A Bayesian Approach for a Stock Market Instability Index: Application to the Korean Stock Market

by Kyong Joo Oh, Jae Joon Ahn, Zhiming Luo, Chiho Kim, Tae Yoon Kim*

Summary

Recently, Kim and others (2009) proposed a stock market instability index (SMII) and a corresponding p-value using a nonparametric model fitted to a stable period. However, this SMII has failed to explain the 2008 global financial market crisis because its data-driven approach relies excessively on data from the stable period upon which it is based. To resolve this problem, this study considers the random walk model and combines it with the nonparametric model using a Bayesian approach. The integrated stock market instability index (iSMII) and its p-value are derived as a posterior expectation of the two models.

Key words and Phrases: *Instability stock market index, Bayesian model, nonparametric model, random walk model*

* Kyong Joo Oh: Dept. of Information and Industrial Engineering, Yonsei Univ., Seoul 120-749, Korea (e-mail: johanoh@yonsei.ac.kr); Jae Joon Ahn: Dept. of Information and Industrial Engineering, Yonsei Univ., Seoul 120-749, Korea (e-mail: redsamzang@yonsei.ac.kr); Zhiming Luo: Dept. of Statistics, Keimyung Univ., Daegu 704-701, Korea (e-mail: zhimingluo@kmu.ac.kr); Chiho Kim: Hana Institute of Finance, Seoul 150-705, Korea (e-mail: gojeongsil@naver.com); Tae Yoon Kim (corresponding author): Dept. of Statistics, Keimyung Univ., Daegu 704-701, Korea (e-mail: tykim@kmu.ac.kr Tel:+82-53-580-5533 Fax:+82-53-580-5164).

The financial crisis triggered by the downfall of Thailand's Baht in the late 1990s caused tremendous social and economic damage in several Asian countries. This financial tsunami led many researchers to analyze and attempt to forecast financial crises. Landmark findings (see, e.g., Frankel and Rose, 1996; Kaminsky, Lizondo, and Reinhart, 1998; Kaminsky and Reinhart, 1999) have resulted from these collective efforts. Many of these results were found by studying financial crises in the context of crisis-related financial and economic variables. Conventionally, financial crisis studies have focused on fundamental economic conditions (see, e.g., Krugman, 1979; Obestfed, 1986; Eichengreen, Rose, and Wyplosz, 1996). In the 1990s, however, new theories emerged (see, e.g., Velasco, 1987; Ozkan and Sutherland, 1995) that identified self-fulfilling market processes as a reasonable alternative explanation for crises. The financial crisis in 2008, which began with the subprime mortgage problem in the USA, could be an example of such a crisis. Since the emergence of these theories, many studies of financial crises have begun to analyze the stock markets carefully because of the pivotal role it plays in such a self-fulfilling process. In the meantime, since the 1997 Asian crisis, international stock markets have expanded their territory at an unprecedented rate because of rapid capital flow and increasingly dynamic domestic and international transactions. Not surprisingly, they are becoming key indicators of current economic or financial situations, including financial crises..

Thus far most crisis studies have compared between past unstable periods and current periods of stock market to define whether current periods are crisis or not (see, e.g., Eichengreen, Rose, and Wyplosz, 1996; Frankel and Rose, 1996; Ahluwalia, 2000; Caramazza, Ricci, and Salgado, 2000). This approach, however, faces the challenge of limited available data because crises, as periods of instability, are relatively rare. To resolve this issue, Kim and others (2009) introduced a statistical testing approach (or stability-oriented approach) in which they compared data from the current period to a set of data from a past stable period, providing a much more abundant source of data to use in statistical comparison.

This statistical testing approach is comprised of four steps: (i) select the stable period from the past, (ii) fit an appropriate model to the stable period, (iii) build test statistics (or stock market instability index, *SMII*) from the residuals of the model fit and find its distribution under the null hypothesis that the given period is stable, and (iv) apply test statistics or *SMII* to the current situation. The core consideration when implementing this approach is to have the appropriate model fit for the selected stable period. Kim and others (2009) proposed a nonparametric fit to the stable period, which exhibited weaknesses due to the excessive reliance of its data-driven methodology on stable period data. In this paper, we resolve this by considering an additional model for the stable period fit. Because stock markets are known to behave like a non-stationary random walk (random walk hypothesis or efficient market hypothesis), the additional model we incorporate is the random walk model. Because it is desirable to build a single set of test statistics verifying the null hypothesis that the given period is stable, the Bayesian approach is employed to combine the two models to form the integrated stock market instability index (*iSMII*). More precisely, two *SMII* s are obtained separately from the nonlinear autoregressive (or nonparametric) and linear autoregressive (or random walk) model fits, and the Bayesian approach is then applied to obtain *iSMII* from the two *SMII* s.

The rest of this study is constructed as follows: Section I provides the statistical underpinnings of *iSMII* . Section II is devoted to a detailed description of the *iSMII* construction procedure, and Section III presents an empirical application of the *iSMII* to the Korean stock market. Lastly, concluding remarks are made in Section IV.

I. Theoretical Underpinnings of *iSMII*

The central idea of $SMII$, as described by Kim and others (2009), is to find a period of stock market stability from the past and measure the discrepancy between the current and stable periods. As mentioned in the Introduction, this is accomplished by fitting an appropriate statistical model to data from the stable period, applying the fitted model to data from the current period and deriving an appropriate measure of the discrepancy to serve as an instability index (or test statistics) based on the residuals from the application of the fitted model. This approach presupposes the statistical testing problem

$$H_0 : \text{the given period is stable.}$$
$$\text{versus} \quad H_1 : \text{the given period is not stable.} \qquad (1)$$

One remarkable advantage of this approach over the conventional approach is increased statistical reliability because test statistics (or instability indices) and their null distributions may be derived with abundant data from stable periods from statistical point of view.

The principal task for building the $SMII$ is proper modeling of the stable period. The famous random walk or efficient market hypothesis asserts intrinsic non-stationary stock market behavior. Recently, however, it has been found that there are some periods over which stock markets violate such hypotheses (see, e.g., Lo and MacKinlay, 1988). Periods not characterized by the random walk model may be divided into two different types: stationary periods caused by temporarily imperfect stock market information and periods of significant instability triggered by a financial crisis or a sudden socio-economic event. The behavior of the periods in the second category is often referred to as Levy flight, a weak form of random walk that is occasionally disrupted by large movements. See, e.g., Cheng and Savit (1987). Therefore, H_0 of (1) is established by demonstrating that the stable market follows either random walk or a stationary model. Also refer to Sornette

(2003). This implies that at least two models compete to describe H_0 over the stable period. The Bayesian approach is desirable when modeling this situation because in Bayesian statistics a set of prior distributions for competing models could be assumed and such prior distributions are as much part of the competing models as the part that expresses the probability distribution of observations given the model. Refer to Hoetting and others (1999).

We consider two competing autoregressive models for stable periods. The first is a nearly stationary autoregressive (NSAR) model that incorporates the stationary AR model as a subset. Given a set of financial time series data $Z_1, Z_2, \ldots, Z_n$, the NSAR model of order $p\ (>0)$ can be expressed as

$$Z_t = f_n(Z_{t-1}, Z_{t-2}, ..., Z_{t-p}) + e_t, \qquad t = 1, 2, \ldots, n \tag{2}$$

where e_t is an independent error process and $\lim_{n\to\infty} f_n = f$. The NSAR model may represent a non-stationary process because the autoregressive function f_n depends on n. It is worth mentioning, however, that the NSAR model exhibits some characteristics of stationary processes because of the condition $\lim_{n\to\infty} f_n = f$ and because it incorporates stationary processes Z_t as a subset. The second model is a non-stationary model that can be represented by

$$Z_t = \phi_{n1} Z_{t-1} + \cdots + \phi_{np} Z_{t-p} + e_t \tag{3}$$

where the roots of the polynomial $1 - \phi_{n1} z - \cdots - \phi_{np} z^p = 0$ lie on or outside the unit circle for any n. Note that the random walk model does not belong to the NSAR model but to the non-stationary model (3), so the two models are mutually exclusive. (2) and (3) may

yield significantly different estimates because (2) is a non-parametric stationary model while (3) fits the non-stationary linear model to existing data. Because there are two competing models for a given stable period S, one may assume that $z_S = \{Z_t : t \in S\}$ proceeds from the first NSAR model M_1 of (2) with probability w_1 and from the second random walk model M_2 of (3) with probability $w_2 = 1 - w_1$. In other words, we select each model randomly with probability w_1 and w_2. Suppose that z_s (or e_s equivalently) from model M_1 and M_2 has probability density function $z_S \sim p_1(z_S \mid M_1)$ and $z_S \sim p_2(z_S \mid M_2)$ respectively. Then z_S proceeds with finite mixture probability density function

$$z_S \sim p(z_S) = p_1(z_S \mid M_1)w_1 + p_1(z_S \mid M_2)w_2 \tag{4}$$

where w_1 and w_2 are priors, i.e., (w_1, w_2) is a discrete prior over the models. Then posterior probability of the $k-th$ ($k = 1,2$) model selection given z_S can be rewritten as

$$\hat{w}_k = P(M_k \mid z_S) = \frac{p_k(z_S \mid M_k)p(M_k)}{p(z_S)} = \frac{w_k p_k(z_S \mid M_k)}{\sum_{k=1}^{2} w_k p_k(z_S \mid M_k)} \tag{5}$$

In other words, we can update or estimate w_k by using Bayesian approach given z_S. Refer to Hoeting and others (1999).

Now we are ready to apply each model to independently derive an *SMII* for each. Indeed, by fitting each model to S, we obtain $\hat{Z}_{1t} = \hat{f}_n(Z_{t-1}, \cdots, Z_{t-p})$ or

$\hat{Z}_{2t} = \hat{\phi}_{n1} Z_{t-1} + \cdots + \hat{\phi}_{np} Z_{t-p}$. Then $SMII_1$ and $SMII_2$ are derived based on the residuals

$$\hat{e}_{1t} = Z_{1t} - \hat{Z}_{1t} \quad \text{or} \quad \hat{e}_{2t} = Z_{2t} - \hat{Z}_{2t} \tag{6}$$

Since the residuals are obtained through transform of z, they follow the same mixture distribution, i.e., $\hat{e}_{1t} = Z_{1t} - \hat{Z}_{1t}$ and $\hat{e}_{2t} = Z_{2t} - \hat{Z}_{2t}$ are selected with priors w_1 and $w_2 = 1 - w_1$. Using the residuals of the fixed time interval of test periods (e.g., one month), one may independently obtain $SMII_1$ and $SMII_2$ that measure prediction accuracies of model M_1 and M_2 fitted to the stable period S when applied to test period. See (10) for possible measures of accuracies for $SMII_1$ and $SMII_2$. Also one may easily notice that they are types of averages of residuals. For instance, $MAE(t) = \sum_{i=0}^{q} \frac{\left| Z_{t-i} - \hat{Z}_{t-i} \right|}{q}$ is a sample average of the past q absolute residuals. Since the residual proceeds with model selection probabilities w_1 and $w_2 = 1 - w_1$, a desirable measure of accuracy would be

$$iMAE(t) = \frac{1}{q} \sum_{i=0}^{q} (w_1 \left| \hat{e}_{1,t-i} \right| + w_2 \left| \hat{e}_{2,t-i} \right|) = w_1 MAE_1(t) + w_2 MAE_2(t) \tag{7}$$

where MAE_1 and MAE_2 are $SMII_1$ and $SMII_2$ developed independently. If one replaces w_1 and w_2 by $\hat{w}_1$ and $\hat{w}_2$ for $t \in S$ in (7), then $iMAE(t)$ becomes posterior mean of $MAE(t)$. This argument straightforwardly applies to *SMII*s using other measures of accuracies given in (10). Thus we have

$$iSMII(t) = \hat{w}_1 SMII_1(t) + (1 - \hat{w}_1) SMII_2(t) \tag{8}$$

In addition, we could obtain p-values as the posterior mean of $I(SMII(t) > observed\, SMII(t) \mid H_0)$ (or p-value) as a function of $\hat{w}_1$

$$iSMII(t) = \hat{w}_1(1 - F_1(SMII_1(t)) + (1 - \hat{w}_1)P(1 - F_2(SMII_2(t)) \tag{9}$$

where F_i is the null distribution function of $SMII_i$.

Finally an artificial neural network (ANN) is recommended for the nonparametric model (2) because ANN tends to over-fit to training data. This is desirable because the over-fitting of $\hat{f}_n$ to the stable period facilitates a more sensitive detection of market instability (see, e.g., Kim and others, 2004). Additionally, ANN is a well-known network for application to nonlinear financial time series data as long as they show some underlying stable character (Chatfield, 1993). In most financial applications, multi-layer perceptron (MLP) is used as a structure of ANN and the backpropagation neural network (BPN), which has become dominant since the publication of Parallel Distributed Processing (Rumelhart and others, 1986), is applied as an algorithm of ANN.

II. $i - SMII$ Development Procedure

This section describes the three technical stages for developing the *iSMII* and corresponding p-value: i) selection of the base period; ii) derivation of $SMII_1$ and $SMII_2$ using the nonparametric and random walk models and calculation of their respective $p-$values; and iii) construction of *iSMII* and the corresponding $p-$value using a Bayesian approach.

Stage 1: Selecting the Stable Period

The stable period is the period S in which the stock price index remains stable. It is desirable to find the period S during which the stock price index shows "stable movement similar to the stationary process with a fixed mean." Domestic and international stock market conditions are taken into consideration when seeking S. Exogenous variables, including economic fundamentals, foreign exchange rates and interest rates, can also be considered. A more detailed discussion about the selection of S and an example are given in Section III. Note that S can be updated as more data accumulate.

Stage 2: *SMII* Calculation

As shown in (6), two *SMII* s ($SMII_1$ and $SMII_2$) are derived from the residuals, i.e., the difference between observed value (Z_i) and predicted value ($\hat{Z}_{1i}$ or $\hat{Z}_{2i}$) for the stable period S. The following four measures are used to establish the *SMII* s: root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error type-1 (MAPE1), and mean absolute percentage error type-2 (MAPE2).

$$RMSE = \sqrt{\sum_{i=0}^{q} \frac{(Z_i - \hat{Z}_i)^2}{q}} \quad , \quad MAE = \sum_{i=0}^{q} \frac{\left|Z_i - \hat{Z}_i\right|}{q}$$

$$MAPE\ 1 = \frac{1}{q}\sum_{i=0}^{q} \frac{\left|Z_i - \hat{Z}_i\right|}{\hat{Z}_i} \times 100, \quad MAPE\ 2 = \frac{1}{q}\sum_{i=0}^{q} \frac{\left|Z_i - \hat{Z}_i\right|}{Z_i} \times 100 \qquad (10)$$

Stage 3: Integrating $SMII_1$ and $SMII_2$ into *iSMII*

In this stage, a proper prior distribution, $(w_1, w_2) = (w_1, 1 - w_1)$, is determined first.

For this, we'd better examine validity of the random walk theory to the market. Indeed one may implement some random walk test such as the Dickey and Fuller test to S and refer to existing random walk test results for the stock market under consideration. Recall that the random walk hypothesis is a stock market theory stating that the chance of a stock's future price going up is the same as it going down and hence it is impossible to predict its future movement (see Malkiel (1973)).

Next, a posterior distribution, $(\hat{w}_1, \hat{w}_2) = (\hat{w}_1, 1-\hat{w}_1)$, is estimated from the period S to establish an $iSMII$ that integrates $SMII_1$ and $SMII_2$ using a Bayesian approach. To estimate posterior probabilities $(\hat{w}_1, \hat{w}_2) = (\hat{w}_1, 1-\hat{w}_1)$, we use

$$\hat{w}_k = P(M_k \mid e_S) = \frac{p_k(e_S \mid M_k)p(M_k)}{p(e_S)} = \frac{w_k p_k(e_S \mid M_k)}{\sum_{k=1}^{2} w_k p_k(e_S \mid M_k)} . \tag{11}$$

Here $p_k(e_S \mid M_k)$ might be estimated by nonparametric or parametric method from the residuals $\hat{e}_{1t}(S)$ or $\hat{e}_{2t}(S)$. In addition, the null distributions of the respective $SMII$ s in (9) can be constructed using an empirical distribution from the stable period S, i.e.,

$$\hat{F}_{SMII}(x) = \frac{\text{the number of SMIIs smaller than x in S}}{\text{the number of total SMIIs in S}} \tag{12}$$

The $p-$value calculated from (12) can be used to issue stock market warnings because it indicates the probability that the stock market is unstable.

III. Application to the Korean Stock Market

Local government authorities monitoring a stock market are always concerned about managing its stability. During the last few decades, local governments have experienced the unprecedented phenomena of globalization and economic deregulation combined with remarkable advancement of electronic trading systems, which have greatly enhanced international transfers of funds. Consequently, stock market stability management has grown in difficulty. Since the 1997 Asian financial crisis, the Korean government has been very concerned about developing a stock market instability index. The government's concern is particularly understandable because the Korean stock market has opened its door to participation by other nations to encourage economic growth and overcome future crises. In this section, the *iSMII* is developed for the Korean stock market as an empirical example. This illustrates the various technical aspects of the Bayesian approach for *iSMII* construction.

The Stable period

The movements of the Korea Stock Price Index (KOSPI) after 1994 are examined to identify the appropriate stable period S. Only movements after 1994 are used because of the need for sufficient volume and efficiency. After scrutinizing the movements of the KOSPI, the year 1995 was chosen as the stable period S (see Figure 1). The KOSPI during 1995 was quite stable, staying between 900 and 1,000. Furthermore, the movement of the 1995 KOSPI neither rose nor fell significantly from its yearly average. Based on these considerations, the year 1995 was selected as the stable period S. Note that 1995 was also known to be a relatively stable year with respect to economic fundamentals such as GDP growth rates and internal and external equilibriums.

Insert Figure 1 here

$SMII_1$ and $SMII_2$ development

Let $Z_1, Z_2, \cdots, Z_n$ represent the 1995 KOSPI time series. From Figure 1, the movement of $Z_1, Z_2, \cdots, Z_n$ is stable; i.e., its trend can be regarded as an approximately constant mean. The sample autocorrelation function (SACF), sample partial autocorrelation function (SPACF) of $Z_1, Z_2, \cdots, Z_n$ and scatter plot of (Z_{t-1}, Z_t) are shown in Figure 2 and Figure 3. Figure 2 shows that the SACF slowly approaches zero as lag increases and SPACF cuts off at lag 3. Figure 3 shows that the underlying autoregressive function is a straight line with a slope of 1. Figure 3 is unexpected in the sense that the random walk model does not usually reveal its underlying autoregressive straight line due to its increased variance. Thus Figure 2 and Figure 3 together suggest that KOSPI is essentially characterized by a non-stationary random walk but also exhibits some stationary attributes.

Insert Figure 2 here

Insert Figure 3 here

Now we employ the following NSAR model

$$Z_{1t,n} = f_n(Z_{t-1}, Z_{t-2}, Z_{t-3}) + e_t, \qquad t = 1, 2, \ldots, n \tag{13}$$

and the non-stationary linear AR model

$$Z_{2t,n} = \phi_{n1} Z_{t-1} + \phi_{n2} Z_{t-2} + \phi_{n3} Z_{t-3} + e_t \tag{14}$$

where order-3 is selected by SPACF. As discussed in Section I, ANN is implemented using a BPN called a $3 \times 3 \times 1$ multilayer (3 input layers, 3 hidden layers, and 1 output layer) perceptron to estimate f_n according to (13). The least squares method is used to estimate the ϕ_n s of model (14). The fitted models are

$$\hat{Z}_{1t,289} = \hat{f}_{289}(Z_{t-1}, Z_{t-2}, Z_{t-3}) + 0.1415\hat{\varepsilon}_{t-1} \tag{15}$$

$$\hat{Z}_{2t,289} = 1.0164 Z_{t-1} - 0.0069 Z_{t-2} - 0.01 Z_{t-3} \tag{16}$$

Note that (15) and (16) are estimated using $n = 289$, the number of data points for the year 1995. Moreover, the residuals in (15) are fitted using the AR(1) error model after ANN fitting. Recall that ANN fitting cannot be explicitly expressed because it is a completely data-driven fitting. From (16), it is easy to see that the autocorrelation coefficient $\hat{\phi}_{n1}$ is almost 1 with $\hat{\phi}_{n2}$ and $\hat{\phi}_{n3}$ almost zero, which is indicative of the random walk model. We perform unit root test to verify this conclusion. In Table 1, a Dickey-Fuller unit root test is conducted for the null hypothesis that "the underlying model is the pure random walk." Under the significant level 0.05, the Dickey-Fuller test indicates that the null hypothesis cannot be rejected because $p-$ value is 0.30905.

Insert Table 1 here

To derive $SMII_1$ and $SMII_2$ based on residuals $\hat{e}_{1t}$ and $\hat{e}_{2t}$ over the stable period S, we consider the four measures given in (10) with q equal to the number of data points in one month, resulting in a series of monthly $SMII_1$ and $SMII_2$ values. Furthermore, we normalize their values between 0 and 1 to remove the scale effect for each measure, i.e.,

$$SMII^*(t,t_0) = \frac{SMII(t) - \min_{1\le t\le t_0} SMII(t)}{\max_{1\le t\le t_0} SMII(t) - \min_{1\le t\le t_0} SMII(t)} \tag{17}$$

where t_0 is the last day for $SMII(t)$. Figure 4 shows the movement of $SMII_1$ when RMSE, MAE, MAPE1 or MAPE2 is employed. It is easy to see that each measure provides nearly identical movements before 2006, but MAPE2 yields different movements after 2006.

Before going further into construction of the $iSMII$, let us discuss the history of the Korean stock market after 1995 to provide a basic criterion for evaluating the accuracy of the $SMII$ s. There were five major crises in Korea prior to 2011, i.e., the 1997 crisis due to the Asian crisis, the 2001 crisis due to the bankruptcies of major conglomerates, the 2003 credit card crisis, the 2008 crisis due to the US subprime mortgage crisis and the 2011 crisis due to the European financial crisis. Among them, the 1997 and 2008 crises are considered major in Korea. In Figure 4, all five crises are clearly depicted as five distinct peaks by each of the four measures. A problem with RMSE, MAE and MAPE1 is that they overemphasize or produce early peak for the last two crises via the two highest peaks. This is related to the fact that these measures tend to identify a market as unstable when the level of the KOSPI is significantly different from that of 1995, regardless of its relative direction. MAPE2 smoothes out the peaks in 2007 and 2011 and is therefore selected as an appropriate

measure for $SMII_1$. Using similar reasoning, MAPE2 is also selected for calculation of $SMII_2$. Figure 5 shows the movements of $SMII_2$. It is interesting to note that $SMII_2$ successfully detects the two major crises (1997 and 2008) and one minor crisis (2011) but fails to respond well to the other two crises. $SMII_2$ not only fails to identify the second and third small peaks (2001 and 2003 crises) but also frequently issues wrong signals during stable periods. This results from the fact that stock markets frequently become stationary when they are stable but the fitted model (16) regards them as departures from the (non-stationary) random walk.

Insert Figure 4 here

Insert Figure 5 here

Now we calculate the $p-$values of individual $SMII$ using the empirical distribution function in (12). Because the $p-$value can be interpreted as the probability that the market is stable, it can be used to issue financial crisis warning signals. The $p-$value plots of $SMII_1$ and $SMII_2$ are presented in Figure 6. Note that the $p-$value plot of $SMII_1$ shows 5 valleys corresponding to the 5 peaks of $SMII_1$ and there is some readily observable correlation between $SMII_2$ and its $p-$value. It is also easy to see that the $p-$value of $SMII_2$ produces many deep valleys that do not correspond to actual financial crises. The merits and drawbacks of each of the two $SMII$ s warrant integration of the two measures.

Insert Figure 6 here

Bayesian Integration of $SMII_1$ and $SMII_2$

In this section, $SMII_1$ and $SMII_2$ are integrated into $iSMII$ using the Bayesian approach of Section I. Recall that, as seen in Table 1, the Dickey-Fuller unit root test verifies that "the underlying model is the pure random walk." In addition Ayadi and Pyun (1994) showed that South Korean market does follow random walk when the test statistic is corrected for heteroscedasticity. Considering these results, we give the random walk model quite high prior, i.e.,

$$(w_1, w_2) = (0.04, 0.96) \tag{18}$$

Using (11) and (18), we calculate posterior probabilities. For implementing (11) effectively, we assume normal distributions for the residuals and then estimate means and variances. Refer to Figure 7 which provides two residual histograms with respective mean and standard error from ANN and AR fitting . Now we have the resulting posterior probabilities

$$\hat{w}_1 = P(ANN \mid e_S) = \frac{p_1(e_S \mid ANN)p(ANN)}{p(e_S)}$$

$$= \frac{p_1(e_S \mid ANN)\times 0.04}{p_1(e_S \mid ANN)\times 0.04 + p_2(e_S \mid AR)\times 0.96} = \frac{1}{1+0.035\times 24} = 0.543$$

where

$$p_1(e_S \mid ANN) = \Pi_{i=1}^{n(s)} \frac{1}{9.743\sqrt{2\pi}} \exp((e_{1i}(S)+0.0342)^2 / 2(9.743)^2)$$

$$p_2(e_S \mid AR) = \Pi_{i=1}^{n(S)} \frac{1}{9.856\sqrt{2\pi}} \exp((e_{2i}(S)+0.017)^2 / 2(9.856)^2)$$

where $n(s)$ is the number of residuals in S and $e_{1i}(S)$ and $e_{2i}(S)$ are respectively the residuals from ANN and AR fitting. Thus we have

$$(\hat{w}_1, \hat{w}_2) = (0.543, 0.467)\ . \tag{19}$$

This shows that even though random walk hypothesis is assumed to be true by giving high prior, the market appears to follow nonrandom walk frequently over the year 1995. Technically our result demonstrates that Bayesian approach developed here might be useful for checking the random walk hypothesis validity against the real situation.

Insert Figure 7 here

Figures 8 and 9 provide plots of the $iSMII$ for various choices of posterior for $\hat{w}_1$. For example, 9_1 means $(\hat{w}_1, \hat{w}_2) = (0.9, 0.1)$ and

$$iSMII(t) = 0.9\ SMII_1(t) + 0.1\ SMII_2(t) \tag{20}$$

where t is a given month and both $SMII_1$ and $SMII_2$ are scaled to between 0 and 1 as in (17).

Insert Figure 8 here

Insert Figure 9 here

It is interesting to see that $iSMII$ with $\hat{w}_1 = 0.543$ yields the correct peaks with clear crisis detections for the two major crises. The other $iSMII$ s with different values of $\hat{w}_1$ fail to accurately reflect actual Korean stock market history.

$p-$values of $iSMII$ are obtained from null distributions of $SMII_1$ and $SMII_2$ with $\hat{w}_1 = 0.543$ using formula (9). In addition to the $p-$value's utility as a probabilistic interpretation of the Korean financial market environment, it has the added technical advantage of being automatically normalized from 0 to 1. The $p-$values of $iSMII$ and KOSPI are given in Figures 10 and 11 for various values of $\hat{w}_1$. As with $iSMII$, $\hat{w}_1 = 0.543$ is justified by examining Figure 12. First, it is clear that the $p-$value decreases when the severity of the US subprime mortgage problem was initially recognized in August 2007 and sharply decreases again when the full crisis hits the market in September 2008. Second, the recent crisis resulting from the European financial crisis has been detected last year in 2011. Third, the 1997 IMF crisis appears to be the most severe among the crises that Korea has experienced. Upon comparison of Figure 12 with Korean financial market history, it appears that it might have been helpful if a warning had been issued if the $p-$value dropped to less than 0.2.

Insert Figure 10 here

Insert Figure 11 here

Insert Figure 12 here

IV. Concluding Remarks

This article proposes a statistical testing approach for developing stock market instability index for efficient monitoring of stock market. It shows that an index with a

good accuracy could be developed in a statistically rigorous fashion. This technique is expected to extend to other market such as exchange and interest rates market. However, what has been done here with stock market index might be technically enhanced further. Indeed the training period S may be updated by using monthly KOSPI data whose p-value from $iSMII$ is greater than a particular value of β (e.g., $\beta = 0.5$). More models appropriate for fitting the stable period S might be introduced such as ARCH (auto-regression conditional heteroscedasticity) model. Empirical null distribution function estimation in estimation might be improved by using advanced technique such as wild bootstrap. These further enhancements are left for future research. Finally, it is worth noting that nonparametric model fits the nearly stationary period effectively and therefore successfully complements the random walk model in this Bayesian approach.

References

Ahluwalia P. 2000, Discriminating contagion: An alternative explanation of contagious currency crises in emerging market. *IMF Working Paper* No. WP/00/14.

Ayadi OF, Pyun C S. 1994, An application of variance ratio tests to the Korean securities market. *Journal of Banking and Finance* **18**, 643-658.

Caramazza, F, Ricci LA, Salgado R. 2000, Trade and financial contagion in currency crises. *IMF Working Paper* No. WP/00/55.

Chatfield, C. 1993, Neural networks: Forecasting breakthrough or passing fad? *International Journal of Forecasting* **9**, 1-3.

Cheng Z, Savit R. 1987, Fractal and nonfractal behavior in Levy flights. *Journal of Mathematical Physics* **28**(3), 592-597.

Eichengreen B, Rose AK, Wyplosz, C. 1996, Contagious currency crisis. *Scandinavian Financial Market Review* **98**, 463-484.

Frankel JA, Rose AK. 1996, Currency crashes in emerging markets: An empirical treatment. *Journal of International Financial Markets* **41,** 351-366.

Hoeting JA, Madigan D, Raftery, AE, Volinsky, CE. 1999, Bayesian model averaging: A tutorial. *Statistical Science* **14**, 382-417.

Kaminsky GL, Lizondo S, Reinhart, CM. 1998, Leading indicators of currency crises. *IMF Staff Papers* **45** (1), 1-48.

Kaminsky GL, Reinhart CM. 1999, The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review* **89**(3), 473-500.

Kim DH, Lee SJ, Oh KJ, Kim, TY, 2009, An early warning system for financial crisis using stock market instability index. *Expert Systems* **26**(3), 260-273.

Kim TY, Oh KJ, Sohn I, Hwang C. 2004, Usefulness of artificial neural networks for early warning system of economic crisis. *Expert Systems with Applications* **26** (4), 583-590.

Krugman PR, 1979, A model of balance-of-payment crises. *International Journal of Money, Credit and Banking* **11** (3) 527-529.

Lo AW, MacKinlay AC. 1988, Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies* **1** (1), 41-66.

Malkiel BG, 1973, *A Random Walk Down Wall Street.* W.W. Norton and Company Inc.

Obestfed, M. 1986, Rational and self-fulfilling balance-of-payment crisis. *American Economic Review* **76** (1), 72-81.

Ozkan GF, Sutherland, AJ. 1995, Policy measures to avoid a currency crisis. *Economic Journal* **105** (429), 510-519.

Rumelhart DE., McClelland JL., the PDP Research Group, 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* MIT Press, Cambridge.

Sornette, D. 2003, *Why stock markets crash.* Princeton University Press, Princeton and Oxford.

Velasco, A. 1987, Financial and balance of payment crises: A simple model of the Southern Cone Experience. *Journal of Development Economics* **27,** 263-83.

Figures and Tables:

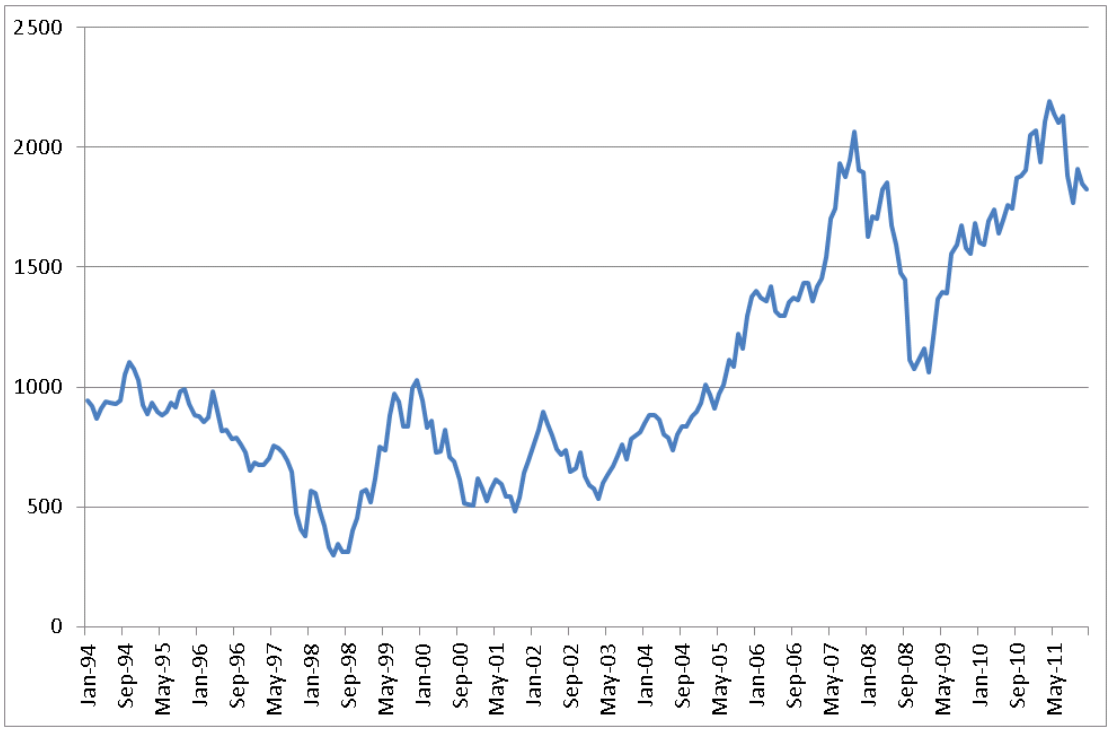
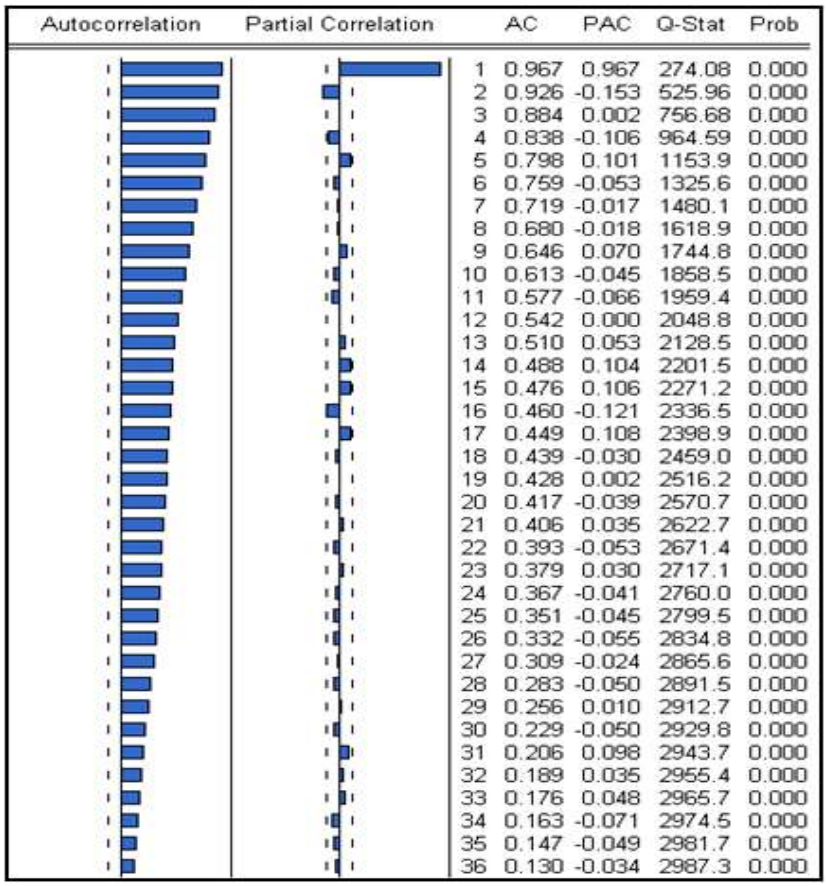


Figure 1. KOSPI from January 1994 to December 2011



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.967	0.967	274.08	0.000
		2	0.926	-0.153	525.96	0.000
		3	0.884	0.002	756.68	0.000
		4	0.838	-0.106	964.59	0.000
		5	0.798	0.101	1153.9	0.000
		6	0.759	-0.053	1325.6	0.000
		7	0.719	-0.017	1480.1	0.000
		8	0.680	-0.018	1618.9	0.000
		9	0.646	0.070	1744.8	0.000
		10	0.613	-0.045	1858.5	0.000
		11	0.577	-0.066	1959.4	0.000
		12	0.542	0.000	2048.8	0.000
		13	0.510	0.053	2128.5	0.000
		14	0.488	0.104	2201.5	0.000
		15	0.476	0.106	2271.2	0.000
		16	0.460	-0.121	2336.5	0.000
		17	0.449	0.108	2398.9	0.000
		18	0.439	-0.030	2459.0	0.000
		19	0.428	0.002	2516.2	0.000
		20	0.417	-0.039	2570.7	0.000
		21	0.406	0.035	2622.7	0.000
		22	0.393	-0.053	2671.4	0.000
		23	0.379	0.030	2717.1	0.000
		24	0.367	-0.041	2760.0	0.000
		25	0.351	-0.045	2799.5	0.000
		26	0.332	-0.055	2834.8	0.000
		27	0.309	-0.024	2865.6	0.000
		28	0.283	-0.050	2891.5	0.000
		29	0.256	0.010	2912.7	0.000
		30	0.229	-0.050	2929.8	0.000
		31	0.206	0.098	2943.7	0.000
		32	0.189	0.035	2955.4	0.000
		33	0.176	0.048	2965.7	0.000
		34	0.163	-0.071	2974.5	0.000
		35	0.147	-0.049	2981.7	0.000
		36	0.130	-0.034	2987.3	0.000

Figure 2. Sample autocorrelation and partial autocorrelation function of 1995 KOSPI

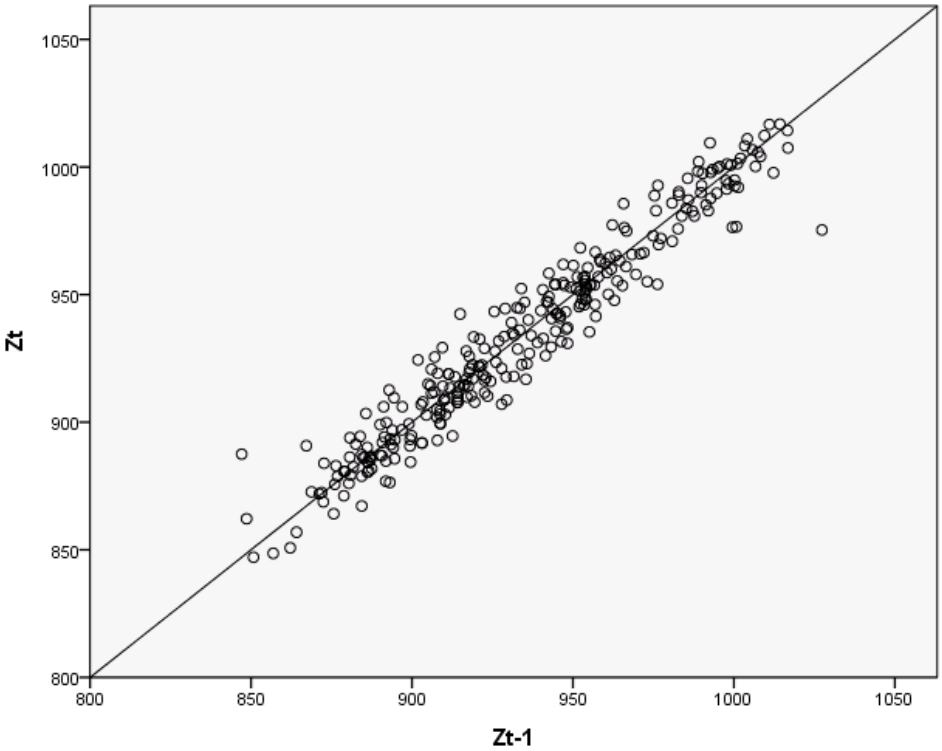
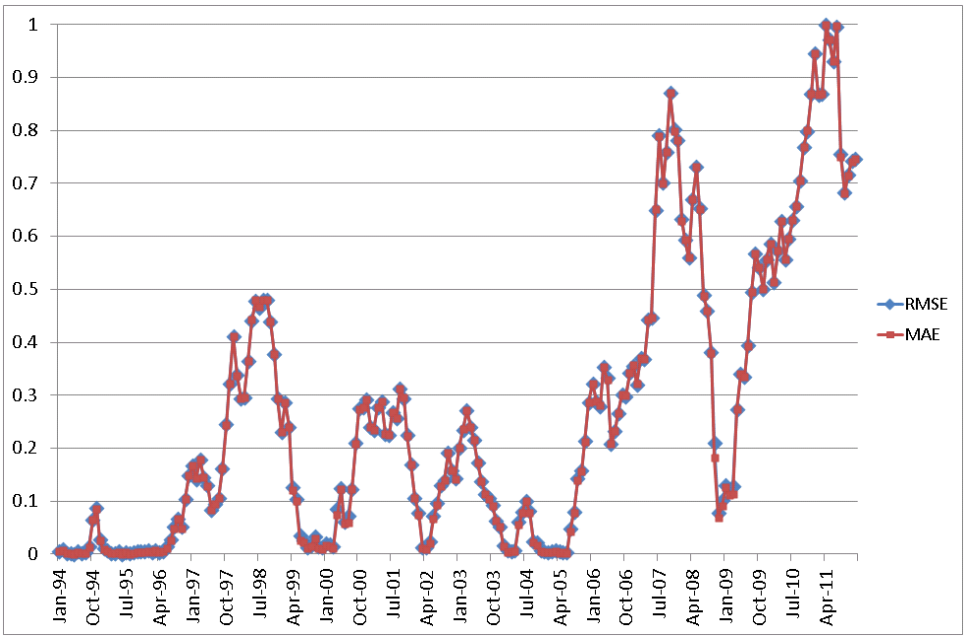


*Figure 3. Scatter plot of Z_{t-1} vs Z_t of 1995 KOSPI

Table 1. Dickey-Fuller test of 1995 KOSPI

OBS	TYPE	NAME	MSE	Intercept	DLAG_V	TAU	TREND	$p-$value
1	OLS		87.6248	25.054	-0.02715	-1.9493	1	0.30905
2	COV	Intercept	87.6248	169.744	-0.18132	-1.9493	1	0.30905
3	COV	DLAG_V	87.6248	-0.181	0.00019	-1.9493	1	0.30905

(a) Plot of $SMII_1$ (monthly RMSE and MAE)



(b) Plot of $SMII_1$ (monthly MAPE1 and MAPE2)

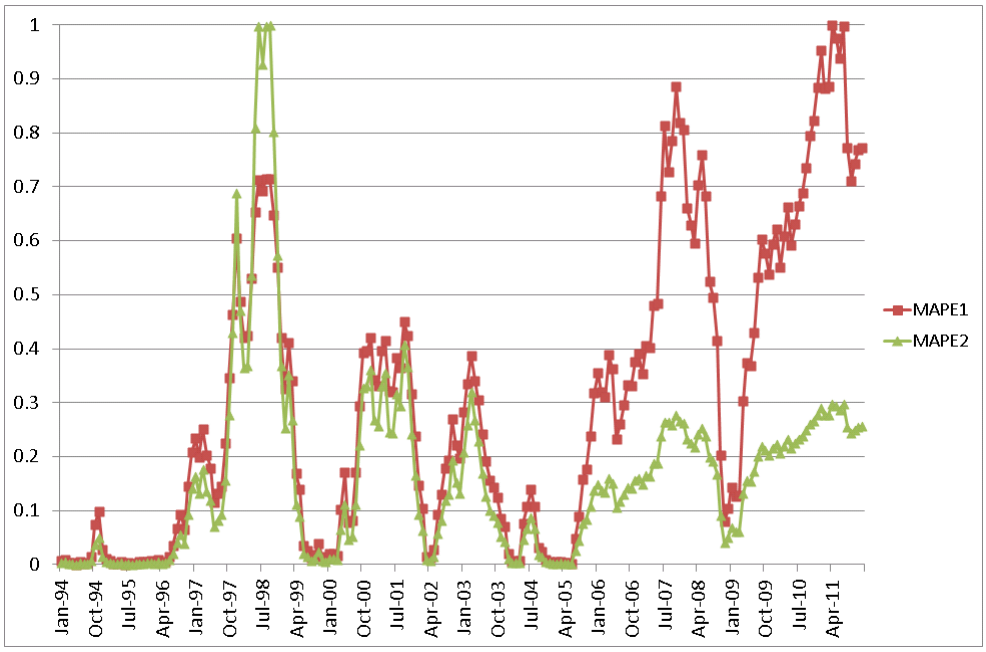


Figure 4. Plots of $SMII_1$ from January 1994 to December 2011

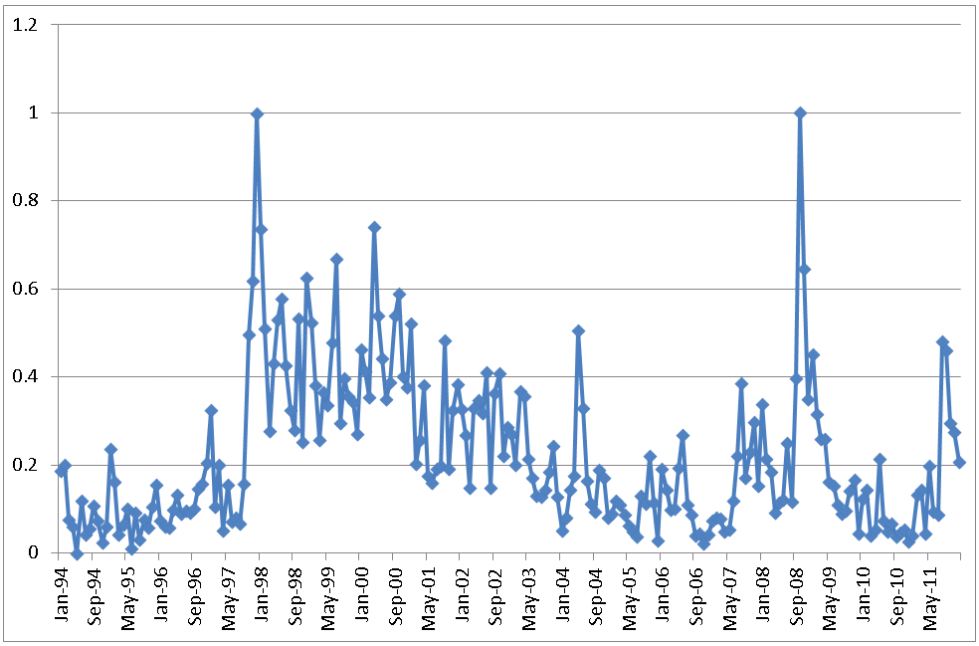
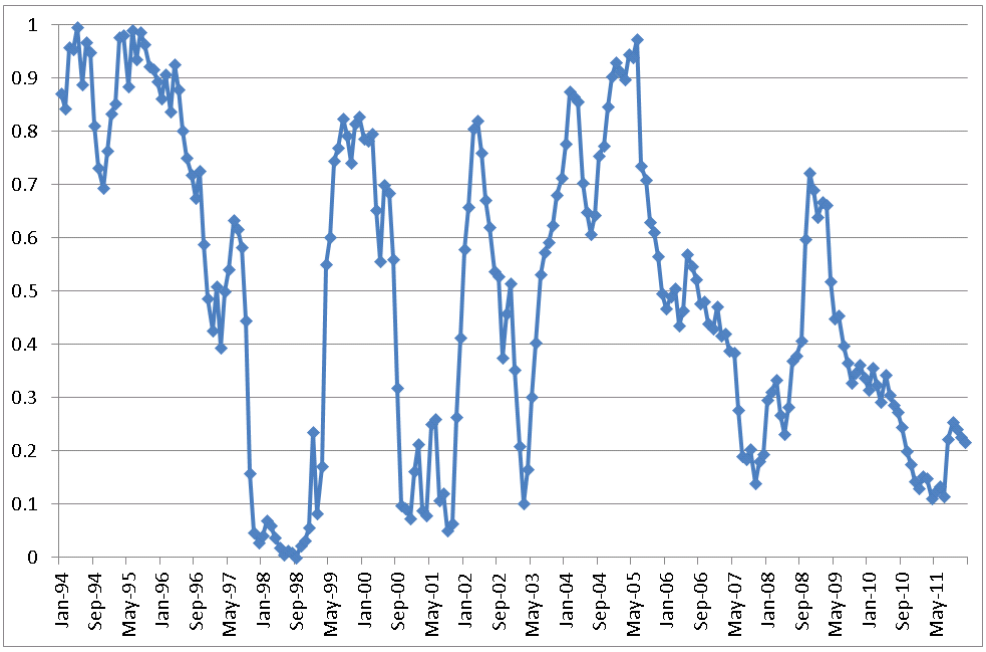


Figure 5. Plot of $SMII_2$ from January 1994 to December 2011 (monthly MAPE2)

(a) The $p-$value plot of $SMII_1$



(b) The $p-$value plot of $SMII_2$

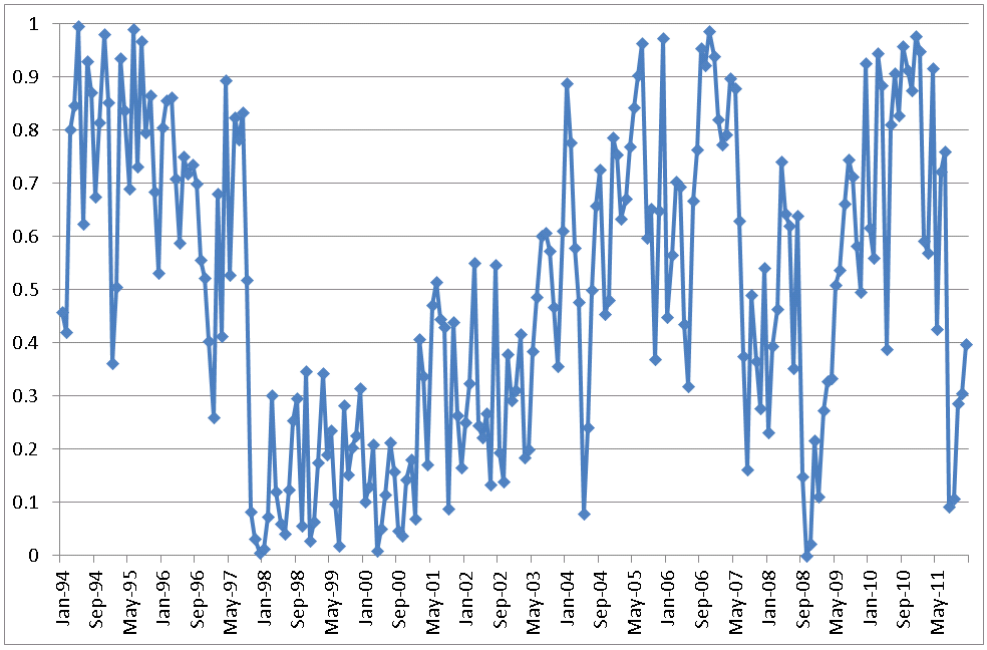
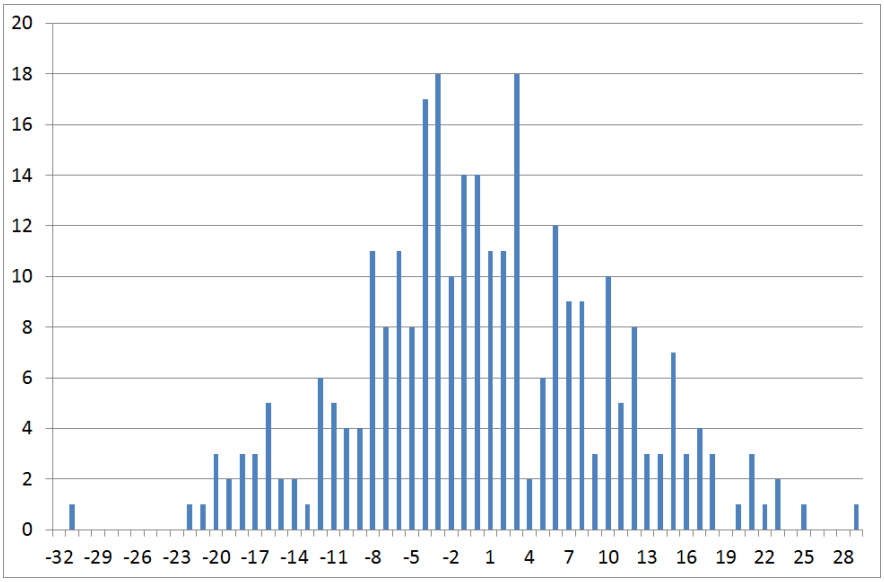


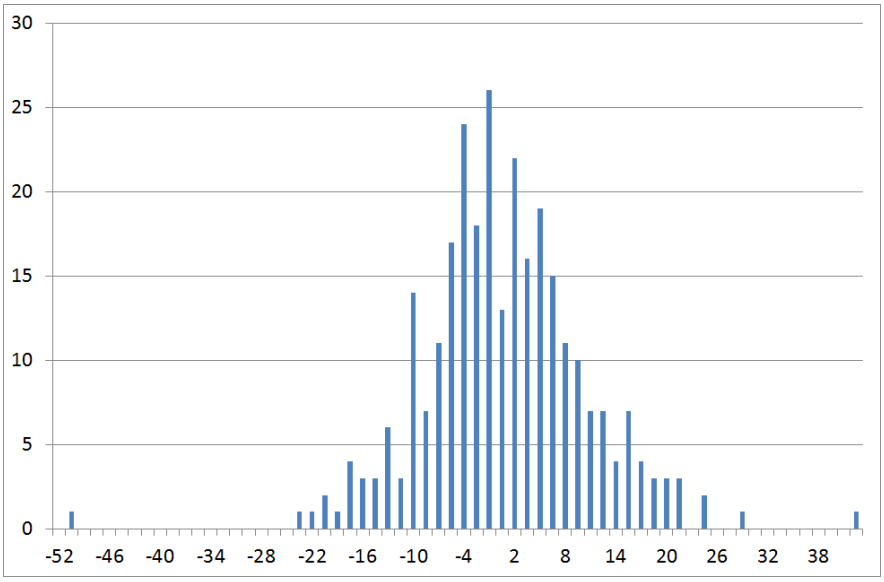
Figure 6. $p-$value plots of $SMII_1$ and $SMII_2$

(a) Residual histogram from ANN fitting



mean: -0.034 / s.e.: 9.743

(b) Residual histogram from AR(3) fitting



mean: -0.017 / s.e.: 9.856

Figure 7. Residual plots

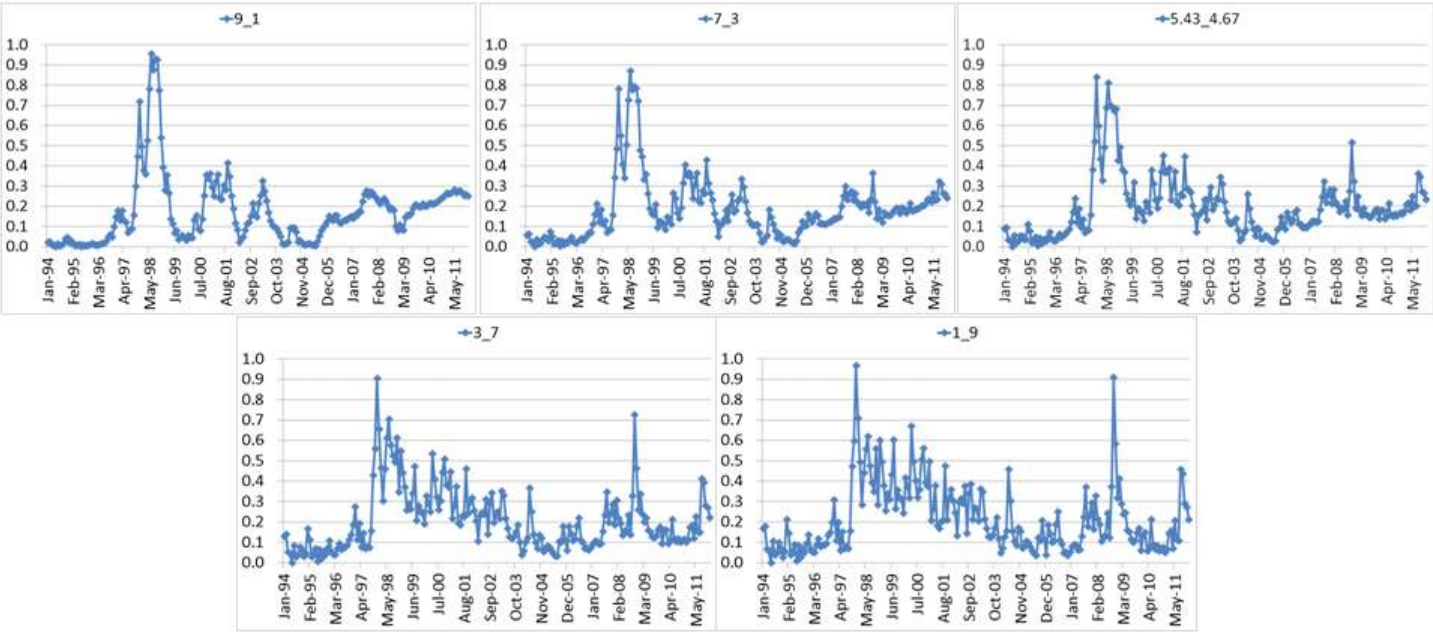


Figure 8. Plots of $iSMII = \hat{w}_1(T)SMII_1 + (1-\hat{w}_1(T))SMII_2$ for various $(\hat{w}_1(T), \hat{w}_2(T))$ s.

For instance, 9_1 means $\hat{w}_1(T) = 0.9$.

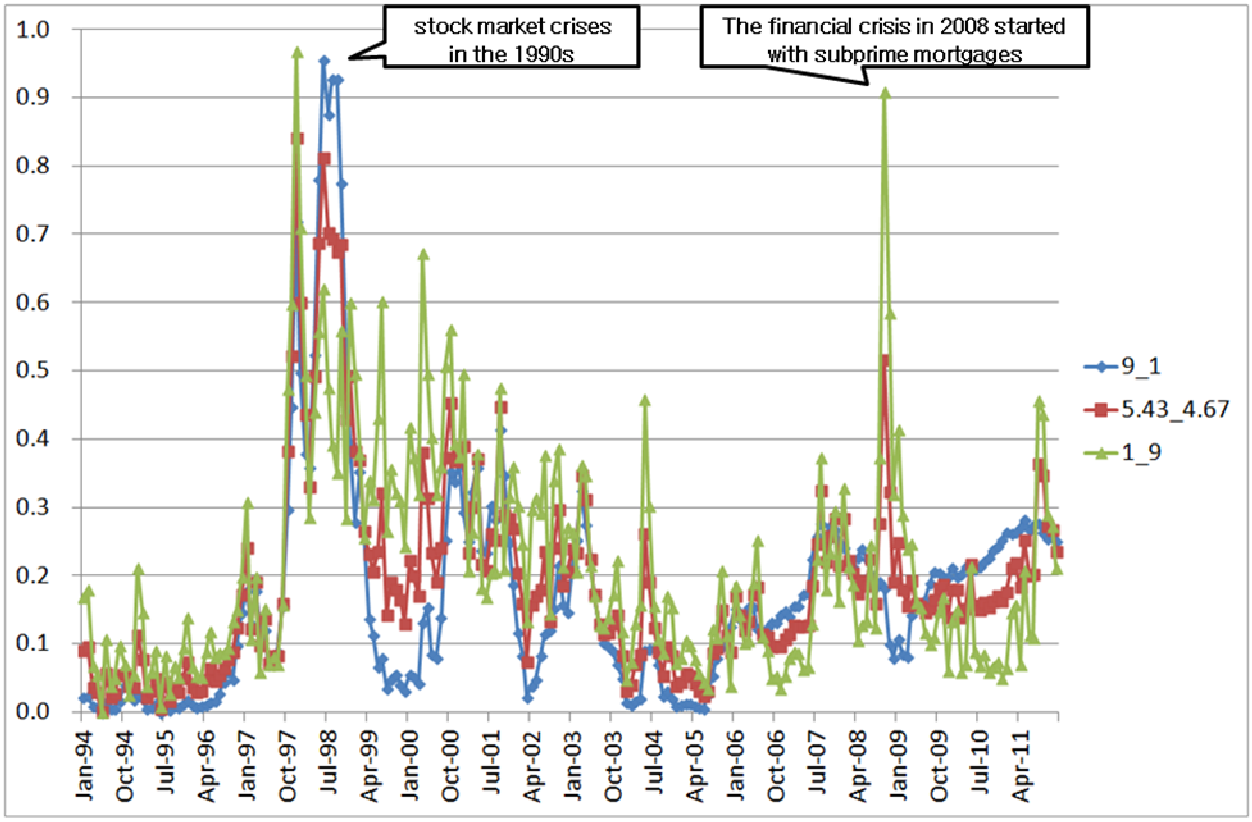


Figure 9. Plot of *iSMII* with the two major financial market crises

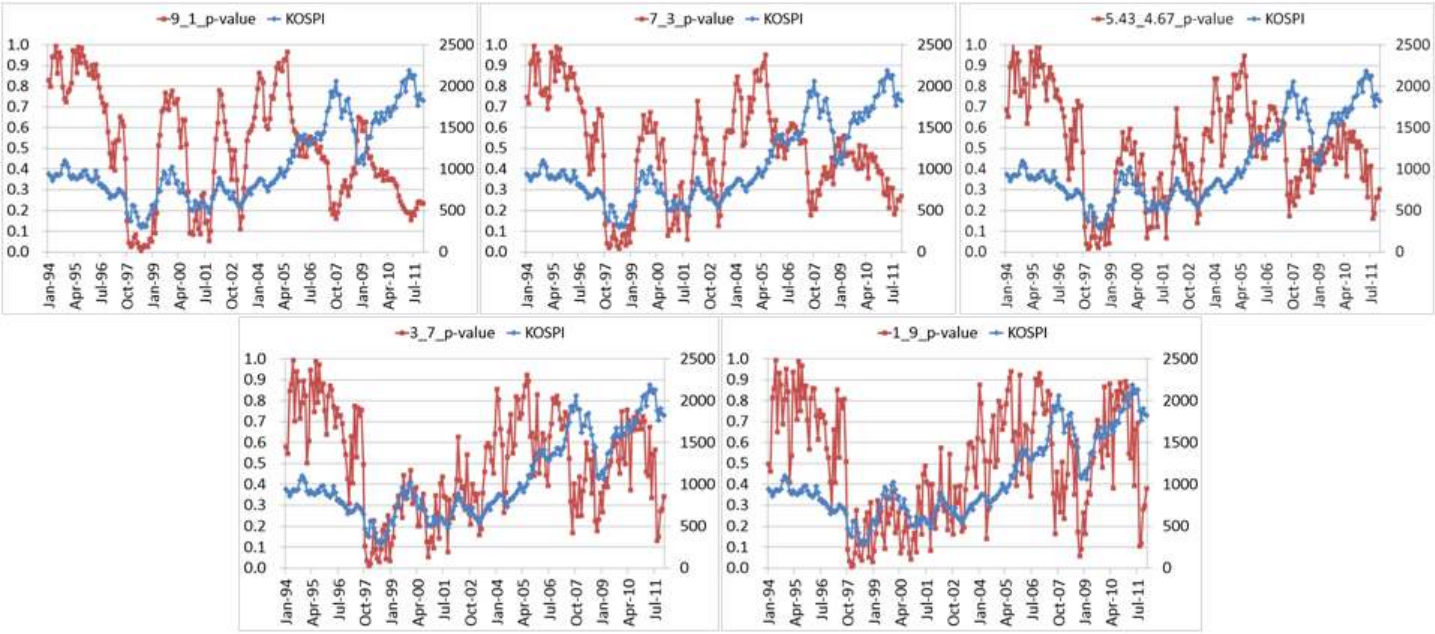


Figure 10. $p-$ value plots of $iSMII$ for various $(w_1(T), w_2(T))$

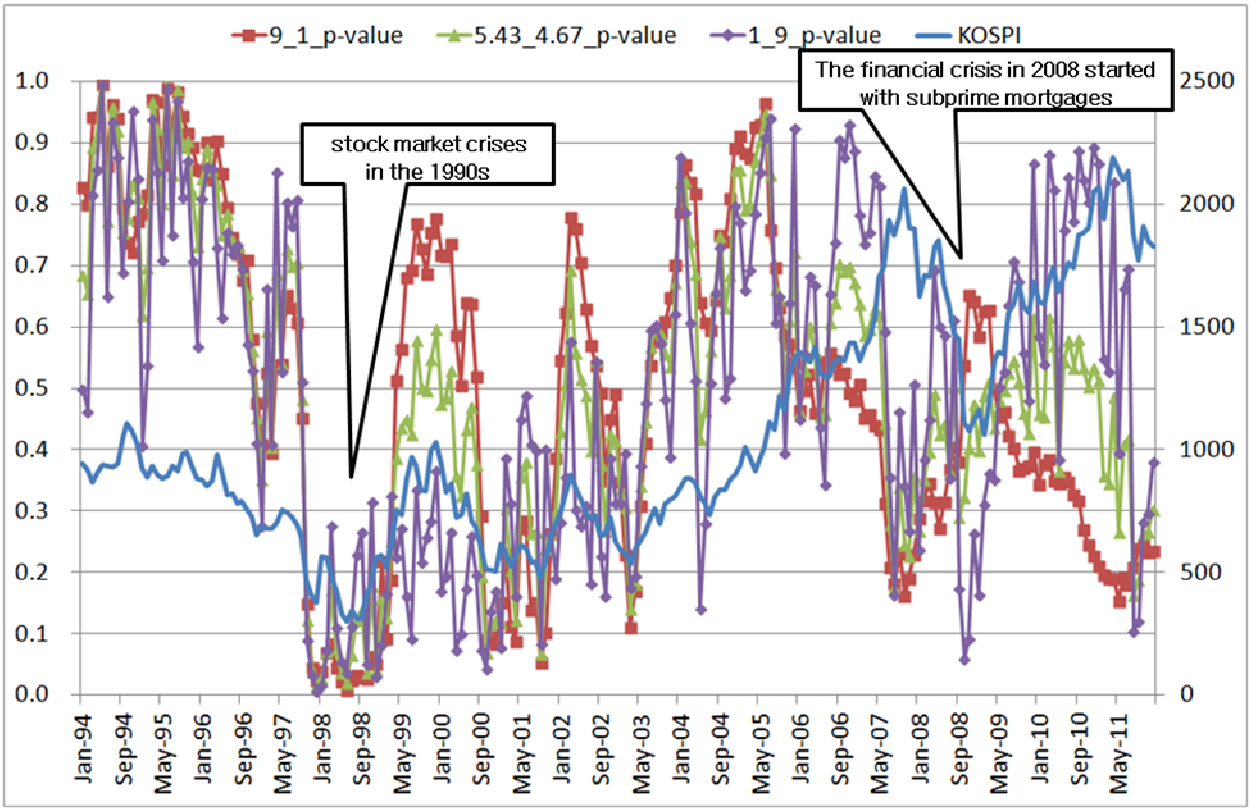


Figure 11. p – value plots of $iSMII$ for various $(w_1(T), w_2(T))$ and KOSPI

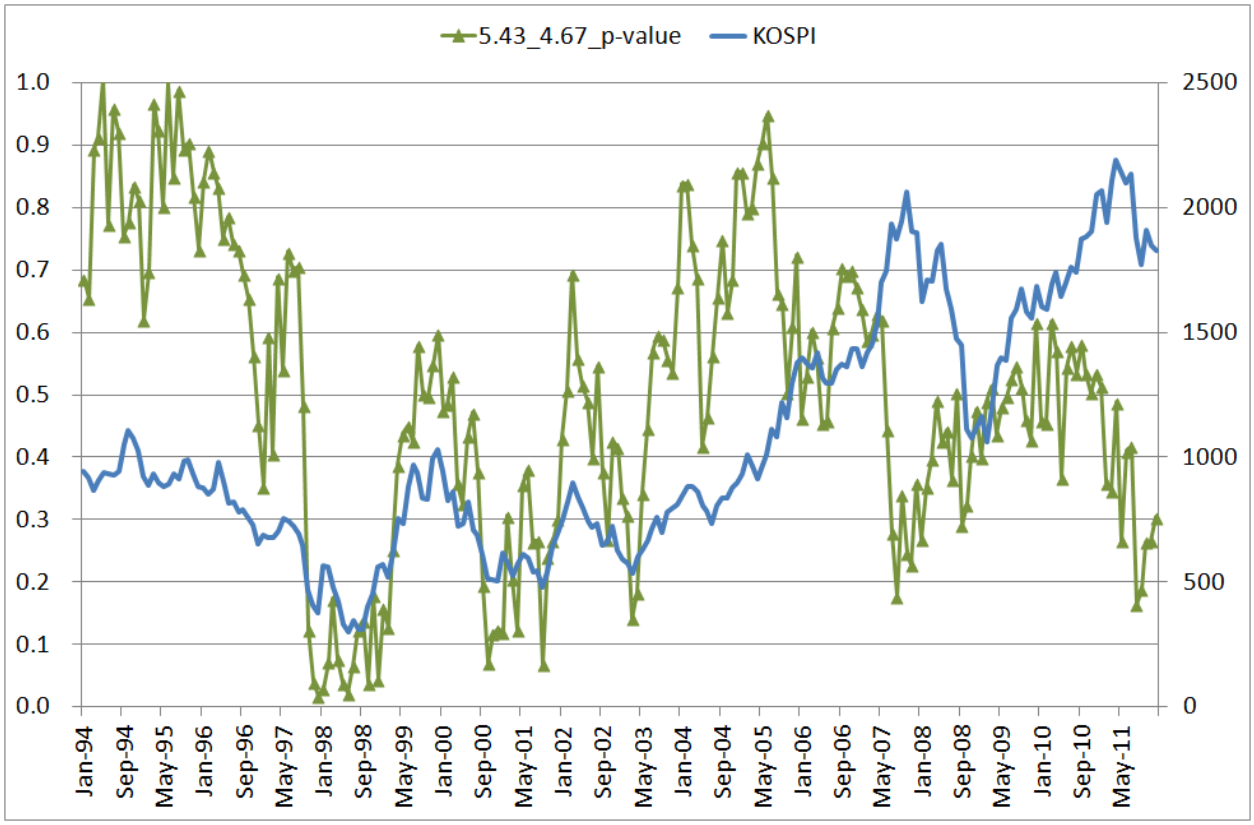


Figure 12. 5.43_4.67 p – value plot of *iSMII* and KOSPI