude oil prices data [12], daily equity prices [13], or inflation data [14].

In this paper, the combination of ARIMA and GARCH is applied to mobile traffic in the condition of Vietnam, which has never been discussed before. Also, the comparison is made to evaluate the performance of the achieved ARIMA and ARIMA/GARCH models, in term of short-term prediction. This kind of time series data analysis is done using Eviews, which can be found in references [15 - 17]. The paper is organized as follows: Section II introduces and analyses EVN traffic to present the characteristics of wireless traffic in Vietnam. Section III proposes to build a multiplicative seasonal

ARIMA/GARCH model to fit and forecast EVN traffic. In section IV, the experiment results and discussions are presented. And finally, conclusions are given in section V.

II. EVN TRAFFIC

EVN Telecom is the international transaction name of Electricity Telecommunications Company which is an independent accounting member that belongs to Vietnam Electricity Corporations. EVN Telecom has been offering CDMA mobile communication service E-Mobile since 2006 March, with the prefix number of 096. Then it has official started offering 3G services since June 9, 2010 with high quality and continuously increasing number of subscribers. EVN Telecom is now merged into Viettel Telecom which is one of the biggest mobile communication providers in Vietnam. Data of mobile traffic used in this study is the amount of Erlang exchanged between MSC Danang and BSC 1462. The number of E1s between the two points is 112, number of channel is 3360 and allowed Erlang is 3337.08. 636 hourly data trace is used to model and forecast.

Figure 1 shows the statistical diagram of mobile communication traffic (Erlang) between MSCDN and BSC 1462 within 27 days, from Saturday, September 5th to Thursday, October 1st, 2009. This traffic stream is in hourly form, from 0:00 a.m. to 11:00 p.m. every day. We have totally 636 values, last from 0:00 a.m. of September 5th to 11:00 a.m. of October 1st, 2009.

Within one day (24 hours), the traffic varies according to several periods of time, as shown in Fig. 2. The traffic is low from 0:00 to 4:00 a.m., which is the sleeping time. The traffic roughly increases from 5:00 a.m. to 7:00 a.m., which is the time of waking up and starting a working day. From 8:00 a.m. to 11:00 a.m., which is working or studying time, the traffic keeps stable with a little fluctuation. After that, the traffic gradually decreases at the relax time ranging from 12:00 to 14:00. In the afternoon working time which starting from 14:00, the traffic again gradually increases until 17:00, which is the time off work or study. It is followed by the period of time from 18:00 and 20:00 (rush hour) that the traffic reaches the top of the day. This is the time for going home, parties, and relaxing after a stressful working or studying day. The traffic trace gradually decreases after that until sleeping time 23:00.

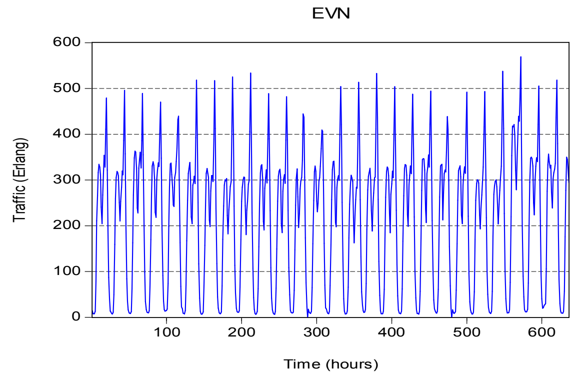


Figure 1. 27-day hourly traffic trace

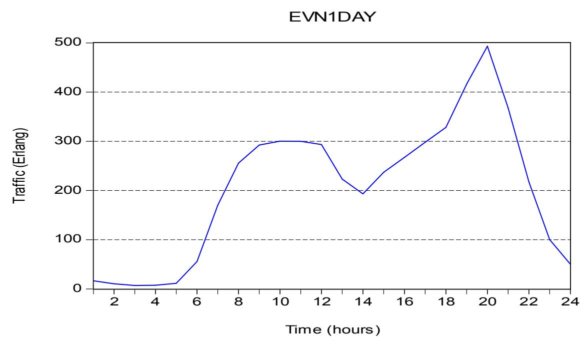


Figure 2. Hourly traffic trace of September 26th, 2009

It can be seen that, the characteristic of mobile traffic depends so much on the daily habit of people in every area in every country. Also, it depends on the way of charging by the service providers. This can be proved by a brief comparison between the traffic of China Mobile in Tianjin and that of EVN in Danang. The research in [1] shows that the traffic of China Mobile in Tianjin reaches the top of the day in the morning, inversely to that of EVN in Danang which reaches the top in the evening. The difference is due to the habit of Danang people, who prefers to make phone calls more in the evening. For mobile communication networks in Vietnam, the call charging is only cheaper from 10p.m. to 6a.m.

III. PROPOSE TO BUILD A MULTIPLICATIVE SEASONAL ARIMA/GARCH MODEL

The detail explanations of a multiplicative seasonal ARIMA model and a GARCH model can be found in references [1][14][18]. Below is the briefly description of a multiplicative seasonal ARIMA model which is derived from [1]:

$$\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D X_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (1)$$

or

$$W_t = \nabla^d\nabla_s^D X_t \quad (2)$$

where,

$$W_t = \phi_p^{-1}(B)\Phi_P^{-1}(B^s)\theta_q(B)\Theta_Q(B^s)a_t \quad (3)$$

To build a multiplicative seasonal ARIMA/GARCH model, we first construct a multiplicative seasonal ARIMA to present the mean component using the past values of the EVN traffic, then incorporate a GARCH model to represent its volatility. The whole progress is described in the flowchart in Fig. 3 and expressed step by step as follow:

Step 1: Using spectrum analysis to determine the period s of the traffic trace

This step is very important to give a consideration to a seasonal ARIMA model. If ***s*** is found, then we can make a decision of a multiplicative seasonal ARIMA which is in the form of $(p,d,q)\mathrm{x}(P,D,Q)_s$.

Step 2: Identification of stationary, determine d and D

The second step is also very important in fitting an ARIMA model. It is the determination of the order of differencing needed to make the series stationary.

Step 3: Model identification, determining all the orders

Propose to begin with candidate parameter sets that have small (*p, q*), (*P, Q*) values such as 0, 1, or 2 but where *p, P* and *q, Q* should not be 0 simultaneously in one set. Then, we can select the best *(p, q), (P, Q)* combination according to the known model identification such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion).

Step 4: Model estimation

Estimating all the parameters using approximate maximum likelihood parameter estimation methods, so that we obtain:

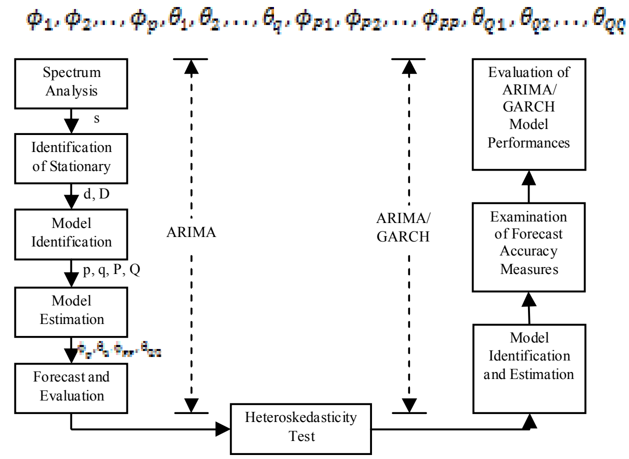


Figure 3. Flowchart of building ARIMA/GARCH process

Step 5: Forecast and evaluation

We use the fitted multiplicative seasonal ARIMA models obtained from (3) to forecast the series.

Step 6: Heteroskedasticity test

The existence of heteroscedasticity in hourly EVN traffic series must be examined before starting to estimate the GARCH model.

Step 7: Model identification and estimation for multiplicative seasonal ARIMA/GARCH model

We now verify the adequacy of AR and MA terms of the mean equation by implementing the correlogram Q-test, Jarque Bera test and ARCH test on the stationary series achieved from Step 2.

The result of no serial correlation under the correlogram Q-test will indicate that we can proceed with the estimation of the conditional variance for the errors using GARCH. We limit the order of GARCH(*p, q*) to 4, that is we use different orders of *p, q* = 0, 1, 2, 3 and 4, since GARCH is used for short-term forecasting. Incorporating the stationary series achieved from step 2 and the mean equation with AR and MA terms achieved from step 3, we estimate a GARCH model by finding a significant order combination under a specific error distribution (p-values should all be less than 0.10 level of significance and coefficient of the variance equation should all be positive).

Step 8: Examination of forecast accuracy measures

Static forecasting on the model is performed to show measures of forecast accuracy over the estimation period. The model with the smallest measure of forecast error will be chosen as the one with the most accurate fit of the time series model. Then, some more tests will be performed, such as correlogram of standardized residuals squared which consists of autocorrelation and partial auto-correlation, test for presenting of conditional hetero-scedasticity in the data with ARCH-LM test on the residuals, standardized residuals.

Step 9: Evaluation of multiplicative seasonal ARIMA/GARCH model performances

The final step is to evaluate the forecast performances by our achieved multiplicative seasonal ARIMA/GARCH model. The evaluation includes the information criterion, i.e. AIC and SIC values in the estimation stage, and forecast performances in the forecasting stage.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

1) Building a multiplicative seasonal ARIMA model

The period s of the traffic trace can be determined using spectrum analysis in Matlab. In Fig. 4, there is a peak when the Normalized Frequency is about (rad/sample), which means the frequency is about 0.042. At this main frequency, we can get the periodicity 1/0.042 = 24. This result presents the periodicity of 24 hours or one day and so $s = 24$.

In the next step, correlogram is used to clarify whether the series needs to be differenced or not. One-time non-seasonal difference and then one-time non-seasonal together with one-time seasonal differences are performed on the series which show that it needs to take the logarithm transformation (*EVNLOG*) to become variance stationary as shown in Fig. 5. The ACFs is in a sinusoidal pattern form with the peak at lag 24. It implies that the series may be needed to take the 24-period seasonal difference (*EVNLOGd0D1*) to achieve the stationary as shown in Fig. 6. Also, the ADF test for *EVNLOGd0D1* is shown in Fig. 7.

The unit root test result in Fig. 7 shows that the series now becomes stationary so we can further search the best ARIMA model. Refer to the equation described in (2) we have:

$$W_t = \nabla^1 \nabla^1_{24} \ln X_t \tag{4}$$

Where X_t is our series *EVN*, and W_t is *EVNLOGd0D1*.

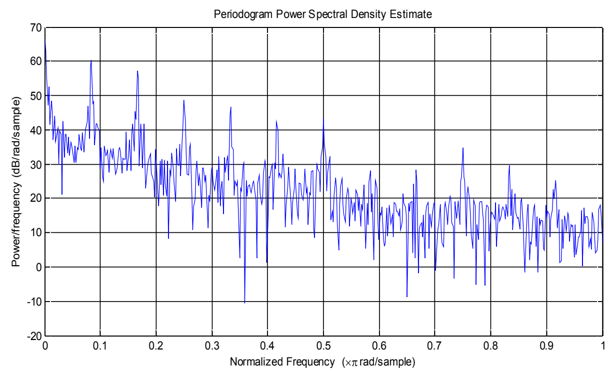
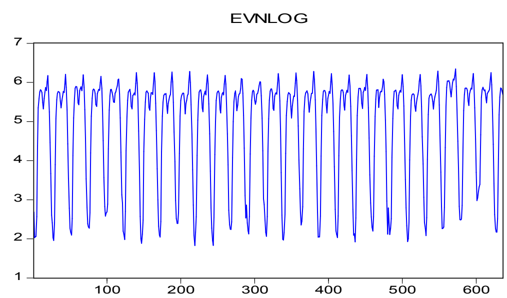
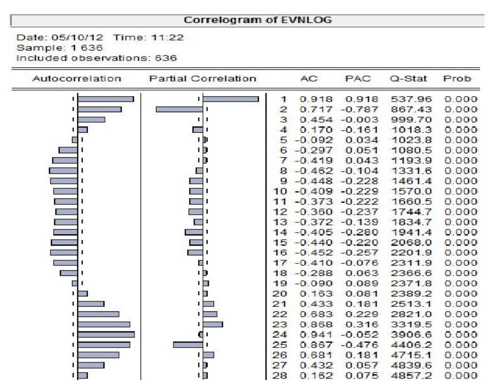


Figure 4. Periodogram based on hourly traffic trace



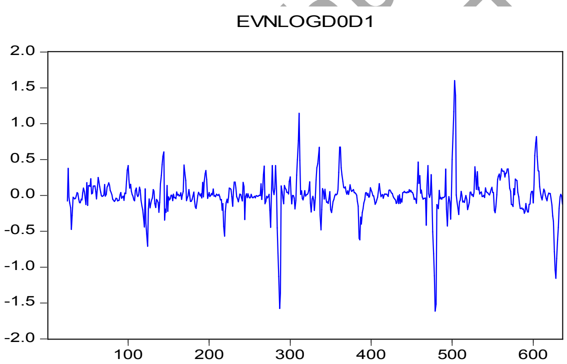


Correlogram of EVNLOG

Date: 05/10/12 Time: 11:22
Sample: 1 636
Included observations: 636

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.918	0.918	537.96	0.000
		2	0.717	-0.787	867.43	0.000
		3	0.454	-0.003	999.70	0.000
		4	0.170	-0.161	1018.3	0.000
		5	-0.092	0.034	1023.8	0.000
		6	-0.297	0.051	1080.5	0.000
		7	-0.419	0.043	1193.9	0.000
		8	-0.462	-0.104	1331.6	0.000
		9	-0.448	-0.228	1461.4	0.000
		10	-0.409	-0.229	1570.0	0.000
		11	-0.373	-0.222	1660.5	0.000
		12	-0.360	-0.237	1744.7	0.000
		13	-0.372	-0.139	1834.7	0.000
		14	-0.405	-0.280	1941.4	0.000
		15	-0.440	-0.220	2068.0	0.000
		16	-0.452	-0.257	2201.9	0.000
		17	-0.410	-0.076	2311.9	0.000
		18	-0.288	0.063	2366.6	0.000
		19	-0.090	0.089	2371.8	0.000
		20	0.163	0.081	2389.2	0.000
		21	0.433	0.181	2513.1	0.000
		22	0.683	0.229	2821.0	0.000
		23	0.868	0.316	3319.5	0.000
		24	0.941	-0.052	3906.6	0.000
		25	0.867	-0.476	4406.2	0.000
		26	0.681	0.181	4715.1	0.000
		27	0.432	0.057	4839.6	0.000
		28	0.162	0.075	4857.2	0.000

Figure 5. Logarithm form of EVN traffic series (EVNLOG) and its correlogram



Correlogram of EVNLOGD0D1

Date: 05/10/12 Time: 11:44
Sample: 1 636
Included observations: 612

	AC	PAC	Q-Stat	Prob
1	0.753	0.753	348.57	0.000
2	0.472	-0.219	485.63	0.000
3	0.274	0.016	532.11	0.000
4	0.127	-0.068	542.04	0.000
5	0.004	-0.074	542.04	0.000
6	-0.046	0.050	543.36	0.000
7	-0.036	0.039	544.16	0.000
8	-0.045	-0.073	545.39	0.000
9	-0.053	0.003	547.13	0.000
10	-0.022	0.059	547.42	0.000
11	0.037	0.065	548.29	0.000
12	0.025	-0.113	548.69	0.000
13	-0.029	-0.065	549.23	0.000
14	-0.025	0.092	549.64	0.000
15	-0.023	-0.041	549.98	0.000

Figure 6. First order of 24-period seasonal difference of EVNLOG (EVNLOGd0D1) and its correlogram

Augmented Dickey-Fuller Unit Root Test on EVNLOGD0D1

Null Hypothesis: EVNLOGD0D1 has a unit root
Exogenous: None
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.267016	0.0000
Test critical values:	1% level	-2.568752	
	5% level	-1.941342	
	10% level	-1.616351	

*MacKinnon (1996) one-sided p-values.

Figure 7. Unit root test of EVNLOGd0D1

From the correlogram in Fig. 6, we may try to estimate with AR(1) MA(1) MA(2) MA(3) and SMA(24). The estimation performing by EViews shows that the coefficients of MA(2) and MA(3) are insignificant, so we may try to drop the MA(2) and MA(3) in the model as shown in Fig. 8.

Dependent Variable: EVNLOGD0D1
Method: Least Squares
Date: 05/10/12 Time: 13:52
Sample (adjusted): 26 636
Included observations: 611 after adjustments
Convergence achieved after 17 iterations
MA Backcast: 1 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.636844	0.038914	16.36537	0.0000
MA(1)	0.316609	0.048240	6.563248	0.0000
SMA(24)	-0.941553	0.017310	-54.39274	0.0000

Figure 8. Statistics of estimating with AR(1), MA(1) and SMA(24)

Correlogram of Residuals

Date: 05/10/12 Time: 13:54
Sample: 26 636
Included observations: 611
Q-statistic probabilities adjusted for 3 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	0.022	0.022	0.2890	
2	0.013	0.012	0.3878	
3	0.001	0.001	0.3889	
4	0.013	0.013	0.5004	0.479
5	-0.073	-0.074	3.8329	0.147
6	-0.094	-0.092	9.3303	0.025
7	0.041	0.047	10.377	0.035
8	0.017	0.019	10.566	0.061
9	-0.045	-0.046	11.853	0.065
10	-0.059	-0.062	14.032	0.051
11	0.097	0.088	19.912	0.011
12	0.123	0.123	29.396	0.001
13	-0.068	-0.068	32.325	0.000
14	-0.007	-0.017	32.354	0.001
15	0.004	-0.012	32.365	0.001
16	0.030	0.039	32.944	0.002
17	0.010	0.053	33.010	0.003
18	-0.011	-0.014	33.091	0.005
19	0.002	-0.036	33.093	0.007
20	-0.054	-0.051	34.913	0.006
21	-0.025	0.010	35.317	0.009
22	0.023	0.042	35.650	0.012
23	-0.028	-0.064	36.140	0.015
24	0.060	0.048	38.462	0.011
25	0.101	0.122	45.032	0.003
26	-0.040	-0.051	46.038	0.003
27	0.039	0.038	47.012	0.003
28	0.022	0.006	47.328	0.004
29	0.014	0.002	47.450	0.006
30	-0.013	0.021	47.557	0.009
31	-0.011	0.008	47.639	0.012
32	0.025	0.023	48.039	0.015

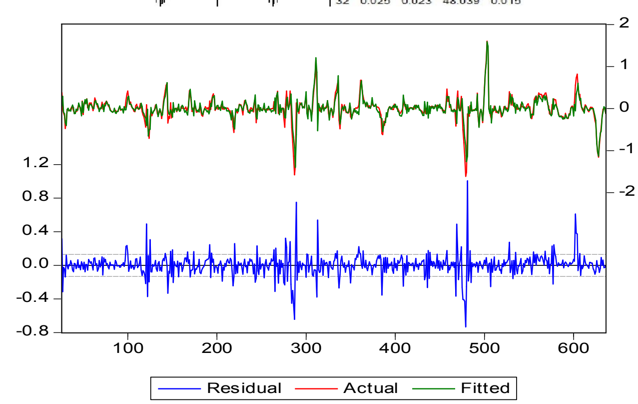


Figure 9. Correlogram of Residuals and Residual graph by estimating with AR(1), MA(1) and SMA(24)

It can be seen in Fig. 8 that the coefficients of AR(1), MA(1) and SMA(24) are all significant. Also, the AR root is satisfied the invertibility condition. Moreover, the residual of this model in Fig. 9 has achieved white noise. So this model should be reasonable and may be chosen. We can further look at the residual graph in Fig. 9 to see the fitness of the model to the series. The fitted line is nearly fitted to the actual line, so the model with AR(1), MA(1) and SMA(24) can be chosen and be used to forecast the future values base on the observation samples.

The achieved model can be expressed as: ARIMA-$(1,0,1)x(0,1,1)_{24}$ which is implemented on the logarithm form of the original series. Also, the coefficients are:

EVNLOGD0D1 = 0 + [AR(1)=0.636844024029, MA(1)=0.316609103164,SMA(24)=-0.94155323762]

According to (1), the obtained fitted multiplicative seasonal ARIMA model can be expressed detail as:

$$\left(1-\phi_1(B)\right)\nabla_{24}^{1}\ln X_t=\left(1-\theta_1(B)\right)\left(1-\Theta_1\left(B^{24}\right)\right)a_t \tag{5}$$

$$\begin{aligned}&\Leftrightarrow \nabla_{24}^{1}\ln X_t-\phi_1\nabla_{24}^{1}\ln X_{t-1}\\&\quad=\left(1-\Theta_1\left(B^{24}\right)\right)\left(a_t-\theta_1 a_{t-1}\right)\end{aligned} \tag{6}$$

$$\begin{aligned}&\Leftrightarrow \nabla_{24}^{1}\ln X_t-\phi_1\left(\ln X_{t-1}-\ln X_{t-25}\right)\\&\quad=a_t-\theta_1 a_{t-1}-\Theta_1\left(B^{24}\right)a_t+\theta_1\Theta_1\left(B^{24}\right)a_{t-1}\end{aligned} \tag{7}$$

$$\begin{aligned}&\Leftrightarrow \left(\ln X_t-\ln X_{t-24}\right)-\phi_1\left(\ln X_{t-1}-\ln X_{t-25}\right)\\&\quad=a_t-\theta_1 a_{t-1}-\Theta_1\left(a_t-a_{t-24}\right)\\&\qquad+\theta_1\Theta_1\left(a_{t-1}-a_{t-25}\right)\end{aligned} \tag{8}$$

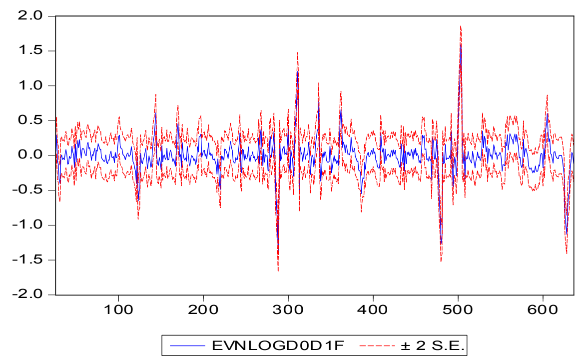
$$\begin{aligned}&\Leftrightarrow \ln X_t=\phi_1\ln X_{t-1}+\ln X_{t-24}-\phi_1\ln X_{t-25}\\&\qquad+\left(1-\Theta_1\right)a_t-\theta_1\left(1-\Theta_1\right)a_{t-1}\\&\qquad+\Theta_1 a_{t-24}+\theta_1\Theta_1 a_{t-25}\end{aligned} \tag{9}$$

where, $\phi_1=0.6368,\theta_1=0.3166,\Theta_1=0.9416$

$$\begin{aligned}&\Rightarrow \ln\hat{X}_t=\hat{\beta}_1\ln\hat{X}_{t-1}+\hat{\beta}_2\ln\hat{X}_{t-24}-\hat{\beta}_3\ln\hat{X}_{t-25}\\&\qquad-\hat{\theta}_1\hat{a}_{t-1}+\hat{\theta}_2\hat{a}_{t-24}+\hat{\theta}_3\hat{a}_{t-25}\end{aligned} \tag{10}$$

where, $\hat{a}_{t-1}=\ln X_{t-1}-\ln\hat{X}_{t-1}$, $\ln X_{t-1}$ is actual values and $\ln\hat{X}_{t-1}$ is forecast values.

The forecast of EVN traffic stream using multiplicative seasonal ARIMA$(1,0,1)x(0,1,1)_{24}$ model is now conducted. EViews provides the one-step ahead static forecasts which are more accurate than the dynamic forecasts. Static forecasting extends the forward recursion through the end of the estimation sample, allowing for a series of one-step ahead forecasts of both the structural model and the innovations. When computing static forecasts, EViews uses the entire estimation sample to backcast the innovations. In Fig. 10, the solid line represents the forecast value of hourly EVN traffic and the dotted lines which are above or below the forecasted hourly EVN traffic stream show the forecast values with ±2 of standard errors. In Fig. 11, the actual hourly EVN traffic stream is plotted using a solid red line while blue line represents the forecasted hourly EVN traffic stream by ARIMA$(1,0,1)x(0,1,1)_{24}$. The forecast series follow the actual series closely.



Forecast: EVNLOGD0D1F	
Actual: EVNLOGD0D1	
Forecast sample: 1 636	
Adjusted sample: 26 636	
Included observations: 611	
Root Mean Squared Error	0.130919
Mean Absolute Error	0.080128
Mean Abs. Percent Error	281.6309
Theil Inequality Coefficient	0.250261
Bias Proportion	0.000767
Variance Proportion	0.033886
Covariance Proportion	0.965347

Figure 10. Static forecast of EVN traffic using ARIMA$(1,0,1)x(0,1,1)_{24}$ model

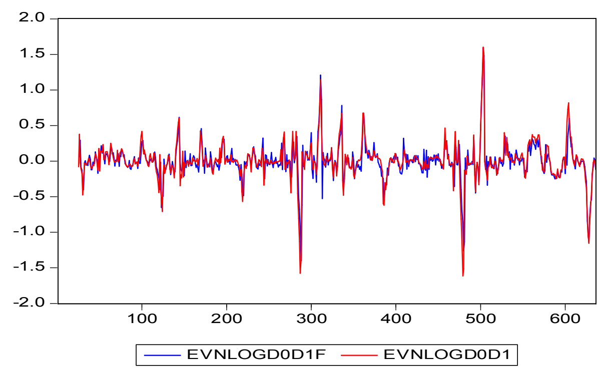


Figure 11. The plot of actual values and forecast values by ARIMA$(1,0,1)x(0,1,1)_{24}$ model

2) Building a multiplicative seasonal ARIMA/-GARCH model

In the next part, the existence of heteroscedasticity in hourly EVN traffic series is examined before starting to estimate the GARCH model. ARCH-LM

Test and Diagnostic Checking for Residuals Squared in EViews show that hourly EVN traffic data used in this study contains volatility periods. Thus, it is suitable to apply heteroscedasticity where conditional variance is not constant throughout the time trend.

The GARCH model is tested using the stationary first order seasonal difference series *EVNLOGd0D1*. To generate parameter estimates for the GARCH model, the adequacy of AR(1), MA(1) and SMA(24) terms of the mean equation are first verified. The correlogram Q-test, Jarque Bera test and ARCH test are used to test the residuals.

The correlogram Q-test in Fig. 9 indicates that there is no serial correlation on the residuals since the autocorrelations and partial autocorrelations are approximately equal to zero. Also, the first p-value of 0.479, which is testing the null hypothesis that correlations of residuals from lags 1 to 4 are all zero, is larger than 10%. Thus, Ho is not rejected and these correlations are concluded to be all zero.

On the other hand, the Jarque-Bera test in Figure 12 suggests non-normality since the statistic is large, and slight positive skewness on the residuals. Lastly, the output on the ARCH test signifies to reject the null hypothesis that there is no ARCH up to order q in the residuals because of the insignificant squared residual term.

The result of no serial correlation under the correlogram Q-test, using the AR(1), MA(1) and SMA(24) terms for the mean equation, indicates that the estimation of the conditional variance for the errors using GARCH can be implemented. The order of GARCH(p, q) is limited to 4, i.e. p and q=0,1,2,3 and 4, since GARCH is used for short-term forecasting.

Incorporating the stationary series *EVNLOGd0D1* and the mean equation with terms AR(1), MA(1) and SMA(24), a GARCH model is now estimated by finding a significant order combination under a specific error distribution (p-values should all be less than .10 level of significance and coefficient of the variance equation should all be positive). After testing different orders of p and q, it is found that the significant orders at the 10% level includes GARCH(1,0) and GARCH(1,1) assuming the error distributions of Normal Distribution, Student's t, Generalized Error Distribution and Generalized Error Distribution (GED) with fixed parameter (f.p.) r = 1.5 (default value).

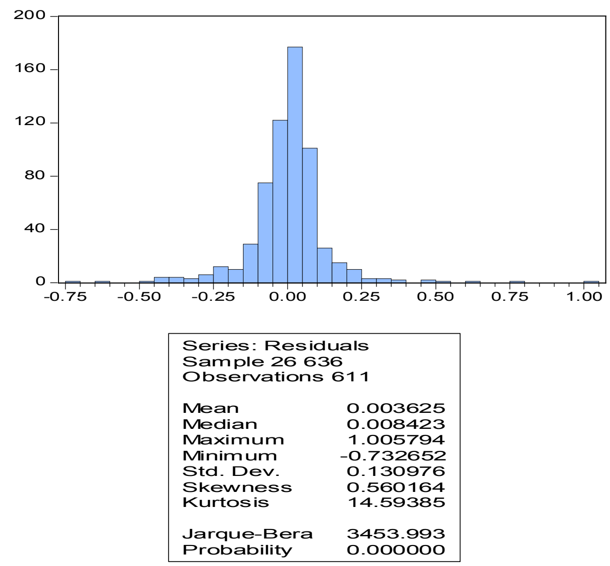


Figure 12. Jarque-Bera test for ARIMA(1,0,1)x(0,1,1)$_{24}$ model

It is now needed to validate the goodness of fit of GARCH(1,0) and GARCH(1,1) models assuming normal, Student's t, GED and GED with fixed parameter r = 1.5 distributions for the error terms by performing static forecasting on the models to show measures of forecast accuracy over the estimation period. In terms of RMSE and MAE, GARCH(1,1) assuming GED formulates the model with the smallest measure of forecast error, so this is the one with the most accurate fit of the time series model. MAE indicates that the average difference between the forecast and the observed value of the model is 0.080042, while RMSE and MAPE are 0.131390 and 276.0843, respectively.

Incorporating the most adequate choice for the volatility model, the forecast for the mean and error variance of the EVN traffic is implemented and presented in Fig. 13 using the in-sample observations under static forecasting. The figure implies that volatile values are evident during the values between about 280 to 290 and 480 to 490. This is evident in the wide confidence intervals on the GARCH model under the forecast of mean. For the other values, however, a stable and predictable traffic can be recognized, as shown in the low values of the forecast of error variance.

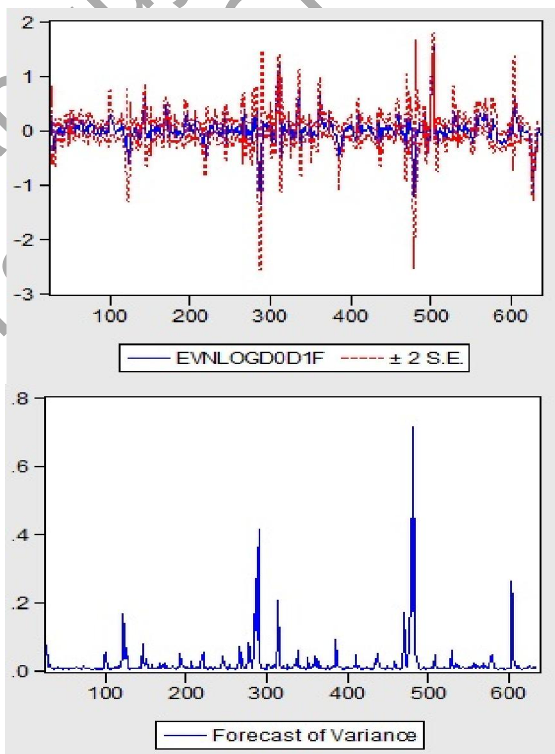


Figure 13. Forecast for the Mean and Error Variance of EVN traffic

Furthermore, some other test results can be included to make our decision more convincing. The first one is the correlogram of standardized residuals squared which consists of autocorrelation and partial autocorrelation. In Fig. 14, ACF and PACF of residuals are approximately zero. The insignificant Ljung-Box Q-statistic also provides the same evidence with p-value that GARCH(1,1) model is adequate. So we again can conclude that the model is adequate. The second one is a test for the present of conditional heteroscedasticity in the data with ARCH-LM test on the residuals as tabulated in Fig. 15. There is computed one lag difference from the residuals squared in the ARCH-LM test. The ARCH-LM for one lag difference of residuals squared is 0.2134 under $X^2(1)$. But, the null hypothesis is not rejected since the p-value is 0.6447 where it has greater than 5% of significance level. On the other hand, F-statistic the test is 0.2129 also not rejected the null hypothesis at the same condition. The ARCH-LM test on the residuals of this model indicates that the conditional heteroscedasticity is no longer present in the data.

Correlogram of Standardized Residuals Squared

Date: 05/28/12 Time: 16:43
Sample: 26 636
Included observations: 611
Q-statistic probabilities adjusted for 3 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.019	-0.019	0.2148	
		2	-0.006	-0.006	0.2335	
		3	-0.029	-0.029	0.7526	
		4	-0.026	-0.027	1.1621	0.281
		5	0.004	0.003	1.1739	0.556
		6	0.009	0.008	1.2290	0.746
		7	-0.005	-0.006	1.2444	0.871
		8	-0.033	-0.033	1.9028	0.862
		9	-0.035	-0.036	2.6584	0.850
		10	0.010	0.009	2.7263	0.909
		11	0.009	0.007	2.7769	0.948
		12	0.002	-0.002	2.7786	0.972
		13	0.006	0.005	2.8032	0.986
		14	-0.016	-0.014	2.9680	0.991
		15	-0.049	-0.049	4.4555	0.974
		16	-0.040	-0.044	5.4697	0.963
		17	-0.024	-0.029	5.8180	0.971
		18	-0.030	-0.036	6.3920	0.972
		19	-0.035	-0.041	7.1565	0.970
		20	-0.032	-0.038	7.8153	0.970
		21	0.028	0.023	8.3037	0.974
		22	-0.002	-0.006	8.3060	0.983
		23	0.009	0.000	8.3558	0.989
		24	0.016	0.010	8.5180	0.993
		25	0.042	0.040	9.6192	0.989

Figure 14. Correlogram of standardized residuals squared for ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) model

Heteroskedasticity Test: ARCH

F-statistic	0.212865	Prob. F(1,608)	0.6447
Obs*R-squared	0.213491	Prob. Chi-Square(1)	0.6440

Test Equation:
Dependent Variable: WGT_RESID^2
Method: Least Squares
Date: 05/28/12 Time: 16:36
Sample (adjusted): 27 636
Included observations: 610 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.069336	0.134996	7.921221	0.0000
WGT_RESID^2(-1)	-0.018708	0.040549	-0.461373	0.6447

R-squared	0.000350	Mean dependent var	1.049641
Adjusted R-squared	-0.001294	S.D. dependent var	3.161032
S.E. of regression	3.163076	Akaike info criterion	5.144241
Sum squared resid	6083.072	Schwarz criterion	5.158711
Log likelihood	-1566.993	Hannan-Quinn criter.	5.149869
F-statistic	0.212865	Durbin-Watson stat	2.000060
Prob(F-statistic)	0.644696		

Figure 15. ARCH-LM test for ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) model

Next, the standardized residuals for ARIMA-(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) model is plotted as shown in Fig. 16. A band of lines are joined together around mean zero with little spikes throughout the time series. The plot can be observed to have a uniform mean and a unity variance. Finally, the actual and forecast hourly EVN traffic value by

ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) model are plotted as shown in Fig. 17. The trend of forecast values follows the actual EVN traffic values closely.

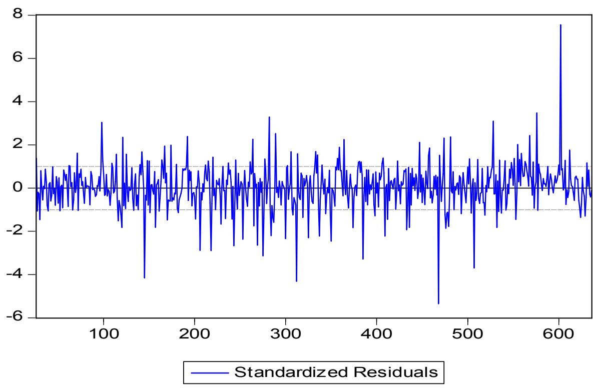


Figure 16. Standardized residuals graph for ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) model

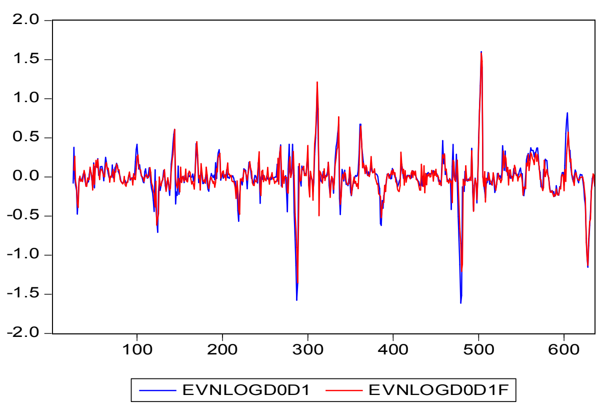


Figure 17. The plot of actual values against forecast values by ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) model

3) Evaluation of the achieved multiplicative seasonal ARIMA and ARIMA/GARCH models

The achieved results can be used to evaluate the ARIMA(1,0,1)x(0,1,1)$_{24}$ and *ARIMA(1,0,1)x(0,1,1)$_{24}$/-* GARCH(1,1) models in terms of their AIC and SIC values in the estimation stage, and forecast performances in the forecasting stage.

a. Information Criterion for ARIMA(1,0,1)x(0,1,1)$_{24}$ and ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) Models

In the model estimation step, the AIC and SIC values from ARIMA(1,0,1)x(0,1,1)$_{24}$ and ARIMA-(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) models are calculated to investigate which model is a better one for hourly EVN traffic values. In this context, the model with smaller AIC and SIC values are concluded to be the better estimation model. Table 1 tabulates the AIC and SIC values obtained from equation estimation of both ARIMA(1,0,1)x(0,1,1)$_{24}$ and ARIMA(1,0,1)x-(0,1,1)$_{24}$/GARCH(1,1) models using EViews. It can be seen that both the AIC and SIC values from ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) model are smaller than those from ARIMA(1,0,1)x(0,1,1)$_{24}$ model. This leads to the conclusion that ARIMA-(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) is a better model than ARIMA(1,0,1)x(0,1,1)$_{24}$ for estimating hourly EVN traffic values.

Table 1. Information criterion for ARIMA(1,0,1)x(0,1,1)$_{24}$ and ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) models

Model	AIC	SIC
ARIMA(1,0,1)x(0,1,1)$_{24}$	-1.2187	-1.1970
ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1)	-1.9226	-1.8721

b. Forecasting Performances of ARIMA(1,0,1)-x(0,1,1)$_{24}$ and ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH-(1,1) Models

Table 2. Forecasting performances of ARIMA(1,0,1)x(0,1,1)$_{24}$ and ARIMA(1,0,1)x-(0,1,1)$_{24}$/GARCH(1,1) models

Forecast Performance		**ARIMA**	**ARIMA/ GARCH**
RMSE		0.130919	0.131390
MAE		0.080128	0.080042
MAPE		281.6309	276.0843
Theil-U		0.250261	0.253228
MSFE	Bias Proportion	0.000767	0.000848
	Variance Proportion	0.033886	0.046625
	Covariance Proportion	0.965347	0.952527

In the forecasting stage, RMSE, MAE, MAPE, Theil-U and MSFE values of ARIMA(1,0,1)x(0,1,1)$_{24}$ and ARIMA(1,0,1)x(0,1,1)$_{24}$/GARCH(1,1) models are calculated as shown in Table 2. If the actual values and forecast values are closer to each other, a small forecast error will be obtained. Thus, smaller RMSE, MAE, MAPE, Theil-U and MSFE values are preferred.

V. CONCLUSIONS

Our study figures out that the mobile communication traffic in the condition of Vietnam has its own characteristics and can be captured as well as forecasted by using statistical models. In this case, ARIMA(1,0,1)x$(0,1,1)_{24}$ shows a good result when applied to describe and forecast EVN traffic. Concurrently, a GARCH model, which is normally used in dealing with the volatility clustering in the data series, was also constructed basing on the estimated ARIMA model to create a ARIMA(1,0,1)x-$(0,1,1)_{24}$/GARCH(1,1) model. The performance of the two models is evaluated based on some estimation criterions and forecasting errors. In term of estimation criterions, ARIMA(1,0,1)x$(0,1,1)_{24}$/GARCH(1,1) is better than ARIMA(1,0,1)x$(0,1,1)_{24}$ in estimating hourly EVN traffic values. However, not all forecasting errors by ARIMA(1,0,1)x$(0,1,1)_{24}$/-GARCH(1,1) model are smaller than those by ARIMA(1,0,1)x$(0,1,1)_{24}$ which means that volatility does exist in EVN traffic, but has an insignificant influence to the forecasting result. Therefore, in short-term prediction, ARIMA(1,0,1)x$(0,1,1)_{24}$ is also a good choice to describe and forecast EVN traffic.

REFERENCES

[1] Yantai Shu, Minfang Yu, Jiakun Liu, Oliver W.W. Yang, *"Wireless Traffic Modeling and Prediction Using Seasonal ARIMA Models"*, IEEE, 2003.

[2] Yanhua Yu, Jun Wang, Meina Song, Junde Song, "*Network Traffic Prediction and Result Analysis Based on Seasonal ARIMA and Correlation Coefficient*", 2010 International Conference on Intelligent System Design and Engineering Application, IEEE, 2010.

[3] Jia Guo, Yu Peng, Xiyuan Peng, Qiang Chen, Jiang Yu, Yufeng Dai, "*Traffic Forecasting for Mobile Networks with Multiplicative Seasonal ARIMA Models*", The Ninth International Conference on Electronic Measurement & Instruments, IEEE, 2009.

[4] Yu-Wei Chang, Meng-Yuan Liao, "*A Seasonal ARIMA Model of Tourism Forecasting: The Case of Taiwan*", Asia Pacific Journal of Tourism Research, Vol. 15, No. 2, June 2010.

[5] Weiqiang Wang, Ying Guo, "*Air Pollution PM2.5 Data Analysis in Los Angeles Long Beach with Seasonal ARIMA Model*", International Conference on Energy and Environment Technology, IEEE, 2009.

[6] Hao Chen, Qiulan Wan, Bing Zhang, Fangxing Li, Yurong Wang, "*Short-term Load Forecasting Based on Asymmetric ARCH Models*", IEEE, 2010.

[7] Sahm Kim, "*Forecasting Internet Traffic by Using Seasonal GARCH Models*", Journal of Communications and Networks, Vol. 13, No. 6, December 2011.

[8] Chenyi Chen, Jianming Hu, Qiang Meng, Yi Zhang, *"Short-time Traffic Flow Prediction with ARIMA-GARCH Model"*, IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, June 5-9, 2011.

[9] Bo Zhou, Dan He, Zhili Sun, *"Traffic Predictability based on ARIMA/GARCH Model"*, Next Generation Internet Design and Engineering, 2006. NGI '06. 2006 2nd Conference on, ISBN: 0-7803-9455-0, IEEE, 2006.

[10] Radha S, M. Thenmozhi, *"Forecasting short term interest rates using ARMA, ARMA-GARCH and ARMA-EGARCH models"*, 2006.

[11] Shan Gao, Yu He, Hao Chen, "*Wind Speed Forecast for Wind Farms Based on ARMA-ARCH Model*", IEEE, 2009.

[12] Lee Chee Nian, "*Application of ARIMA and GARCH models in forecasting crude oil prices*", a dissertation submitted in partial fulfillment of the requirement for the award of the degree of Master of Science (Mathematics), November 2009.

[13] John J. Sparks, Yuliya V. Yurova, "*Comparative Performance of ARIMA and ARCH/GARCH Models on Time Series of Daily Equity Prices for Large Companies*", Recipient of 2006 McGraw/Hill Distinguished Paper Award, 2006.

[14] Haydee L. Ramon, "*Forecasting the Volatility of Philippine Inflation using GARCH Models*", BSP Working Paper Series, Series No. 2008-01, January 2008.

[15] Nguyen Quang Dong, *"Econometrics - Advanced Program"*, Science and Technique Publisher, 2008.

[16] Nguyen Quang Dong, *"Econometrics Assignments - with EVIEWS"*, Science and Technique Publisher, 2008.

[17] I Gusti Ngurah Agung, "*Time Series Data Analysis Using EViews*", John Wiley & Sons (Asia) Pte Ltd, 2009.

[18] Nguyen Ho Quynh, *"Time series analysis and identification"*, Science and Technique Publisher, 2004.

AUTHORS' BIOGRAPHIES

Tran Quang Thanh is a lecturer at the University of Transport and Communications, Hanoi, VietNam. He is doing his Ph.D. at Southwest Jiaotong University, China. His research interests include applying statistical models to wireless traffic prediction and LTE scheduling in high velocities scenarios.

Trinh Quang Khai is with Telecommunication Department, University of Transport and Communications, Hanoi, Vietnam. His research interests include sequences design and its applications in communication systems, wireless traffic prediction.