

Application of ARIMA Model for Testing “Serial Independence” of Stock Prices at the HSEC

Cao Hao Thi – Pham Phu – Pham Ngoc Thuy
School of Industrial Management
HoChiMinh City University of Technology

ABSTRACT

The paper is an attempt to test the “serial independence” of stock prices at HoChiMinh City Stock Exchange Center (HSEC) in Vietnam by applying the ARIMA model for preliminary assessment in terms of its market efficiency. From findings derived, it appears to be that: (a) ARIMA model could be applied for testing the serial independence of stock prices at the HSEC; (b) It is failed to prove that the HSEC market is not a weak-form efficient one; and (c) the “sheep flock effect” psychology is a factor dominated at the HSEC during the past two years.

I. INTRODUCTION

The first stock exchange floor of Vietnam named “HoChiMinh City Stock Exchange Center” (HSEC) was officially opened in HCM City on July 20, 2000. After two and half years of operation, there are now 21 listed stocks and 41 bonds traded on the HSEC with a total capitalization of about VND 5,200 billion. This would be seen as a first success on the way of setting-up SEC in Vietnam.

VN-Index of the HSEC, however, has experienced a truly ups and downs movement and changed considerable during almost last two years. In the first section on July 28, 2000, VN-Index was 100 points and increased to a peak of 571.04 points in June 25, 2001 before sliding to lower 150 points in the first months of the year 2003. Therefore, the problem is: Why they fluctuated so much, lack of orientation, and whether or not they reflected to some extent the real health of the related stock companies?

From a common sense, some experts from the brokerage firms said that the stock prices at the HSEC has been fluctuated by “sheep flock effect” psychology of “naive speculators” (short time but lack of professional skills!) but not their estimates on the “intrinsic values” of stocks. However, in investment domain, this is probably the most controversial issue of “market efficiency” which has been and is likely to continue to be a topic of intensive debate in the investment community. Furthermore, the problem is much more difficult in the emerging markets like the case of the HSEC.

A market is efficient with respect to some particular information if that information is not useful in earning a positive excess return. “With respect to what information”, it could be defined three forms of market efficiency: weak, semistrong, and strong. For the case of the HSEC, it seems to be to suspect that the market is a weak-form efficient one. In that case, the hypothesis states that “stock prices reflect historical price information and,

therefore, an investor cannot “beat the market” by studying historical prices. There are many types of empirical tests for the above hypothesis and the first one is tests for “*serial independence*”. In other words, test of weak forms market efficiency would be failed if knowing how stock prices moved in the past cannot be translated into accurate predictions of future stock prices.

The available *ARIMA model* (Auto Regressive Integrated Moving Average) is an one using for forecasting future values of time series and then forecasted values depend on the own past values and the weighted success of current and lagged random disturbances. Therefore, it appears to be that the model could be seen as a tool for testing the serial independence in the issue of market efficiency.

From the above assessments, the problem of this study is that: (a) Would it be possible using the ARIMA model as a one among the tools for testing weakly efficient market hypothesis? And (b) Would it be possible providing an evidence (among others) on the weak-form efficiency of the HSEC’s market through applying ARIMA model for its historical records of stock prices?

II. ARIMA MODEL

• Stochastic process and its stationary

Any time series data can be thought of as being generated by a stochastic or random process. Broadly speaking, a stochastic process is said to be stationary if its mean and variance are constant over time and the value of covariance between two time periods depend only on the distance or lag between the two time periods and not on the actual time at which covariance is computed.

Using the graph of time series $Y_t = f(t)$, graph of autocorrelation function (Correlogram) or Dickey-Fuller test can identify the stationary of a time series.

If the time series is not stationary, difference it one or more times to achieve stationarity. If the original time series is Y_t , then the first order difference is $W_t = Y_t - Y_{t-1}$ and the second order difference is $V_t = W_t - W_{t-1}$.

• Seasonality

Seasonality is just a cyclical behavior that occurs on a regular calendar basis. Often seasonal peaks and troughs are easy to spot by direct observation of the time series. However, if the time series fluctuates considerably, seasonal peaks and troughs might not be distinguishable from other fluctuations. Fortunately, the recognition of seasonality can be made easier with the help of autocorrelation function ($SAC = f(t)$). If the graph of SAC has peaks after m time period, then it is the sign of seasonality. A crude method of “deseasonalizing” data would be to a m^{th} difference: $Z_t = Y_t - Y_{t-m}$.

• **Autoregressive Models**

According to Box-Jenkins, any of stochastic processes can be modeled as and an ARIMA processes (**AutoRegressive Integrated Moving Average**). In which, the autoregressive process is an order of p if the current observation Y_t is generated by a weighted average of past observations going back p periods, together with a random disturbance in the current period. This process is denoted as AR(p) and its equation is written as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + \varepsilon_t$$

• **Moving Average Models**

In the moving average process of order q the current observation Y_t is generated by a weighted average of random disturbances going back q periods, together with a random disturbance in the current period. This process is denoted as MA(q) and its equation is written as

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

• **AutoRegressive Integrated Moving Average Models**

Many stochastic processes cannot be modeled as purely autoregressive or as purely moving average, since they have the qualities of both types of processes. The logical extension of the models is autoregressive integrated moving average process of order (p,q). This process is denoted as ARIMA(p, d, q) and its equation is written as

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

• **Specification of ARIMA models**

Specification of ARIMA model is to choose the most appropriate values for p, d, q; in which p denotes the number of autoregressive terms, d the number of times the series has to be differenced before it becomes stationary, and q the number of moving average terms. This problem is partly resolved by examining both autocorrelation function (SAC) and the partial autocorrelation function (SPAC). SAC was defined in the previous part. SPAC is the correlation between Y_t and Y_{t-k} after removing the effect of intermediate Y's.

- Model AR(p) is chosen if SPAC correlogram has significant spikes through lags p and cuts off after p; and SAC correlogram dies down.
- Model MA(q) is chosen if SAC correlogram has significant spikes through lags q and cuts off after p; and SPAC correlogram dies down. In summary,

Type of model	Typical pattern of SAC = f(t)	Typical pattern of SPAC = f(t)
AR(p)	Decays exponentially with damped sine wave pattern or both	Significant spikes through lags p
MA(q)	Significant spikes through lags q	Declines exponentially
ARMA(p, q)	Exponential decay	Exponential decay

- **Estimation of ARIMA models**

Estimation of ARIMA model is to choose the most appropriate values for model parameters ϕ_i and θ_j . These parameters can be estimated by Ordinary Least Square Method:

$$\sum (Y_t - \hat{Y}_t)^2 \rightarrow \text{Min}$$

where

$$\varepsilon_t = (Y_t - \hat{Y}_t)$$

- **Diagnostic checking**

After determining of p, d, q and ϕ_i, θ_j , it is necessary to test whether residual terms ε_t of ARIMA model are white noise. This is a requirement of a good model. In theory, residual terms ε_t are generated by white noise process if:

$$+ \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \begin{cases} E(\varepsilon_t) = 0 \\ \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 = \text{const} \end{cases}$$

$$+ \gamma_k = \text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0$$

The testing of white noise is based on SAC correlogram and SPAC correlogram of residual terms ε_t .

- **Forecasting**

Based on the equation of ARIMA model, the values of point forecasting and confident interval will be determined

- Point Forecasting: \hat{Y}_t
- Confident interval: $\hat{Y}_t - k\sigma(\varepsilon_t) < Y_t < \hat{Y}_t + k\sigma(\varepsilon_t)$
For 95% confident level, $k=2$.

III. APPLICATION OF ARIMA MODEL IN FORECASTING THE STOCK PRICES AT THE HSEC.

- **Stock price information**

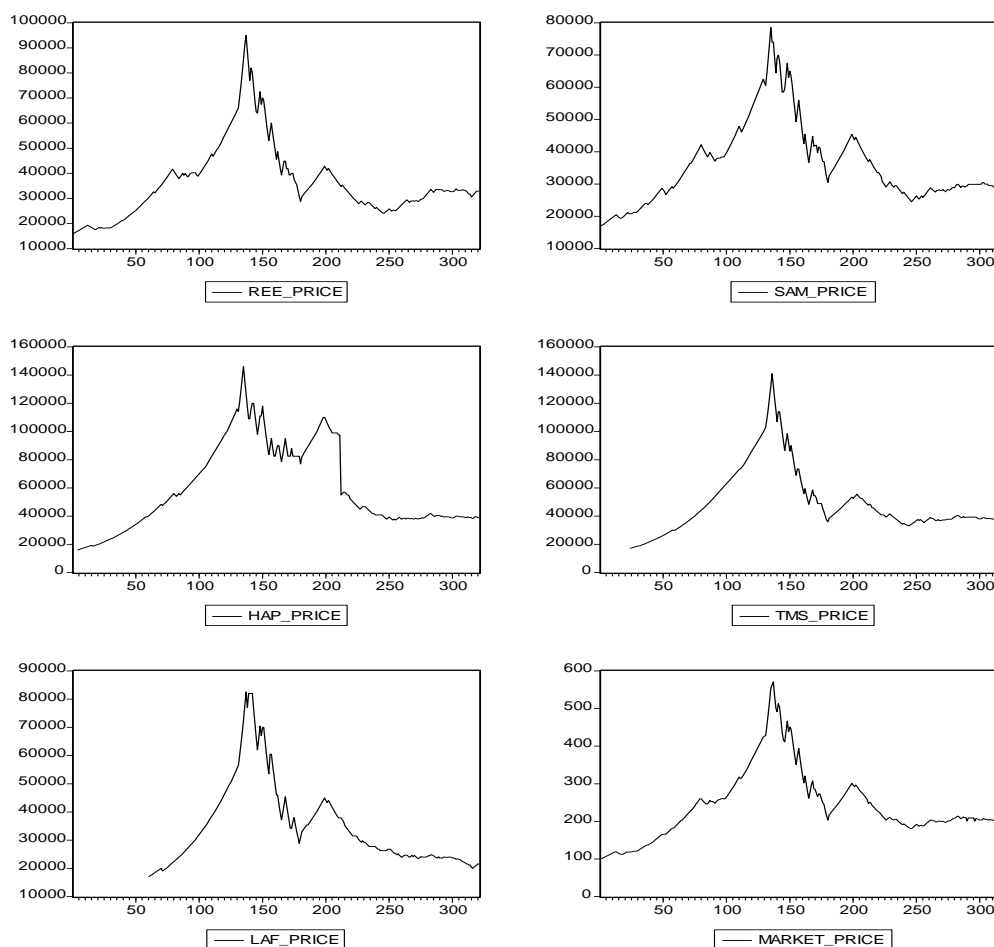
As mentioned above, up to March 2003, there are 21 stocks listed in the HSEC. However, in terms of time series of their prices, most of them are small sample sizes and are non-stationary. Therefore, in this research, only 5 stocks that have been listed and traded for a relatively long period including REE, SAM, HAP, TMS and LAF were selected for assessment. In addition, VN-Index, named MARKET PRICE, was also taken into account for comparison and assessing on the “sheep flock effect”. The main statistic parameters and the graph of these time series are as follows:

Table 1: THE STATISTIC PARAMETERS OF 5 COMPANY STOCK PRICES and MARKET PRICE

	REE_PRICE	SAM_PRICE	HAP_PRICE	TMS_PRICE	LAF_PRICE	MARKET_PRICE
Mean	36361.37	35800.00	59712.26	49498.83	34721.18	241.4512
Median	33100.00	30800.00	46800.00	40100.00	29700.00	210.5600
Maximum	95000.00	78500.00	146000.0	141000.0	82500.00	571.0400
Minimum	16000.00	17000.00	16000.00	17100.00	17000.00	100.0000
Std. Dev.	14701.51	12425.52	30103.32	23355.33	14717.06	93.01314
Skewness	1.435736	1.122369	0.618982	1.504105	1.433017	1.197553
Kurtosis	5.302573	3.908302	2.222365	5.069448	4.608104	4.506994
Observations	321	321	318	298	262	321

Source: Saigon Security Incorporation (July 2000 – June 2002)

GRAPHS OF 5 COMPANY STOCK PRICES and MARKET PRICE



• Applying ARIMA for time series of the REE’s stock prices

For illustration, the time series of REE_PRICE was selected to construct the ARIMA model. In this part, 320 past values of REE_PRICE were used to specify the equation of ARIMA model. The 321st actual value of REE_PRICE is used to compare with the forecasted value that will be estimated from the equation of ARIMA model.

By using EViews software, the optimal model, which is purely autoregressive model ARIMA (2,0,0), was finally specified. The results of autoregressive model parameters ϕ_1 is presented in following table:

Table 2: THE RESULT OF AUTOREGRESSIVE MODEL ARIMA(2,0,0) OF REE_PRICE

Dependent Variable: **REE_PRICE**

Method: Least Squares

Date: 08/01/02 Time: 16:09

Sample(adjusted): 3 321

Included observations: 319 after adjusting endpoints

Convergence achieved after 8 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	39484.06	8364.996	4.720153	0.0000
AR(1)	1.426261	0.050598	28.18831	0.0000
AR(2)	-0.435760	0.050449	-8.637551	0.0000
R-squared	0.991086	Mean dependent var		36488.09
Adjusted R-squared	0.991030	S.D. dependent var		14659.75
S.E. of regression	1388.452	Akaike info criterion		17.31913
Sum squared resid	6.09E+08	Schwarz criterion		17.35454
Log likelihood	-2759.401	F-statistic		17567.10
Durbin-Watson stat	1.917363	Prob(F-statistic)		0.000000
Inverted AR Roots	.98	.44		

After determining of ARIMA model equation, it is necessary to conduct white noise testing of residual terms ε_t . Based on SAC correlogram of ε_t , the testing demonstrated residual terms ε_t are white noise. The results of testing is shown as follow:

Table 3: THE CORRELOGRAM OF REE_PRICE RESIDUALS

Date: 08/01/02 Time: 16:11

Sample: 3 321

Included observations: 319

Q-statistic

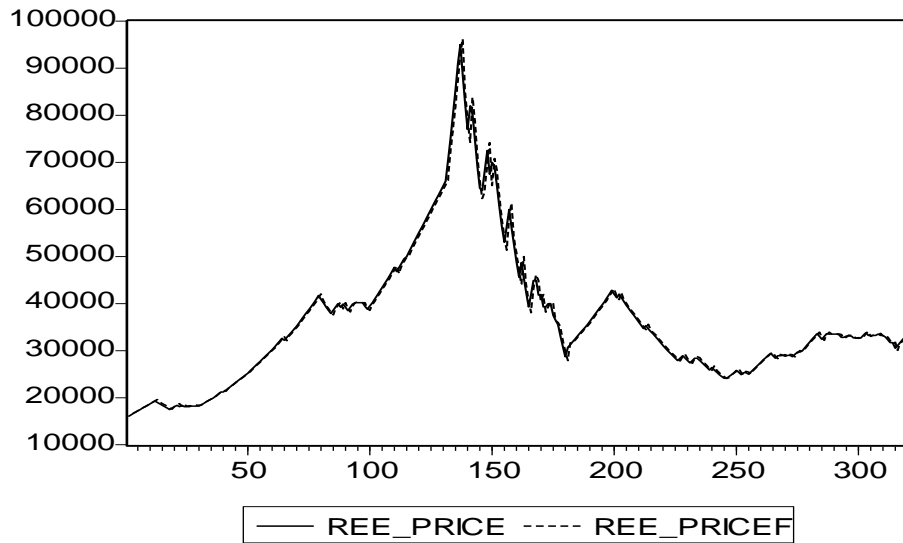
probabilities adjusted

for 2 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. .	. .	1	0.041	0.041	0.5496	
. .	. .	2	-0.055	-0.056	1.5109	
* .	* .	3	-0.137	-0.133	7.5749	0.006
. .	. .	4	0.037	0.045	8.0134	0.018
. *	. *	5	0.106	0.090	11.647	0.009
. *	. *	6	0.189	0.173	23.386	0.000
. .	. .	7	-0.049	-0.044	24.183	0.000
** .	* .	8	-0.206	-0.176	38.175	0.000
. *	. **	9	0.153	0.214	45.899	0.000
. .	. .	10	0.058	0.000	47.032	0.000
. *	. .	11	0.115	0.053	51.417	0.000
* .	* .	12	-0.078	-0.062	53.442	0.000
* .	* .	13	-0.117	-0.080	58.009	0.000
. *	. **	14	0.111	0.210	62.154	0.000
. **	. *	15	0.202	0.088	75.901	0.000
. *	. *	16	0.143	0.084	82.797	0.000
* .	* .	17	-0.133	-0.075	88.838	0.000
* .	* .	18	-0.122	-0.093	93.879	0.000
. .	. *	19	-0.011	0.080	93.920	0.000
. *	. .	20	0.152	-0.007	101.83	0.000

The results of forecasting model are stored in file name of REE_PRICEF. The graphs of REE_PRICE and REE_PRICEF are presented together as follow:

GRAPH OF ACTUAL REE_PRICE AND FORECASTED REE_PRICE



Based on the equation of forecasting model, which is ARIMA(2,0,0), the value of REE_PRICE at 321st trading time (27 June 2002) is forecasted:

Forecasting point is $\hat{Y}_t = 32,962.75$ VNĐ and

95% confident interval is [30,176.31 VNĐ, 35,749.19 VNĐ]

By comparing the actual value and forecasted value of REE_PRICE at 321st trading time (27 June 2002), the forecasting error can be evaluated. The actual value is $Y_t = 32,900.00$ VNĐ. This value falls within 95% confident interval. The forecasting error is $(\hat{Y}_t - Y_t) / Y_t * 100 = (32,962.75 - 32,900.00) / 32,900.00 * 100 = 0.19\%$

• **Applying ARIMA for time series of other stock prices**

Similarly, this research also constructed the models for SAM_PRICE, HAP_PRICE, TMS_PRICE, LAF_PRICE and MARKET_PRICE. The main results of these time series models are shown in the following table; the graphs and the other results are presented more details in appendix:

Table 4: THE RESULTS OF TIME SERIES MODELS

	REE_PRICE	SAM_PRICE	HAP_PRICE	TMS_PRICE	LAF_PRICE	MARKET PRICE
ARIMA (p,d,q) MODEL	ARIMA(2,0,0)	ARIMA(2,0,0)	ARIMA(2,0,0)	ARIMA(2,0,0)	ARIMA(2,0,0)	ARIMA(2,0,0)
Actual Price Y_t	32900.00	28900.00	38900.00	37900.00	21600.00	200.50
Forecasted Price \hat{Y}_t	32962.75	28994.22	39017.85	37944.13	21885.78	200.14
Forecasted Error ($\hat{Y}_t - Y_t$) / $Y_t * 100$	0.19%	0.33%	0.30%	0.12%	1.32%	- 0.18%
95% Confidence Interval						
Maximum	35749.19	31577.40	45805.51	41984.35	24881.52	216.02
Minimum	30176.31	26411.04	32230.19	33903.91	18890.04	184.26

• **Relationship between stock prices and VN-Index**

In addition, this research also introduced regression techniques to establish the relationship between 5 company stock prices (REE-PRICE, SAM_PRICE, HAP_PRICE, TMS_PRICE, LAF_PRICE) and MARKET_PRICE. The forecasted values of 5 company stock prices depend on the forecasted values of MARKET_PRICE, which are estimated from ARIMA model. The main results of these regression models are shown in the following table; the graphs and the other results are presented more details in appendix:

Table 5: THE RESULTS OF LINEAR REGRESSION MODELS
COMPANY STOCK PRICE = a + b* MARKET PRICE

	REE_PRICE	SAM_PRICE	HAP_PRICE	TMS_PRICE	LAF_PRICE	MARKET PRICE
Correlation Coefficient Between Company Price And Market Price	0.988974	0.989002	0.8714750	0.976444	0.944794	1.000000
Linear Regression Model						
Intercept a	-1464.15	3885.02	-12056.36	-15154.70	-8599.74	
Slope b	156.62	132.20	295.73	257.15	163.07	
Actual Price Y_t	32900.00	28900.00	38900.00	37900.00	21600.00	200.50
Forecasted Price \hat{Y}_t	29881.78	30343.53	47131.04	36311.30	24037.09	200.14
Forecasted Error ($\hat{Y}_t - Y_t$)/ Y_t *100	9.17%	4.99%	21.00%	-4.19%	11.28%	- 0.18%

By comparing the results of time series models and regression models, the forecasting errors of time series are smaller than regression models.

IV. DISCUSSIONS AND CONCLUSIONS

- (1) The graphs of 5 forecasted time series of stock prices in the HSEC are fitted well with the graphs of 5 historical records of their “actual prices”. This is resulted from that the forecasting errors are very small and all of 5 actual prices at the 321st trading section (June 27, 2002) are fallen within 95% confident intervals. Moreover, the 5 ARIMA models of the 5 time series of stock prices are in the same form of ARIMA (2,0,0). It means that they are purely autoregressive, they are “not serial independence” and having known how stock prices moved in the past it could be translated into accurate predictions of future stock prices. From the above observations, it could be said that:
 - + It would be possible using ARIMA model as an one among the tools for testing the serial independence of time series of stock prices; this is the first test for weakly efficient market hypothesis, at least in the case of the HSEC.
 - + All of 5 stock prices have reflected historical price information. So it is failed to prove that the HSEC market is not a weak-form efficient one.
- (2) In addition, linear regression models between 5 stock prices and VN-Index have been formulated with correlation coefficients of about 0.94 to 0.99. It means that the movement

of the particular price of a stock is followed the price of the other stocks and VN-Index but not the real health of the related company, or a good/ bad information disclosed from corresponding one. Consequently, it appears to be that the “sheep flock effect” psychology in securities trading is a factor dominated at the HSEC during the past two years. Moreover, with these highly significant serial correlations, a preferable trading rule might be developed.

- (3) It should be said that the above mentioned issues have been over simplified for preliminary assessments of the applying of ARIMA model in the HSEC and three year historical records of stock prices are probably not long enough to properly judge the outcomes but the basic point of the rational inference seems to be rather reasonable.

TÓM TẮT

Bài báo trình bày việc kiểm định “tính độc lập theo chuỗi thời gian” của giá chứng khoán tại Trung Tâm Giao Dịch Chứng Khoán Thành Phố Hồ Chí Minh (HSEC), Việt Nam bằng cách áp dụng mô hình ARIMA cho các đánh giá ban đầu về tính hiệu quả của HSEC. Từ các kết quả tìm được, có thể có một số kết luận như sau: (a) Có thể áp dụng mô hình ARIMA để kiểm định “tính độc lập theo chuỗi thời gian” của giá chứng khoán tại HSEC; (b) Chưa thể chứng minh được thị trường HSEC không phải là một thị trường có tính hiệu quả yếu; (c) Tại HSEC, tâm lý “Hiệu ứng bầy đàn” là nhân tố có ảnh hưởng lớn trong suốt hai năm qua.

REFERENCES

- (1) Bowerman B.L., and O’Connell R.T., 1993. *Forecasting and Time Series*. 3rd ed., Wadsworth, Inc.
- (2) Charles J. Corrado and Bradford D. Jordan, 2002. *Fundamentals of Investments – Valuation and Management*. 2nd ed., McGraw-Hill.
- (3) Eviews, 2000. Quantitative Micro Software.
- (4) Gordon J. Alexander, William F. Sharpe and Jeffery V. Bailey, 2001. *Fundamentals of Investments*. 3rd ed., Prentice-Hall, Inc.
- (5) Jack Clark Francis and Richard W. Taylor, 2000. *Theory and Problems of Investments*. 2nd ed., McGraw-Hill.
- (6) Pindyck R.S., and Rubinfeld D.L., 1991. *Econometric Models and Economic Forecast*. 3rd ed., McGraw-Hill.
- (7) Ramanathan R., 2001. *Introductory Econometrics with Applications*. 5th ed., Harcourt College Publishers.
- (8) Thi Cao Hào, và Các Cộng Sự, 1998. *Bản Dịch Kinh Tế Lượng Cơ Sở (Basic Econometrics của Gujarati D.N.)*. Chương Trình Fulbright về Giảng Dạy Kinh Tế tại Việt Nam.

Contact : Mr. Cao Hao Thi
Address : School of Industrial Management, HoChiMinh City University of Technology
268 Ly Thuong Kiet Street, District 10, HoChiMinh City, Vietnam
Tel : 84 - 8 - 8650460
Fax : 84 - 8 – 8635058
Email : chthi@sim.hcmut.edu.vn