FEA Working Paper No. 2001-7

FORECASTING THE INTRADAY KUALA LUMPUR STOCK EXCHANGE COMPOSITE INDEX

Kim-Leng Goh*
&
Kok Kim Lian**

Department of Applied Statistics
Faculty of Economics and Administration
University of Malaya
50603 Kuala Lumpur, Malaysia

*Email: g7klgoh@umcsd.um.edu.my
**Email: g7kklian@umcsd.um.edu.my

July 2001

FORECASTING THE INTRADAY KUALA LUMPUR STOCK EXCHANGE COMPOSITE INDEX

Abstract

This paper conducts time-series comparisons between a few models on the basis of forecast performance. The random walk and a simple regression, generalized autoregressive conditional heteroscedasticity (GARCH) and ARCH-in-mean models with a lag dependent structure are used for forecasting the Kuala Lumpur Stock Exchange Composite Index ten minutes ahead for three out-of-sample periods of five trading days each. On average, the difference between the actual and the forecast index is in the range of 1.4 to 2.0 points, or about 0.2 per cent. The random walk model is out-performed for majority of the cases. The GARCH model that incorporates the seasonal pattern in the mean as well as volatility of returns has the best forecast performance.

1. Introduction

Trading in stock market has broadened the dimension of investment opportunity to both individual and institutional investors, and thus, the ability to accurately forecast stock market movements has profound implications. The stock price movements in the Kuala Lumpur Stock Exchange (KLSE) are reported to exhibit a random walk behaviour by Laurence (1986), Saw and Tan (1989), Mansor (1989) and Kok and Goh (1994a), among others. This implies that historical prices do not provide additional information for predicting the future prices. Interestingly, Goh and Gui (2000) found that a random walk model is difficult to beat in their assessment of the performance of various univariate and multivariate models for forecasting the KLSE sectoral indices. The superior forecast power of random walk is also observed by Meese and Rogoff (1983) and Chinn and Meese (1995) for the foreign exchange market.

On the contrary, Lo and MacKinlay (1987) claimed that stock returns are to some extent predictable in that they are mean-reverting in the long run. Mean reversion also exists for the case of the Malaysian stock market (see, e.g., Kok and Goh 1994b). Further, significant empirical regularities pertaining to the calendar effect were found for KLSE suggesting that returns are not uncorrelated altogether. These include the day-of-the-week effect (Davidson and Peker 1996, Mansor 1997) and the time-of-the-day effect (Chang et al. 1994, Lim 1996, Goh and Kok 2001). The recent spate of development in autoregressive conditional heteroscedasticity (ARCH) modelling and the empirical support for time-varying volatility that spawned a massive literature posed yet another challenge to the stylized fact of random walk. For the Malaysian case, ARCH effects were reported by Mansor (1999) and Pan et al. (1999).

While the calendar regularities and ARCH effects have explained better the time-series behaviour of the KLSE stock prices than a random walk, their forecasting ability remains to be explored. This paper examines different models that adequately capture the seasonal patterns and time-varying volatility and shows that they can out-perform a random walk model for forecasting the KLSE Composite Index under various market conditions.

2. Data and Models

The KLSE Composite Index recorded at ten-minute intervals for every trading day spanning from September 1998 to April 1999 is used in the estimation of the models considered in the study. This covers the period where capital control is implemented. Goh and Kok (2001) observed that the period since the start of capital control is more relevant for forecasting recent stock price movements in KLSE. Empirical evidence is well established that stock indices are not stationary in levels but stationary in first differences. The intraday ten-minute returns given by:

$$r_t = \ln (p_t/p_{t-1}) \times 100$$

are used for modelling purposes, where p_t is the Composite Index recorded at time t and p_{t-1} is the index observed ten minutes before.

Six different models are considered. The random walk model is as follows:

$$p_t = p_{t-1} + \varepsilon_t \qquad (1)$$

where $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$ and t = 1, ..., n.

The other five models incorporate calendar regularities and the ARCH effects. Goh and Kok (2001) observed significant time-of-the-day effect that does not differ significantly over different days of a week. As there are 36 ten-minute intervals in a trading day, the following model is employed:

$$r_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + ... + \alpha_{36} D_{36,t} + \theta r_{t-1} + \varepsilon_t \qquad (2)$$

where D_{it} = 1 for interval-i and 0 otherwise, i = 1, 2, ..., 36, . To allow for a lag dependent structure, r_{t-1} is included in the equation.

In addition to the mean of the process, a generalized ARCH or GARCH model (see Bollerslev et al. 1992) includes an equation for modelling the time-varying volatility. The GARCH (1, 1) model is:

$$r_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + ... + \alpha_{36} D_{36,t} + \theta r_{t-1} + u_t \qquad (3)$$
$$u_t = z_t \sqrt{h_t}$$
$$h_t = \beta_0 + \beta_1 u^2_{t-1} + \beta_2 h_{t-1}$$

where $u_t | \Omega_{t-1} \sim (0, h_t)$, $z_t \sim$ i.i.d.(0, 1), and Ω_{t-1} is the information set available at time-t.

Seasonality in the returns can be artificially induced by the variation in equity market risks. To avoid this, the conditional standard deviation, a proxy for risks, is included in the mean equation of the ARCH-in-mean or ARCH-M model (see Engle et al. 1987) given by:

$$r_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + ... + \alpha_{36} D_{36,t} + \theta_1 h_t^{1/2} + \theta_2 r_{t-1} + u_t \qquad (4)$$
$$u_t = z_t \sqrt{h_t}$$
$$h_t = \beta_0 + \beta_1 u^2_{t-1} + \beta_2 h_{t-1}$$

The variance equation can be expanded to account for possible systematic time-of-the-day pattern in volatility as follows:

$$r_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + ... + \alpha_{36} D_{36,t} + \theta_1 h_t^{1/2} + \theta_2 r_{t-1} + u_t \quad (5)$$
$$u_t = z_t \sqrt{h_t}$$
$$h_t = \beta_0 + \beta_1 u^2_{t-1} + \beta_2 h_{t-1} + \phi_1 D_{1t} + \phi_2 D_{2t} + \phi_{35} D_{35,t} + \phi_{36} D_{36,t}$$

Four dummies are included in the variance equation following Goh and Kok's (2001) findings that volatility behaves differently at the open and close of the market compared to other times of a trading day.

Dropping the risk-return tradeoff in equation (5), the sixth model considered is:

$$r_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + ... + \alpha_{36} D_{36,t} + \theta_2 r_{t-1} + u_t \quad (6)$$
$$u_t = z_t \sqrt{h_t}$$
$$h_t = \beta_0 + \beta_1 u^2_{t-1} + \beta_2 h_{t-1} + \phi_1 D_{1t} + \phi_2 D_{2t} + \phi_{35} D_{35,t} + \phi_{36} D_{36,t}$$

3. Measures of Forecast Errors

All models are used to forecast one-period (ten minutes) ahead out-of-sample returns. At period t, the forecast $\hat{r}_{t+1}$ is obtained and $\hat{p}_{t+1}$ is worked out accordingly. This is repeated for s periods outside the sample. By comparing to the actual index, the model that has the lowest forecast error is deemed best. The four criteria given below are employed as forecast performance measures:

$$\text{Root mean squared error (RMSE)} = \sqrt{\frac{1}{s}\sum_{t=n+1}^{n+s}(p_t - \hat{p}_t)^2}$$

$$\text{Mean absolute deviation (MAD)} = \frac{1}{s}\sum_{t=n+1}^{n+s}\left|p_t - \hat{p}_t\right|$$

$$\text{Mean absolute percent error (MAPE)} = \frac{1}{s}\sum_{t=n+1}^{n+s}\left|(p_t - \hat{p}_t)/p_t\right| \text{ x } 100$$

$$\text{Theil inequality coefficient (Theil)} = \sqrt{\frac{1}{s}\sum_{t=n+1}^{n+s}(p_t - \hat{p}_t)^2} \Bigg/ \left[\sqrt{\frac{1}{s}\sum_{t=n+1}^{n+s}\hat{p}_t^{\,2}} + \sqrt{\frac{1}{s}\sum_{t=n+1}^{n+s}p_t^{\,2}}\right]$$

4. Results

The estimated models are given in Table 1. Returns are significantly positive half an hour after the open of the afternoon trading session (interval 25) and the last ten minutes before the market closes across all models. This implies that the seasonal pattern found in the mean process is not attributable to the variation in market risks. The returns depend significantly on the returns for the period before, and this relationship is positive. The Lagrange multiplier (LM) test for presence of ARCH effect (Engle 1982) suggests that Model (2) is statistically inadequate for capturing the time-varying volatility. The inclusion of the GARCH(1,1) terms in Models (3) through (6) is sufficient to model the effect, and both the terms are significant. The risk-return tradeoff is not significant for this period of study.

Table 1: The Estimated Models: The Mean Equation

	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
Interval 1	-0.0279	-0.0037	0.0025	0.0079	0.0018
Interval 2	0.0097	-0.0406	-0.0356	-0.0276	-0.0333
Interval 3	0.0728	0.0124	0.0211	0.0283	0.0189
Interval 4	-0.0009	0.0235	0.0308	0.0247	0.0180
Interval 5	0.0283	-0.0104	-0.0036	-0.0032	-0.0095
Interval 6	-0.0570	-0.0264	-0.0202	-0.0183	-0.0238
Interval 7	-0.0255	-0.0127	-0.0066	-0.0046	-0.0099
Interval 8	0.0199	-0.0056	0.0002	0.0013	-0.0037
Interval 9	-0.0335	-0.0132	-0.0078	-0.0068	-0.0113
Interval 10	0.0196	0.0059	0.0113	0.0124	0.0079
Interval 11	-0.0034	-0.0135	-0.0083	-0.0105	-0.0147
Interval 12	0.0030	0.0009	0.0060	0.0074	0.0032
Interval 13	0.0355	0.0287	0.0335	0.0342	0.0303
Interval 14	0.0208	-0.0099	-0.0049	-0.0077	-0.0118
Interval 15	0.0049	0.0104	0.0153	0.0155	0.0115
Interval 16	-0.0175	-0.0335*	-0.0286	-0.0289	-0.0327*
Interval 17	-0.0382	-0.0229	-0.0180	-0.0197	-0.0237
Interval 18	-0.0176	-0.0116	-0.0068	-0.0079	-0.0117
Interval 19	0.0118	-0.0163	-0.0115	-0.0122	-0.0161
Interval 20	-0.0363*	-0.0255*	-0.0209	-0.0207	-0.0244*
Interval 21	0.0432	0.0186	0.0230	0.0222	0.0187
Interval 22	0.0166	0.0229	0.0274	0.0256	0.0220
Interval 23	-0.0127	-0.0047	0.0001	0.0006	-0.0030
Interval 24	-0.0102	0.0212	0.0261	0.0289	0.0249
Interval 25	0.0575**	0.0347*	0.0395*	0.0408*	0.0372*
Interval 26	0.0468*	0.0321	0.0369	0.0412*	0.0377*
Interval 27	-0.0314	0.0222	0.0270	0.0253	0.0215
Interval 28	0.0315	0.0189	0.0236	0.0260	0.0224
Interval 29	0.0363	0.0132	0.0180	0.0191	0.0154
Interval 30	-0.0152	0.0097	0.0144	0.0196	0.0159
Interval 31	-0.0047	-0.0074	-0.0026	-0.0009	-0.0046
Interval 32	-0.0576	0.0139	0.0187	0.0116	0.0077
Interval 33	0.0009	-0.0602	-0.0555	-0.0402	-0.0436
Interval 34	0.0087	-0.0384	-0.0328	-0.0158	-0.0195
Interval 35	-0.0345	-0.1269**	-0.1232**	-0.0133	-0.0177
Interval 36	0.1295**	0.1048*	0.1066*	0.1644**	0.1589**
r_{t-1}	0.1402**	0.2508**	0.2515**	0.2539**	0.2528**
$h_t^{1/2}$			-0.0229	-0.0211	

Table 1(cont'd): The Estimated Models: The Variance Equation

	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
constant		0.0103**	0.0104**	0.0055**	0.0054**
ARCH u_t^2		0.3255**	0.3325**	0.2333**	0.2280**
GARCH h_t		0.5900**	0.5852**	0.6627**	0.6679**
Interval 1				0.2949**	0.2910**
Interval 2				-0.1689**	-0.1695**
Interval 35				0.0100	0.0100
Interval 36				0.0311*	0.0316*
ARCH LM -5 lags	1773.30**	4.9358	5.1681	1.8133	1.6536
ARCH LM -10 lags	1807.12**	5.6192	5.8624	5.4410	5.1945

Notes: ** and * denote significance at 1% and 5% respectively.

The estimated coefficients of the models are reported. The significance test is based on the White's (1980) heteroscedastic consistent standard errors for Model (2) and the quasi-maximum likelihood standard errors due to Bollerslev and Wooldridge (1992) for Models (3) to (6).

ARCH LM refers to the Engle (1982) LM test for presence of ARCH effects, and the test statistics are reported.

These six models are used to obtain the forecast of the market index for three out-of-sample periods, each consisting of 5 trading days. The samples chosen represent different market conditions so that an average of the different scenarios can be obtained. Table 2 provides some descriptive statistics and an estimated regression that shows the short-term trend. Figure 1 plots the series. All three periods have different mean levels. The volatility in the third period is lower than that of the first two periods, and price movements are within a narrower range. All three periods went through a significant trend change. The first and third periods experienced an increasing and then decreasing trend, with the change in direction found on the fourth day for the former but the third day for the latter. The second period witnessed a reverse pattern in trend with a change in direction on the fourth day.

The forecast performance measures are reported in Table 3. The bias and variance proportions of the mean squared forecast errors are rather close to zero, showing that the average performance of all the models is satisfactory. The difference between the actual and the forecast index is in the range of 1.4 to 2.0 points, or about 0.2 per cent. Based on RMSE, MAD, MAPE and Theil inequality coefficient for the average of the three periods, the random walk has the highest forecast errors. This is followed by Model (2). Generally, modelling of the volatility improves forecast accuracy, and Model (6) shows the best results. The forecast series from this model is plotted in Figure 1. Model (3) seems to be the second best. For this period, inclusion of the tradeoff between risk and return has generally increased forecast errors, rendering the poorer performance of Models (4) and (5).

Looking at the individual periods, the random walk model has lower forecast errors than all the other models only in the first period based on RMSE and the Theil inequality coefficient, and two other models for the second period based on MAD and MAPE. In all the other cases, this model has the highest forecast errors. Model (2) occasionally out-performs the other models, but generally it ranks only better than the random walk. The other four models that take into account time-varying volatility have better forecast ability that these two models.

Table 2: Out-of-sample Forecast Periods - Descriptive Statistics

	Period 1 3-7 May	Period 2 20-21, 24-26 May	Period 3 31 May, 1-4 June
$\hat{p}_t = b_0 + b_1 t + b_2 D_t + b_3 (D_t t)$			
Constant	669.68 (1.01)	787.26 (1.20)	739.72 (0.77)
Time trend (t)	0.41 (0.02)	-0.33 (0.02)	0.20 (0.02)
Trend Change (D)	94.68 (4.41)	-96.50 (5.25)	18.35 (1.51)
D*t	-0.72 (0.03)	0.72 (0.04)	-0.30 (0.02)
Summary Statistics			
Maximum	731.90	784.23	759.37
Minimum	670.62	721.98	738.47
Range	61.28	62.25	20.90
Mean	703.21	760.68	745.67
Standard deviation	18.20	15.58	4.99
Coefficient of variation	2.59	2.05	0.67

Notes: t = 1, 2, ..., 180. Coefficients (standard errors) are reported.
D = 1 for the 4th and 5th day, 0 otherwise – Periods 1 and 2
D = 1 for the 3rd, 4th and 5th day, 0 otherwise – Period 3

Table 3: Measures of Forecast Errors

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
Period 1: 3-7 May 1999						
RMSE	2.183410	2.225548	2.205307	2.204424	2.214931	2.215006
MAD	1.565556	1.564076	1.554612	1.552342	1.562248	1.562258
MAPE	0.221668	0.221496	0.220109	0.219789	0.221197	0.221191
Theil	0.001552	0.001582	0.001568	0.001567	0.001574	0.001574
-bias	0.000227	0.000010	0.000313	0.000002	0.000333	0.000018
-variance	0.001399	0.001665	0.002203	0.002237	0.001925	0.001913
-covariance	0.998372	0.998325	0.997482	0.997761	0.997740	0.998069
Period 2: 20-21, 24-26 May 1999						
RMSE	2.203326	2.201175	2.194385	2.196068	2.191157	2.188720
MAD	1.616222	1.631021	1.623128	1.628106	1.613430	1.608217
MAPE	0.213906	0.215752	0.214759	0.215410	0.213475	0.212794
Theil	0.001448	0.001446	0.001442	0.001443	0.001440	0.001438
-bias	0.002595	0.003487	0.000743	0.002018	0.004405	0.002614
-variance	0.002796	0.003470	0.004045	0.004093	0.004077	0.004030
-covariance	0.994594	0.993024	0.995207	0.993877	0.991493	0.993342
Period 3: 31 May, 1-4 June 1999						
RMSE	1.487164	1.444250	1.423070	1.425965	1.420874	1.417222
MAD	1.068111	1.030148	1.043360	1.046822	1.049040	1.045307
MAPE	0.143200	0.138122	0.139870	0.140336	0.140630	0.140129
Theil	0.000997	0.000968	0.000954	0.000956	0.000953	0.000950
-bias	0.001952	0.003970	0.000399	0.002176	0.006274	0.003155
-variance	0.000021	0.001024	0.002688	0.002768	0.002841	0.002715
-covariance	0.998016	0.994983	0.996911	0.995044	0.990850	0.994113
Average for the 3 periods						
RMSE	1.957967	1.956991	1.940921	1.942152	1.942321	1.940316
MAD	1.416630	1.408415	1.407034	1.409090	1.408239	1.405261
MAPE	0.192924	0.191790	0.191580	0.191845	0.191767	0.191371
Theil	0.001332	0.001332	0.001321	0.001322	0.001322	0.001321
-bias	0.001591	0.002489	0.000485	0.001399	0.003671	0.001929
-variance	0.001406	0.002053	0.002979	0.003033	0.002948	0.002886
-covariance	0.996994	0.995444	0.996533	0.995561	0.993361	0.995175

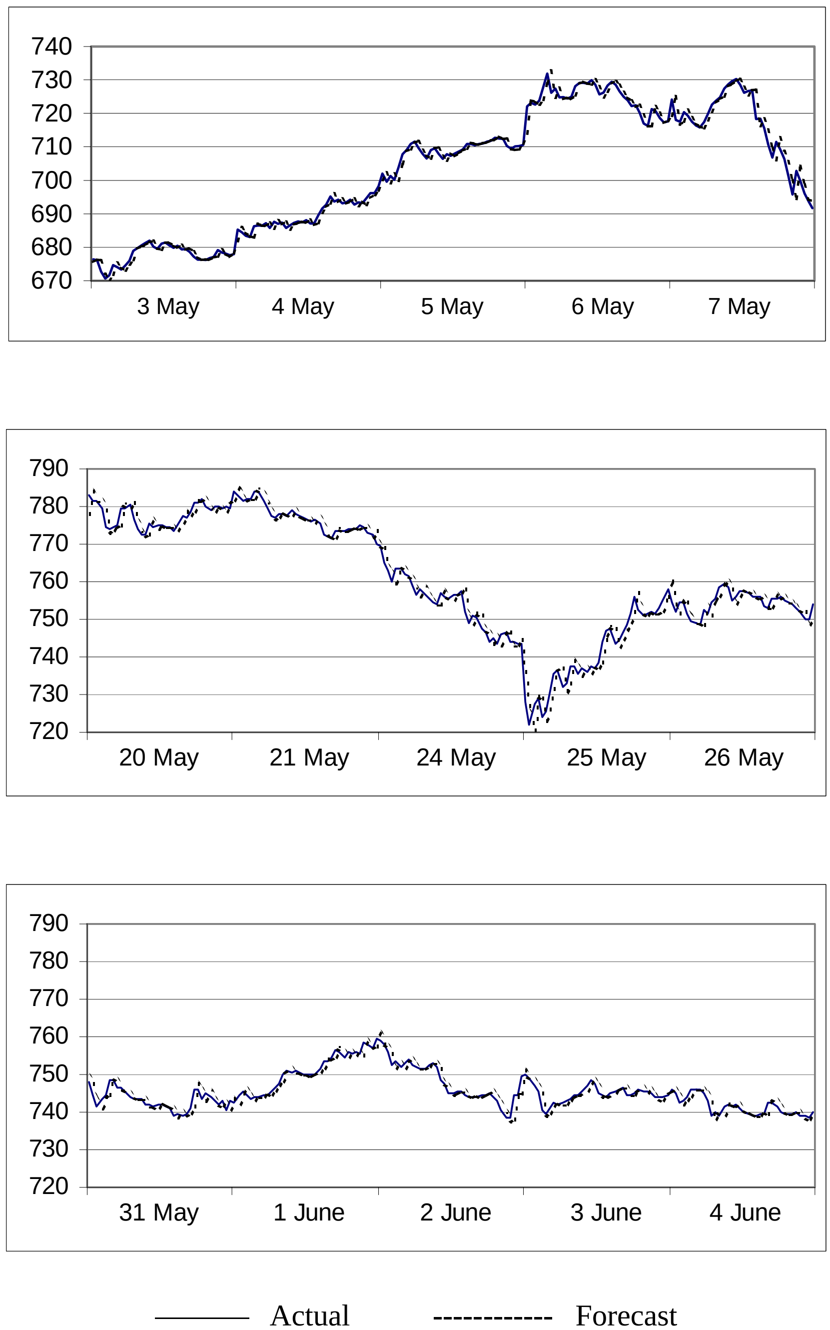
Notes: Bias, variance and covariance refer to the respective proportion of the mean squared forecast error.

Figure 1: Ten-Minute Ahead Forecasts for KLSE Composite Index, 1999

$$\text{Model}: r_t = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \ldots + \alpha_{36} D_{36,t} + \theta_2 r_{t-1} + u_t$$

$$u_t = z_t \sqrt{h_t}$$

$$h_t = \beta_0 + \beta_1 u^2_{t-1} + \beta_2 h_{t-1} + \phi_1 D_{1t} + \phi_2 D_{2t} + \phi_{35} D_{35,t} + \phi_{36} D_{36,t}$$



5. Conclusion

Significant time-of-the-day effects, time-varying volatility and lag dependence are found for ten-minute returns in the KLSE. This paper shows that models that incorporate these findings can out-perform the accuracy of a random walk for obtaining ten-minute ahead market index forecasts. The risk-return tradeoff has been irrelevant, at least for the chosen period of study. Also, the time-of-the-day effect explaining the time-varying volatility can further improve the forecast performance when taken into account.

References

Bollerslev, T. and J.M. Wooldridge, 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances, *Econometric Reviews* 11, 143–172.

Bollerslev, T., R.Y. Chou and K.F. Kroner, 1992. ARCH modelling in finance: A review of the theory and empirical evidence," *Journal of Econometrics* 52, 5–59.

Chang, R.P., J.K. Kang and S.G. Rhee, 1994. The behaviour of Malaysian stock prices, *Capital Markets Reviews* 2(1), 15-44.

Chinn, M.D. and R.A. Meese, 1995. Banking on currency forecasts: How predictable is change in money?, *Journal of International Economics* 38, 161-78.

Davidson, S. and A. Peker, 1996. Malaysian evidence on the robustness of the day-of-the-week effect, *Capital Markets Review* 4(2), 15-29.

Engle, R. F., D.M. Lilien and R.P. Robins, 1987. Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* 55, 391–407.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987–1008.

Goh, K.L. and Gui, H.K., 2000. Forecasting sectoral indices in the Kuala Lumpur Stock Exchange, *Capital Markets Review*, forthcoming.

Goh, K.L. and Kok K.L., 2001. Time-of-the-day effect in the Malaysian stock market, Proceedings for the Third Annual Malaysian Finance Association Workshop, International Islamic University Malaysia, Kuala Lumpur, forthcoming.

Kok K.L. and Goh K.L., 1994a. Weak form efficiency in the KLSE: New Evidence, *Capital Markets Review* 2, 45-60.

Kok K.L. and Goh K.L., 1994b. Weak form efficiency and mean reversion in the Malaysian stock market, *Asia-Pacific Development Journal* 1, 137-52.

Laurence, M.M., 1986. Weak-form efficiency in the Kuala Lumpur and Singapore stock markets, *Journal of Banking and Finance* 10, 431-45.

Lim, E.K., 1996. Intraday stock price and volume behaviour and their relationship in the Kuala Lumpur Stock Exchange, Unpublished M.Ec dissertation, University of Malaya.

Lo, A.W. and MacKinlay, A.C., 1987. Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of Financial Studies* 1, 41-66.

Mansor, H.I., 1997. New evidence on day of the week effect in the Malaysian stock market, *Capital Markets Review* 5(1), 23-33.

Mansor, H.I., 1999. Financial liberalisation and stock market volatility in Malaysia: A GARCH approach, *Capital Markets Review* 7, 75-86.

Mansor, M.I., 1989. Share price behaviour on the Malaysian stock market: Some empirical evidence, *Malaysian Journal of Economic Studies* 26(1), 1-20.

Meese, R.A. and K. Rogoff, 1983. Empirical exchange rate models of the seventies: Do they fit out of sample?, *Journal of International Economics* 14, 3-24.

Pan, M.S., Y.A. Liu and H.J. Roth, 1999. Common stochastic trends and volatility in Asian-Pacific equity markets, *Global Finance Journal* 10 (2), 161-172.

Saw, S.H. and Tan K.C., 1989. Test of random walk hypothesis in the Malaysian stock market, *Securities Industry Review* 15, 45-50.

White, H., 1980. A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.