

Long Term Mean Reversion of Stock Prices Based on Fractional Integration

Duk Bin Jun *

KAIST Business School

Yongjin Kim

Tepper School of Business, Carnegie Mellon University

Dae Keun Park

Accenture Management Consulting

(Received: September 26, 2011 / Revised: November 7, 2011 / Accepted: November 7, 2011)

ABSTRACT

In this study we examine the long term behavior of stock returns. The analysis reveals that negative autocorrelations of the returns exist for a super-long horizon as long as 10 years. This pattern, however, contrasts to predictions of previous stock price models which include random walks. We suggest the introduction of a fractionally integrated process into a nonstationary component of stock prices, and demonstrate empirically the existence of the process in NYSE stock returns. The predicted values of autocorrelation from our stock price model confirm the super-long term behavior of the returns observed in regression, indicating that inefficiency in the stock market could remain for a long time.

Keywords: Fractional Integration, Market Inefficiency, Mean Reversion, Stock Price Model

1. Introduction

Early research in support of the efficiency of stock markets contended that stock prices follow random walks. However, Summers [15] pointed out that stock prices could be far from the intrinsic values of companies, and they have stationary compo-

* Corresponding author, E- mail: dbjun@business.kaist.ac.kr

nents that converge slowly.

This early research was followed by a number of studies using various statistical methods to determine the existence of stationary processes in stock prices. Using the regression of stock price returns, Fama and French [4] (Hereafter FF) observed a mean reversion in the returns, which means the inefficiency of stock markets. Poterba and Summers [13] argued that a stock price is composed of not only a pure random walk but also of a stationary stochastic process.

Even though some are critical of the statistical significance of the above result (e.g., Kim, Nelson and Startz [9]; Richardson and Stock [14]; Malliaropoulos [12]), there is still a growing amount of evidence of the existence of the mean reversion. Balvers, Wu, and Gilliland [2] reported the reverting behavior of stock indexes across 18 countries, using the deviation of a stock price index from a reference instead of the controversial method. Lim and Liew [10] showed that the random walk assumption is rejected in Asian stock markets, which means that the reverting behavior exists, using nonlinear stationary tests.

However, even after these findings, researchers have brought out inconsistent results regarding how long the mean reversion of returns persists. According to FF and Khil and Lee [8] (Hereafter KL), who suggested stock price models consisting of a random walk and an autoregressive process, the mean reversion is the strongest in the 3~5 year return horizon, it starts to decrease after the peak, and it almost disappears after 7 or 8 years. Meanwhile, Poterba and Summers [13] investigated the mean reversion with a variance ratio test and demonstrated that the reversion remained significant for the 8 year, the longest interval in their study. This inconsistency means disagreement regarding the persistence of the inefficiency of stock markets. The recent paper to use fractional differencing with the stock return is Cuñado *et al.* [3] which handles long-term data using nonlinear models and fractional integration to test stock market bubbles.

In this study, we extend FF's regression horizon to 20 years from 10 in order to clarify the extremely long term behavior of stock prices. From the analysis, we report a reoccurrence of negative autocorrelations of the returns over 10 year horizons and call this reoccurrence the super-long term mean reversion. This phenomenon cannot be explained by FF's or KL's stock price models that focus on the behaviors in 'long terms,' at most 10 years, and are based on random walks. For explaining this super-

long term mean reversion, the fractionally integrated process will be a substitute for a random walk in the stock price model, since the process can have both nonstationarity and a decreasing autocorrelation like a stationary stochastic process.

The rest of the paper is organized as follows: Section 2 presents the regression method introduced by FF and describes the super-long term mean reversion. Section 3 briefly explains the theory on fractional integration, demonstrates the necessity of fractional integration, and suggests our stock price model. Section 4 introduces the data and parameter estimation method, and presents empirically the existence of fractional integration in stock returns and consistency between predictions from our model and the observations in regression. Finally, a short summary and conclusion are given in Section 5.

2. Regression Analysis and Previous Stock Price Models

FF introduced auto-regression of stock returns as a way to determine whether stock prices have predictable components. If p_t is the natural log of a stock price at time t , the continuously compounded return from t to $t+T$ is

$$r(t, t+T) = p_{t+T} - p_t. \quad (1)$$

Using New York Stock Exchange (NYSE) monthly stock return data, they regressed $r(t, t+T)$ on $r(t-T, t)$ and estimated the regression slope, $\beta(T)$, for T , varying from 1 to 10 years. They observed negative slopes which reach a minimum for the 3~5 year horizon and approach zero after the negative peak, forming a U-shape pattern.

In an attempt to explain this pattern of stock returns, they suggested the State Space Model of stock prices, which is comprised of the I(1), integrated order of 1, process and the AR(1), autoregressive order of 1, process, i.e.,

$$\begin{aligned} p_t &= q_t + z_t \\ q_t &= \mu + q_{t-1} + \varepsilon_t \\ z_t &= \phi z_{t-1} + \eta_t \end{aligned} \quad (2)$$

$\{q_t\}$ is unobserved component variables and $\{z_t\}$ is stationary cycle of autoregressive order 1, and $\{\varepsilon_t\}$ and $\{\eta_t\}$ are assumed to be normal independent and identical random variables whose expectations are zero. Also $q_0, z_0, \{\varepsilon_t\}$ and $\{\eta_t\}$ are assumed to be mutually independent. The parameters to be estimated are μ, ϕ and $\{\varepsilon_t\}$ and $\{\eta_t\}$ variances.

In this model, the theoretical value of $\beta(T)$ can be calculated as follows:

$$\begin{aligned}
\beta(T) &= \frac{\text{cov}(r(t, t+T), r(t-T, t))}{\text{var}(r(t-T, t))} \\
&= \frac{\text{cov}(q_{t+T} - q_t, q_t - q_{t-T})}{\text{var}(r(t-T, t))} + \frac{\text{cov}(z_{t+T} - z_t, z_t - z_{t-T})}{\text{var}(r(t-T, t))} \\
&= \frac{\text{cov}(z_{t+T} - z_t, z_t - z_{t-T})}{\text{var}(q_{t+T} - q_t) + \text{var}(z_{t+T} - z_t)} \\
&= \frac{-\text{var}(z_t) + 2\text{cov}(z_t, z_{t+T}) - \text{cov}(z_t, z_{t+2T})}{\text{var}(q_{t+T} - q_t) + 2\text{var}(z_t) - 2\text{cov}(z_t, z_{t+T})}.
\end{aligned} \tag{3}$$

Referring to the positive autocorrelations of returns for the short horizon such as 1 year, which were mentioned by Lo and MacKinlay [11] and Poterba and Summers [13], KL showed that the FF [4]'s model was limited in that it can result in only negative autocorrelation. They suggested replacing AR(1) with AR(2) for the stationary component of a stock price.

We repeat the regression analysis of FF, but expand the longest return horizon of regression to 20 years from 10 years in order to investigate the super-long term behavior of stocks. Huizinga [7] pointed out that a regression of returns that have data overlaps encompasses a bias in the slope. Since it is difficult to calculate analytically the bias adjustments, we simulate 1,000 random walks whose theoretical slopes are zero, and we average the slopes to estimate the theoretical bias. Then we subtract this bias from the slope of decile portfolio returns.

Figure 1 shows the regression slopes of portfolios of decile 8-10 whose super-long term mean reversions are noticeable. As earlier studies mentioned, returns on the portfolios until the 10 year horizon are U-shaped. For a horizon longer than 10 year, however, regression coefficients again go down to negative values, although FF expects the values to converge to zero. This result is consistent with Poterba and

Summers [13] in that it supports the existence of mean reversion in extremely long term horizon returns.

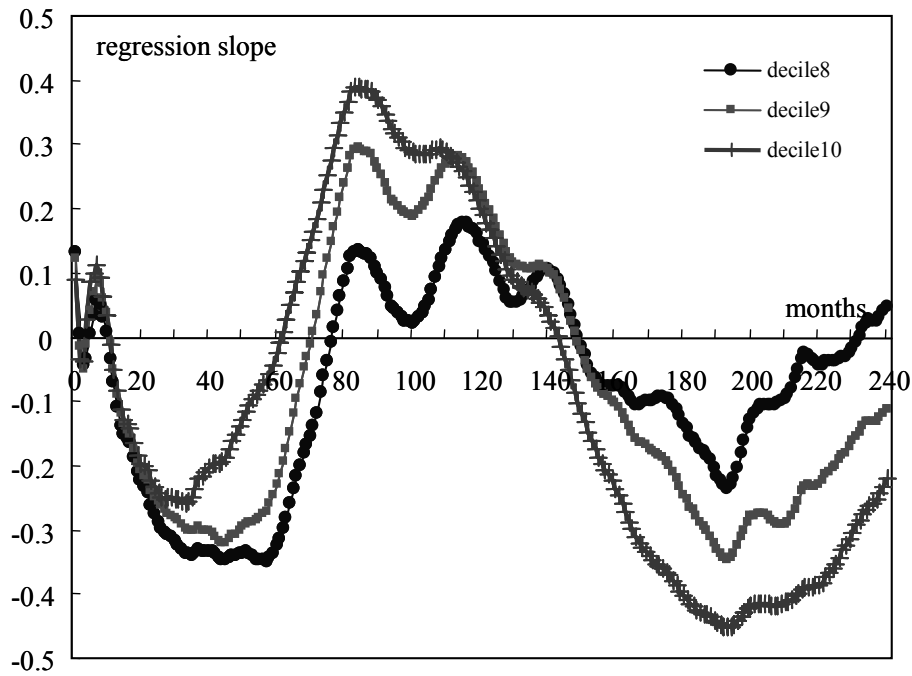


Figure 1. $\beta(T)$ of Decile 8~10 Portfolio Returns from Regression Analysis

Table 1. Values of Autocorrelation of the Returns in Regression Analysis and Predicted Values of KL

	Return horizon (months)											
	20	40	60	80	100	120	140	160	180	200	220	240
Regression	-0.196	-0.203	-0.026	0.351	0.288	0.201	0.035	-0.235	-0.403	-0.417	-0.387	-0.221
KL	-0.058	-0.031	-0.021	-0.016	-0.013	-0.011	-0.009	-0.008	-0.007	-0.006	-0.006	-0.005

In table 1, we compare $\beta(T)$ values of decile 10 from the regression with predicted values calculated from estimated parameters in the KL model. One can see that regression results show that the mean reversion can occur in horizons over 10 years, while the KL model predicts that the value uniformly converges to zero in such long horizons.

The prediction of FF and KL result from a property of a random walk. In the

above equation of $\beta(T)$, $\text{var}(q_{t+T} - q_t)$ a part of the denominator, increases in proportion to T due to permanent memory property of random walks, while other terms come close to constants with T increasing. As a result, $\beta(T)$ becomes almost zero for a sufficiently large T . Therefore, for the purpose of establishing stock price models that bring out negative autocorrelations in over 10 year return horizons, another stochastic process is necessary to replace a random walk in the models.

3. A Stock Price Model Based on a Fractionally Integrated Process

This section presents a brief explanation of fractional integration and suggests our stock price model based on this process. The ARMA process, a frequently used stationary process, assumes that autocorrelation decays exponentially as lag increases. Meanwhile, the fractionally integrated process, FI(d), represents more slowly decaying autocorrelation, and FI(d) is regarded as playing an intermediate role in the dichotomy of the time series between I(1) and I(0) (Baillie [1]).

The FI(d) process is defined as follows:

$$\begin{aligned} (1-L)^d = & 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 \\ & + \dots + (-1)^k \frac{d(d-1)\dots(d-k+1)}{k!}L^k + \dots \end{aligned} \quad (4)$$

where L is the lag operator, and d can have noninteger values such as 0.7 or -0.4. As one can see, fractional integration can be regarded as an operator with infinite integer lags through binomial expansion. In addition, FI(d) is stationary if $-0.5 < d < 0.5$ (Hosking [6]).

If the nonstationarity of the stock price is represented by FI(d) and if the stock price is assumed to have only a nonstationary component for convenience, i.e.,

$$(1-L)^d q_t = \mu + \varepsilon_t. \quad (5)$$

then, q_t can be expressed as follows:

$$\begin{aligned}
q_t &= \underbrace{(1-L)^{-d}}_{\mu'} \mu + (1-L)^{-d} \varepsilon_t \\
&= \mu' + \left(1 + dL + \frac{d(d+1)}{2!} L^2 + \dots \right. \\
&\quad \left. + \frac{d(d+1)\dots(d+k-1)}{k!} L^k + \dots \right) \varepsilon_t \\
&\equiv \mu' + (1 + \pi_1 L + \pi_2 L^2 + \dots + \pi_k L^k + \dots) \varepsilon_t.
\end{aligned} \tag{6}$$

Once the process is defined, we can calculate $\beta(T)$ as follows:

$$\begin{aligned}
\beta(T) &= \frac{\text{cov}(r(t, t+T), r(t-T, t))}{\text{var}(r(t-T, t))} \\
&= \frac{\text{cov}(q_{t+T} - q_t, q_t - q_{t-T})}{\text{var}(q_{t+T} - q_t)}.
\end{aligned} \tag{7}$$

The $q_{t+T} - q_t$ can be expressed as a moving average process with infinite lag, i.e.,

$$\begin{aligned}
q_{t+T} - q_t &= \varepsilon_{t+T} + \pi_1 \varepsilon_{t+T-1} + \dots + \pi_{T-1} \varepsilon_{t+1} \\
&\quad + (\pi_T - 1) \varepsilon_t + (\pi_{T+1} - \pi_1) \varepsilon_{t-1} + \dots.
\end{aligned} \tag{8}$$

where $\{\pi_j\}$ is moving average parameters.

Then, $\text{cov}(q_{t+T} - q_t, q_t - q_{t-T})$ can be derived, i.e.,¹

$$\begin{aligned}
\text{cov}(q_{t+T} - q_t, q_t - q_{t-T}) &= \text{cov} \left((\pi_T - 1) \varepsilon_t + (\pi_{T+1} - \pi_1) \varepsilon_{t-1} + \dots, \right. \\
&\quad \left. \varepsilon_t + \pi_1 \varepsilon_{t-1} + \dots + (\pi_T - 1) \varepsilon_{t-T} \dots \right) \\
&= \sigma_\varepsilon^2 [(\pi_T - 1) + \dots + (\pi_{2T-1} - \pi_{T-1}) \pi_{T-1} + \dots].
\end{aligned} \tag{9}$$

In the same way, we can calculate $\text{var}(q_{t+T} - q_t)$ and finally derive theoretical values of the slope for various values of d .

¹ We used an approximation in summing infinite terms. Note that as a new term is added to the summation, the absolute value of the term decrease. So we continued the summation until the absolute value of new term is larger than 10^{-6} .

Table 2. $\beta(T = 1000)$ s for Various FI(d)s

d	$\beta(T = 1000)$
0.6	-0.4496
0.7	-0.3478
0.8	-0.2440
0.9	-0.1298

Table 2 presents $\beta(T)$ for $T = 1000$ for various values of d in FI(d). Since each d is larger than 0.5, all the stochastic processes of table 2 are nonstationary. It can be seen that even though the return horizon is large enough for $\beta(T)$ to reach almost zero in the random walk model, FI(d)s' $\beta(T)$ for that horizon are negative. From these findings and the regression result of Section 2, we expect that the decile portfolios have FI(d) processes whose integration orders are over 0.5.

On the basis of these considerations, this study suggests a stock price model whose nonstationary component is represented as FI(d) and stationary component is AR(2). We can describe p_t in the State Space format, i.e.,

$$\begin{aligned}
 p_t &= q_t + z_t \\
 (1-L)^d q_t &= \mu + \varepsilon_t \\
 z_t &= \phi_1 z_{t-1} + \phi_2 z_{t-2} + \eta_t
 \end{aligned} \tag{10}$$

It is expected that predicted values of the slope from the parameter estimation of our model will decrease from near zero with an increase in the horizon. As the horizon continuously increases, an influence of the stationary component, AR(2), on the slope will lessen and the effect of FI(d) will be dominant, resulting in a negative value of the slope in the super-long term.

4. Empirical Results

4.1 Data

This section discusses the data, parameter estimation method, and results of the empirical test. For empirical testing, we use the same data as in FF, i.e., monthly returns of NYSE from July 1926 to December 1985 from the Center for Research in Secu-

rity Prices. The data is composed of portfolio returns, 10 portfolios, to which all the securities are classified with the respect to the size of the firms. We focus on the decile 8~10 portfolios which involve relatively large firms, since they show significant negative autocorrelations of returns over 10 year horizons.

4.2 Parameter Estimation

We estimate parameters in two stages. In the first stage, the fractional integration order is estimated by the spectral method introduced by Geweke and Porter-Hudak [5]. It estimates the fractional order, d , from the relationship between population and sample periodogram, and spectral representation theorem using observed $\{p_t\}$. Because even though $\{q_t\}$ is unobserved, but the fractional order of $\{q_t\}$ is the same with $\{p_t\}$, it can be determined from the spectral analysis of $\{p_t\}$.

Using the measured value of the integration order, we estimate all the remaining parameters in the State Space Model by Kalman filtering in the second stage. To specify AR orders, we try alternative ARMA models with different AR and MA orders based on the autocorrelations and partial autocorrelations and then selected the best fitted model in respect of goodness of fit, adjusted R-squares. For the definition of the model, we approximate FI(d) by 20, fully considering long-term effect, i.e.,

$$(1-L)^d q_t \approx \left(1 - dL + \frac{d(d-1)}{2!} L^2 + \dots + \frac{d(d-1) \cdots (d-19)}{20!} L^{20} \right) q_t. \quad (11)$$

4.3 Results

Table 3 reports estimation results of decile portfolios. One can see that all the estimated values of integration order are smaller than 1. This means that random walk models of stock prices could omit important properties of the return behaviors. The integration orders larger than 0.5 are compatible with the nonstationarity of stock prices. From the columns of adjusted R-squares, one can see that the stock price model containing FI(d) outperforms the previous model of KL in respect of goodness of fit.² In addition, the estimated values of integration order, fractional values, will

² For calculation of adjusted R square, we approximated the reduced form of our model as ARFIMA(2, d, 3). After order d differencing of the stock prices, we fitted ARMA(2, 3) model

lead to persistence of the mean reversion of returns in extremely long term horizons as we discuss in Section 3.

Table 3. Estimation Results of Decile 8~10 Portfolios

	d	ϕ_1	ϕ_2	σ_ε	σ_η	Adjusted R-square	Adjusted R-square of KL
Decile 8	0.7636	1.1418	-0.1638	1.6384	6.3172	0.2282	0.0517
Decile 9	0.8079	1.1311	-0.1464	1.4760	6.1585	0.1955	0.0375
Decile 10	0.8682	1.1113	-0.1225	2.1467	4.8114	0.1196	0.0352

To study more closely the behaviors of stock returns, we derive a prediction of $\beta(T)$ from estimated values of parameters in the stock price model. The following expression of $\beta(T)$ can be used:

$$\beta(T) = \frac{\text{cov}(q_{t+T} - q_t, q_t - q_{t-T})}{\text{var}(q_{t+T} - q_t) + 2 \text{var}(z_t) - 2 \text{cov}(z_t, z_{t+T})} + \frac{-\text{var}(z_t) + 2 \text{cov}(z_t, z_{t+T}) - \text{cov}(z_t, z_{t+2T})}{\text{var}(q_{t+T} - q_t) + 2 \text{var}(z_t) - 2 \text{cov}(z_t, z_{t+T})}. \quad (12)$$

Table 4 shows the predicted values of $\beta(T)$ for decile 10. As we expect, the slope starts near zero and approaches a negative value, far from converging to zero, for over 10 year horizons. In other words, the behavior of stock returns derived from our model is consistent with the regression results in return horizons over 10 years, which cannot be explained by previous models. Hence, the result of our model substantiates the extremely long term mean reversion of stock returns whose existence has been suggested by the regression analysis.

Table 4. Predicted Values of $\beta(T)$ for Decile 10 from Our Model

	Return horizon (months)											
	20	40	60	80	100	120	140	160	180	200	220	240
FI(d)+AR(2)	-0.113	-0.197	-0.26	-0.308	-0.345	-0.373	-0.394	-0.41	-0.421	-0.43	-0.436	-0.44

to the data and calculated adjusted R square. We employed the same method for KL, The reduced form of the KL model after differencing is ARMA(2, 1).

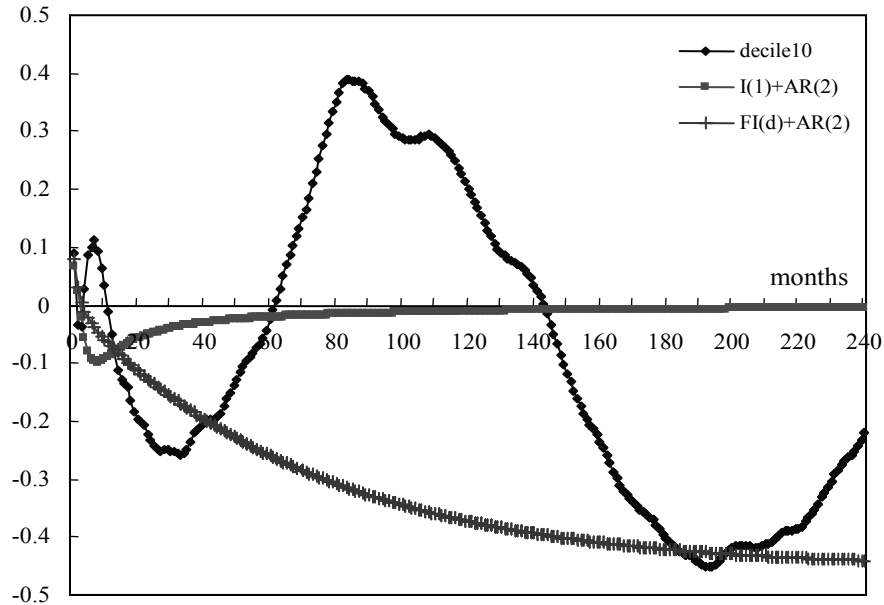


Figure 2. $\beta(T)$ of Decile 10 Portfolio Returns from Regression Analysis, the KL Model, and our model

To make it convenient for readers to compare performance of our model with the previous ones, Figure 2 demonstrates three $\beta(T)$ s for decile 10, namely, values from regression analysis, ones predicted by the KL model, and ones by our model. Compared to the previous model, our model based on FI(d) is observed to better explain the interested long term stock behavior in that it can catch negative autocorrelations of returns in super-long horizons in the regression result.

5. Conclusion

This paper has studied the super-long term behavior of stock returns. Following the regression method suggested by FF, we find negative autocorrelations of the returns in over 10 year horizons. This observation is compatible with Poterba and Summers [13]. However, some previous studies on stock returns such as FF and KL predicted no significant correlation for such a long horizon due to their dependence on random walks. This inconsistency leads to the necessity of modeling the nonsta-

tionary component of stock prices in another way rather than using random walks.

We suggest a stock price model that is the sum of a fractionally integrated process and a stationary component. Stock prices following FI(d) are observed to have an extremely long term mean reversion of returns. Through empirical tests using NYSE portfolio data, we show that integration orders of the portfolios are near 0.8, far from 1. As a result of our modeling, the predicted behaviors of the returns are similar to the observations from regression in the aspect of the super-long term relationships of stock returns. Hence, we conclude that the mean reversion of stock returns can appear for not only a 3-5 year horizon but also for over 10 years, and consequently that inefficiency in the stock market can remain for such long periods.

It should be noted, however, that the increasing pattern of the slope from the 3 to 6 year horizon in the regression analysis could not be explainable with our study. We think that a more precise description of the stationary components of stocks is needed to predict this pattern. In addition, it is likely that a stock once contained in one decile portfolio can be later involved in another decile due to a significant change in company size, hence affecting the return-relations of this portfolio. For the purpose of ruling out this possibility, future study can use another financial asset such as a specific stock or bond to investigate the behavior of returns.

References

- [1] Baillie, R., "Long memory processes and fractional integration in econometrics," *Journal of Econometrics* 73 (1996), 5-59.
- [2] Balvers, R., Y. Wu, and E. Gilliland, "Mean reversion across national stock markets and parametric contrarian investment strategies," *Journal of Finance* 55 (2000), 745-772.
- [3] Cuñado *et al.*, "Testing for stock market bubbles using nonlinear models and fractional integration," *Applied Financial Economics* 17, 16 (2007), 1313-1321.
- [4] Fama, E. and K. French, "Permanent and temporary components of stock prices," *Journal of Political Economy* 96 (1988), 246-273.
- [5] Geweke, J. and S. Porter-Hudak, "The estimation and application of long memory time series models," *Journal of Times Series Analysis* 4 (1983), 221-238.

- [6] Hosking, J., "Fractional differencing," *Biometrika* 68 (1981), 165-176.
- [7] Huizinga, J., "Test of market efficiency in models with multiperiod forecasts," Mimeographed. Chicago: University of Chicago, Graduate School of Business, 1984.
- [8] Khil, J. and B. Lee, "A time-series model of stock returns with a positive short-term correlation and a negative long-term correlation," *Review of Quantitative Finance and Accounting* 18 (2002), 381-404.
- [9] Kim, M., C. Nelson, and R. Startz, "Mean reversion in stock prices? A reappraisal of the empirical evidence," *Review of Economic Studies* 58 (1991), 515-528.
- [10] Lim, K. and V. K. Liew, "Nonlinear mean reversion in stock prices: evidence from Asian markets," *Applied Financial Economics Letters* 3 (2007), 25-29.
- [11] Lo, A. and C. MacKinlay, "Stock market prices do not follow random walks: evidence from a simple specification test," *Review of Financial Studies* 1 (1988), 41-66.
- [12] Malliaropulos, D., "Are long-horizon stock returns predictable? A bootstrap analysis," *Journal of Business Finance and Accounting* 23 (1996), 93-106.
- [13] Poterba, J. and L. Summers, "Mean reversion in stock prices: evidence and implications," *Journal of Financial Economics* 22 (1988), 27-59.
- [14] Richardson, M. and J. Stock, "Drawing inferences from statistics based on multiyear asset returns," *Journal of Financial Economics* 25 (1989), 323-348.
- [15] Summers, L., "Does the stock market rationally reflect fundamental values?," *Journal of Finance* 41 (1986), 591-601.