Université de Montréal
Faculté des études supérieures

Using the Modified FPE Criteria Forecasting the Realized Variance: a Statistical Analysis

Rapport de recherche présenté à la Faculté des études supérieures
en vue de l'obtention du grade de Maîtrise
en sciences économiques

Présenté par
Xuemei Huang

Département de sciences économiques
Faculté des arts et des sciences

Directeur de recherche : **Professeur Benoît Perron**

Décembre 2008

Sommaire

La variance réalisée (RV) joue un rôle crucial dans la finance puisqu'il présente une estimation convergente de la variance réelle. Par conséquent, il est important de choisir un critère approprié afin de prévoir la RV manifestant la caractéristique de mémoire longue. Dans ce rapport, un modèle d'approximation d'autocorrélation (AR) est utilisé pour la prévision, car il correspond au processus de mémoire longue. Pour l'amélioration de la performance des prévisions, selon le principe de l'erreur de prévision finale (FPE), Moon, Perron et Wang (2007) proposent deux nouveaux critères : le FPE1 modifié et le FPE2 modifié pour la détermination de l'ordre optimal, ainsi que prévoir plus efficacement les processus de mémoire longue.

Ces deux nouveaux critères sont utilisés pour de la prévision hors échantillon de la RV afin de sélectionner l'ordre approprié du modèle d'approximation AR. Tous les modèles sont exécutés en utilisant les méthodes des fenêtres de récurrence et les fenêtres de roulement pour les prévisions de *h* périodes. La comparaison avec les prévisions obtenues en utilisant les critères d'information classiques comme le critère d'information d'Akaike (AIC), le critère d'information bayésien (BIC) et le FPE tend à démontrer que le critère FPE1 modifié ou le critère FPE2 modifié est supérieur aux autres critères pour la prévision de RV à long terme.

Abstract

The realized variance (RV) plays a crucial role in finance for it yields a consistent estimation of the actual variance. Hence, it is important to choose an appropriate criterion to forecast the RV displaying the long memory characteristic. In this paper, an autoregressive (AR) approximation model is used for forecasting, since it fits the long memory process. For improving the performance of the forecasts, based on the final prediction error (FPE) principle, Moon, Perron and Wang (2007) proposed two new criteria: the modified FPE1 and the modified FPE2, for determining the optimal lag length thus more effectively forecasting the long memory process.

These two new criteria are used in the RV out-of-sample forecasting to select the suitable lags for the AR approximation model; all models are run under both the recursive windows and rolling windows for the *h*-step-ahead forecasts. The comparison of the forecasting performances of these two criteria and the classical information criteria such as the Akaike's information criterion (AIC), Bayesian information criterion (BIC) and FPE indicates that the modified FPE1 or the modified FPE2 criterion is superior to the other criteria in the RV forecasting for the long-range forecasts.

Remerciements

En premier lieu, je tiens à exprimer mes sincères et profonds remerciements à mon directeur de recherche, le professeur Benoît Perron, pour son encadrement, ses précieux conseils, sa grande disponibilité et sa patience, dont j'ai bénéficié tout au long de ce travail.

En second lieu, je remercie mes amis Weihao Sun, Yubo Guo et Simon Charland, pour leurs encouragements, leurs suggestions et la patience qu'ils ont eue pour la révision de ce rapport.

Enfin, je voudrais remercier ma famille et mes amies, pour leur soutien et leurs encouragements pendant mes études supérieures.

Table of Contents

List of Figures

List of Tables

1 Introduction

Measuring the variance is one of the central topics in financial analyses. The accuracy of estimating and forecasting variances for financial asset pricing, investment portfolio, risk management is of particular importance. Among the ways to evaluate the variances, the realized variance (RV), as demonstrated in the previous studies, presents a consistent measure of the variation of the actual variance.

RV is easily obtained from the high-frequency intra-period returns. It can be computed as the sum of squared high-frequency returns within a day. In conformity with appropriate conditions, RV is a consistent and highly efficient estimator of the return variance (Andersen et al.(2001), Barndorff-Nielsen and Shephard (2001, 2002)).

According to previous studies, the series of logarithmic realized variances (log RV) resembles a long memory process. To construct log RV, the autoregressive fractionally integrated moving average ($ARFIMA$) model is often selected. (Andersen et al.(2001, 2001a), Koopman, Jungbacker and Hol (2005)). However, in real practice, there is a low success rate in choosing a suitable ARFIMA model (Crato and Ray (1996)). Berk (1974), Beran, Bhansali, and Ocker (1998), Hong (1996), as well as Franses and Ooms (1997) argued that an autoregressive (AR) approximation model has superior performance in the prediction of long memory time series. Ray (1993a) also indicated that the AR approximation can be employed in the long-range forecasting of a long memory process.

Model selection, which is obtained by choosing a suitable order for the process, is the fundamental step for establishing any time series, and is the most important factor in achieving the forecasting accuracy. For better forecasting the RV, it is important to choose a suitable order of an AR approximation model. Indeed, the determination of right order, represented by the “k” of an $AR(k)$ approximation model fitted to long memory processes is crucial.

There are many methods for determining the order of an AR model. With the true model is infinite dimension, if an order is selected lower than the true order of the process, the estimated error will be very significant; that is, it is not consistent. In contrast, if a higher order is selected, an undesirable high variance will emerge (Quenouille (1949), Anderson (1963)); although theoretically with the increasing lag length of AR, autoregressive errors can be reduced, it still has a drawback of decreasing the degrees of freedom. Therefore, to achieve high accuracy of estimation, an appropriate order has to be selected.

Many different selection criteria have been employed in estimating the lag length of an AR process for a time series in the literature. Generally, there are two kinds of selection criteria: efficiency and consistency. In real practice, with the true model is finite dimension, the efficient criteria have the drawback of not being able to assist the consistent model selections; whereas, consistent criteria forgo the efficiency. Hence, all the information criteria have their advantages and disadvantages, and neither can be viewed as having superior performance over the others.

According to previous studies, the traditional criteria such as AIC and BIC are widely using for the selection of the long memory models. For instance, Sowell (1992b) used both the AIC and the BIC in his studies; Cheung (1990) exclusively used the AIC; Schmidt and Tschernig (1993) found that the BIC performs better than the other criteria; last but not least, Franses and Ooms (1997) also employed the AIC criterion.

To possibly find an alternative solution other than the traditional criteria, and improve the performance of forecast by the AR approximation methodology, Moon, Perron and Wang (2007) proposed the modified FPE1 criterion. Then, by considering the effective number of observations and the degrees of freedom adjustment for the estimated variance, the authors defined another new criterion: the modified FPE2.

According Moon, Perron and Wang (2007), the two new criteria obtained by applying the suitable value of fractional differencing number d can improve the forecasting accuracy and reduce the lag length of the model. Namely, it can improve both the efficiency and simplicity of using the AR approximation model for long memory series.

In the following analysis, the modified FPE1 and the modified FPE2 are compared to the classical information criteria in the out-of-sample forecasting of RV. Both the recursive windows and rolling windows with *h*-step-ahead forecasts are employed in this paper. The empirical results show that comparing to the performances of the classical information criteria; the modified FPE1 or the modified FPE2 has the best performance for forecasting the RV in the long term.

The paper is organized as the following. Section 2 discusses the concept of RV. Section 3 defines the *h*-step-ahead forecast by using the AR approximation. Section 4 presents the formularies and differences of the information criteria which include the two new criteria: the modified FPE1 and the modified FPE2. Section 5 reports the method to estimate the fractional differencing number *d*. Section 6 presents a theoretical background of forecasting. Section 7 reports the empirical application. Finally, section 8 presents the summary of findings and the according conclusions.

2 Realized Variance

2.1 Realized Variance Theory

In the empirical studies, the subject of how to model and forecast financial market variances has been widely researched and documented. Numerous papers have argued that one could measure latent variance by RV (Merton (1980), Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002)), hence RV has been used extensively in the financial field.

The first formal presentation of RV was by Merton (1980). Merton (1980) found that during a fixed period of time, under the condition of sufficiently high frequency, summing the value of the squared high-frequency returns, an independent random variable with the same distribution of the variance can be accurately estimated. In his research, the data of daily volatility were used to estimate the monthly stock volatilities; and a formal presentation of RV was therefore achieved.

Many researches were conducted in this area: Taylor & Xu (1997) and Andersen et al. (2001) studied the 10-year DJIA30 index by using five-minute data to explore the characteristics of the RV; Areal and Taylor (2002) studied RV in the futures of FTSE-100 index; Bamdorff-Nielsen and Shephard (2002) achieved the asymptotic distribution of RV in the US dollar to German Deutschmark daily exchange rate; Oomen (2003, 2004) considered the autoregressive nature of high-frequency data series and examined the modeling and characteristic of RV by using 10-year FTSE-100 index data; and Koopman et al. (2005) used the ARFIMA model to forecast the RV of S&P 100 stock index. Their results showed that out-of-sample predictability of the ARFIMA model is superior to the other models.

2.2 Long Memory ARFIMA(p,d,q) Model

According to previous studies, the log RV series can be modeled by a long memory processes. For instance, using an ARFIMA to model the long memory properties is presented in Andersen et al. (2001) and Andersen et al. (2003). The log RV is employed rather than RV itself is because that log RV is approximately normally distributed. Also, based on previous literatures, it is indicated that the log RV has better forecasting performance than RV.

Suppose that x_t represent a series of log RV. Following the logic of abovementioned studies, an $ARFIMA(p,d,q)$ model is employed. x_t satisfies the following equation (Theorem 3 of Hosking (1996)):

$$\phi(L)(1-L)^d (x_t - \mu) = \theta(L)e_{x,t} \tag{1}$$

where

- d $\equiv$ the fractional differencing number which called the memory parameter ($0 < d < 0.5$);
- p $\equiv$ the number of autoregressive lags;
- q $\equiv$ the number of moving average lags;
- L $\equiv$ the backward-shift operator (lag operator) such that , $Lx_t = x_{t-1}$;

  $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$;

  where $\phi(L)$ and $\theta(L)$ have distinct zeros and roots lie outside the unit circle;
- $(1-L)^d$ $\equiv$ the fractional differencing operator

  $$(1-L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - \cdots = \sum_{j=0}^{\infty} \pi_j L^j ,$$

  $$\pi_{\mathrm{j}} = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} ,$$ Γ is the gamma function;
- μ $\equiv$ the mean of x_t, as $E(x_t) = \mu$;
- $e_{x,t}$ $\equiv$ white noise, namely $E(e_{x,t}) = 0$, $E(e_{x,t}^2) = \sigma^2$, and $E(e_{x,t}e_{x,s}) = 0, \forall t \neq s$.

In particular, the $ARFIMA(p,d,q)$ model uses parameters p and q to describe the short memory process, and uses the parameter d to reflect the long memory process.

Generally, there are two methods to estimate the $ARFIMA(p,d,q)$ model. The first method employs a two-step estimation, whereas the second employs a simultaneous (one-step) estimation. The method with two-step estimation is commonly used.

For the two-step method, firstly the fractional differencing number, d, is estimated and denoted as $\hat{d}$. Afterwards, $\hat{d}$ is used to transform the series into a standard ARMA model. Then several information criteria are employed to estimate the parameters p and q.

The second method does not require separate estimations on the value of d and the parameters of the ARMA model. It only needs a step to estimate all the parameters, which is therefore called the full maximum likelihood estimation (Hosking (1981) and Sowell (1987)).

Since it is possible for a data set to have both short and long memory, distinguishing the behavior between the two processes is difficult. Consequently, it may pose potential model selection difficulties (Ray, 1993a); as a result, there is a low success rate in choosing a suitable ARFIMA model (Crato and Ray (1996)).

2.3 Using AR Approximation Model to Represent ARFIMA Process

Theoretically speaking, when constructing an ARFIMA model, the estimation of the fractional differencing number, d, is very complicated, and may not be very accurate, especially in finite samples when d is close to 0.5.

Because of its relative simple structure of estimating the parameters, AR model has a comprehensive range of applications in the fields such as economics, finance, and engineering. According to the literature, ARFIMA process can be approximated well by an AR model (Ray (1993a), Crato and Ray (1996)). Furthermore, based on previous studies conducted by Berk (1974), Beran, Bhansali, and Ocker (1998), Hong (1996), as well as Franses and Ooms (1997), it is observed that an AR approximation model has superior performance in the prediction of long memory time series. Ray (1993a) also indicated that the AR approximation could be employed in the long-range forecasting of a long-memory process. Thus, the AR approximation method is used for predicating the long memory processes in the following study.

Based on Moon, Perron and Wang (2007), the expression of AR approximation model that represents the ARFIMA process x_t, is:

$$\mathrm{B(L)}x_t = \beta_0 + e_{x,t}$$

where $\mathrm{B(L)}x_t = (1-L)^d \phi(L)\theta(L)^{-1} = 1 - \sum_{i=1}^{\infty} \beta_i L^i$; and

$E(x_t) = 0$.

3 The *h*-Step-Ahead Forecast Using AR(k) Approximation

Forecasting model is a frequently used econometric tool to predict the future by effectively using historical information. Ensuring the accuracy of the time series forecasts is fundamental for making any further interpretations.

Based on Moon, Perron and Wang (2007), for forecasting the series x_{t+h}, the usual *h*-step-back substitution can be used. The x_{t+h} can be represented as a function of $\{x_t, x_{t-1}, \dots\}$ as

$$x_{t+h} = e_{x,t+h,h} + \beta_{0,h} + \sum_{j=0}^{\infty} \beta_{j+1,h} x_{t-j} \quad .$$

However, in practice, an infinite $AR(\infty)$ model should be truncated to a finite $AR(k)$ to approximate the ARFIMA process.

Suppose that the lag k and the coefficients $\beta_{0,h}, \beta_{1,h}, \dots, \beta_{k,h}$ are known. Then the natural *h*-step-ahead forecast of x_{T+h} at time T with observations $X_{T,k}$ is

$$x_{T+h|T,k} = \beta_{0,h} + \beta_{1,h} x_T + \cdots + \beta_{k,h} x_{T-k+1} \quad .$$

However, in real practice, these coefficients are typically unobservable. To forecast $x_{T+h|T,k}$, first of all, it is necessary to estimate the coefficients $\beta_{j,h}$ (where $j = 0, \dots, k$) based on the historical data.

In this paper, the ordinary least squares (OLS) method is chosen for estimating the coefficients, for its estimation is relatively simple and with high accuracy. Also, the OLS estimations are considered to be consistent.

After estimating the coefficients $\beta_{j,h}$, the estimated values, denoted by $\hat{\beta}_{j,h}$ (where $j = 0,\ldots,k$) are used in the *h*-step-ahead forecasting function. Hence, $\hat{x}_{T+h|T,k}$ can be express as:

$$\hat{x}_{T+h|T,k} = \hat{\beta}_{0,h} + \hat{\beta}_{1,h} x_T + \cdots + \hat{\beta}_{k,h} x_{T-k+1}.$$

4 Information Criteria

Information criteria are formed based on solid theoretical foundation and good statistical properties. They are standard tools for model selection. Specifically, the information criteria are used to select appropriate number of lags when constructing a model. It evaluates the selected order from two aspects: the goodness of fit and the model complexity. In brief, holding everything else constant, under the prerequisite of the same complexity, the model with the higher goodness of fit is better; for similar goodness of fit, the simpler model is better.

Many different information criteria are proposed in the literatures. For example, there are the Akaike's information criterion (AIC) (Akaike (1973)), corrected Akaike information criterion (AICC) (Hurvich and Tsai (1989)), Bayesian information criterion (BIC) or Schwarz information criterion (SIC) (Schwarz (1978)), final prediction error (FPE) (Akaike (1969)), Hannan-Quinn criterion (HQC) (Hannan and Quinn (1979)) and Mallow's C_p (Mallows (1973)). Among all these criteria, the AIC and BIC are very commonly used.

In general, the selection criteria are classified depends on the efficiency and consistency. The efficient criteria (e.g. AIC, AICC, FPE and C_p) emphasize on selecting the model that can produce the least mean square prediction error; hence it helps to create the best finite-dimensional approximation model when the true model is of infinite dimension. In contrast, the consistent criteria (e.g. BIC and HQC) are typically used assuming data coming from a finite order autoregressive process, thus asymptotically select the true order of the process when the true model is of infinite dimension. However, there is normally a trade off between the consistency and the efficiency when the true model is finite dimension.

In the following section, the definitions and relative advantages and disadvantages of the AIC, BIC, FPE, modified FPE1 and modified FPE2 criteria are presented.

4.1 Classical Information Criteria

AIC criterion

AIC criterion is initially proposed by Akaike (1973) and was successfully applied to determine the lag length of an AR model. This method can also be used for selecting the order of an ARMA model.

AIC is an unbiased estimator of the expected Kullback-Leibler divergence between the fitted and true model. When the series is normally distributed, the AIC criterion can be expressed as:

$$AIC(k) = \ln(\hat{\sigma}_k^2) + \frac{2k}{T}$$

Where

$\hat{\sigma}_k^2$ is an estimator of the regression error variance $\hat{\sigma}^2$ for the kth order of the process;

k is the order of the AR process;

T is the sample size.

BIC criterion

The BIC criterion is proposed by Schwarz (1978). It asymptotically approximates the integrated likelihood of the model. The BIC criterion can be expressed as:

$$BIC(k) = \ln(\hat{\sigma}_k^2) + \frac{k}{T}\ln(T)$$

FPE criterion

Akaike (1969) proposed FPE criterion to select the lags that minimize the one-step-ahead mean squared prediction error for the best order of an AR model.

For a time series x_t, the $AR(k)$ is

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_k x_{t-k} + \varepsilon_t \ ,$$

where $E(x_t) = 0$; $t = 1,...,n$; $k = 1,...,m$ and $n > m$.

Akaike used $\phi(k) = E[x_{t+1} - \hat{x}_{t+1}]^2$ as the objective function to find $k = p$ that gives the minimum $\phi(k)$ value. Namely, the minimum mean squared prediction error. As a result, the *k* is the optimal estimation of the lag length. It follows:

$$FPE(k) = \hat{\sigma}^2(k)\frac{T+k}{T-k}$$

4.2 Modified FPE Criteria

Modified FPE1 criterion

To improve the forecast performance of the AR approximation model, Moon, Perron and Wang (2007) proposed a new criterion: the modified FPE1. This criterion focus on the forecasting performance, thus the efficiency of the model selection is more important.

By following the principle of minimizing the one-step-ahead mean squared prediction error, FPE ensures efficiency. However, the data generating process (DGP) that Akaike (1969) considered to structure the FPE criterion was actually a short memory process. As a consequence, Moon, Perron and Wang (2007) extend the FPE criterion by considering the characteristics of a long memory process while applying the value of fractional differencing number *d*.

Instead of minimizing the one-step-ahead prediction error variance (the mean squared error), the modified FPE1 criterion minimizes the *h*-step-ahead prediction error variance. At $h = 1$, a special case is encountered: the modified FPE1 minimizes the one-step-ahead mean squared error.

Based on Moon, Perron and Wang (2007), suppose that the time series x_t is an AR approximation process, the *k*-lag AR approximation of x_{t+h} has the expression as:

$$x_{t+h} = \beta_{0,h} + \beta_{1,h} x_t + \cdots + \beta_{k,h} x_{t-k+1} + e_{x,t+h,h}$$

Define $\sigma_h^2 = \mathrm{Var}(e_{x,t+h,h})$, based on the general theory of OLS, σ_h^2 can be estimated by

$$\hat{S}_{h,k}^2 = \frac{1}{T-k} \sum_{t=k+1}^{T-h} \hat{e}_{x,t+h,h}^2 \qquad (2)$$

Moon, Perron and Wang (2007) proved that as $T \to \infty$, the *h*-step-ahead mean squared prediction error is:

$$E\left[x_{T+h} - \hat{x}_{T+h \mid T,k}\right]^2 \sim \sigma_h^2 \left(1 + \left(\frac{k}{T}\right)^{1-2d}\right) \quad .$$

In this formula, there are two unknown parameters: σ_h^2 and *d*. σ_h^2 can be estimated by $\hat{S}_{h,k}^2$. For large *T*s, it follows $T\hat{S}_{h,k}^2/\sigma_\mathrm{h}^2 \sim \chi_{T-k}$; therefore $\frac{T}{T-k}\hat{S}_{h,k}^2$ can be used to replace σ_h^2. Then, $\hat{d}$, estimated by using the GPH method,[1] can be used to determine the mean squared prediction error of x_{T+h}.

Hence, the modified FPE1 criterion minimizes the mean squared prediction error for *h*-step-ahead prediction:

$$FPE_h^{M1}(k) = \frac{T}{T-k}\ \hat{S}_{h,k}^2 \ \left(1 + \left(\frac{k}{T}\right)^{1-2\hat{d}}\right),$$

where ($0 < \hat{d} < 0.5$) and ($0 \le k < T$)

[1]The method to estimate *d* will be presented in the following section.

If $\hat{d}=0$ and $h=1$, $FPE_1^{M1}(k)=\hat{S}_{1,k}^2\frac{T+k}{T-k}$; this is the same expression as the FPE criterion. In addition, for $0<\hat{d}<0.5$ and $0\leq k<T$, if $\hat{d}$ increases, $FPE_h^{M1}(k)$ increases too.

Using this criterion to select the lag length of an AR approximation model, it follows the same intuition as the FPE: when $FPE_h^{M1}(k)$ is at its minimum, *k* is the suitable lag length for an AR approximation model.

From the expression of the modified FPE1, it is observed that both of the first term $\frac{T}{T-k}$ and the third term $\left(1+\left(\frac{k}{T}\right)^{1-2\hat{d}}\right)$ are increasing functions of *k*. The second term $\hat{S}_{h,k}^2$ can be estimated by equation (2): $\hat{S}_{h,k}^2=\frac{1}{T-k}\sum_{t=k+1}^{T-h}\hat{e}_{x,t+h,h}^2$. According to Wooldridge (2006), when *k* increases, the sum of squared residuals does not increase; hence the value of $\hat{S}_{h,k}^2$ decreases with *k* increasing. Therefore, when the *k* increases, given the size of the data, the second term of the modified FPE1's expression is monotonously decreasing; and the first and the third terms, namely the "penalty terms", are monotonously increasing.

If the second term of the modified FPE1 plays a decisive role, when *k* increases, the value of the modified FPE1 will decrease. On the contrary, if the first and the third term of the modified FPE1 are decisive, when *k* increases, the value of the modified FPE1 will increase. When the *k* reaches a certain value *p*, the modified FPE1(*k*) reaches its minimum value. Thus, the value *k* = *p* is considered the optimal order of an AR process.

Modified FPE2 criterion

Moon, Perron and Wang (2007) used the effective number of observations and a degrees of freedom adjustment of the estimated variance to define another modified FPE criterion, namely modified FPE2:

$$FPE_h^{M2}(k) = \frac{T}{T-k}\left[\hat{S}_{h,k}^2 + \hat{S}_{h,k_{\max}}^2\left(\frac{k}{T}\right)^{1-2\hat{d}}\right]$$

Since $\hat{S}_{h,k}^2 \le \hat{S}_{h,k_{\max}}^2$, the value of $FPE_h^{M2}(k)$ is larger than that of $FPE_h^{M1}(k)$ for the same value of k. Furthermore, because the expression of the FPE2 is equivalent to

$$FPE_h^{M2}(k) = \frac{T}{T-k}\hat{S}_{h,k}^2\left[1 + \frac{\hat{S}_{h,k_{\max}}^2}{\hat{S}_{h,k}^2}\left(\frac{k}{T}\right)^{1-2\hat{d}}\right],$$

it can be observed that $\left[1 + \frac{\hat{S}_{h,k_{\max}}^2}{\hat{S}_{h,k}^2}\left(\frac{k}{T}\right)^{1-2\hat{d}}\right] \ge \left(1 + \left(\frac{k}{T}\right)^{1-2\hat{d}}\right)$. Hence, the modified FPE2 has a heavier penalty term than the modified FPE1. Consequently, the lag order selected via the modified FPE2 is smaller than the lag order selected via the modified FPE1.

4.3 A Comparison of the Performance of the Two Modified FPE Criteria and the Classical Information Criteria

The quality of the performance of a criterion relies not only on the forecasting accuracy but also on the spirit of parsimony.

Since an effective model is the appropriate balance between the underfitting and the overfitting for a given sample size, to select a better model, the concept of parsimony must be considered. Specifically, with a certain level of forecasting accuracy, a good model should be as simple as possible. Generally, simple models are easier to estimate, forecast and analyze. Keeping everything else constant, a simpler, more parsimonious model with fewer estimated parameters is better than a complex one.

When different criteria are used to select the lag length of the model, for equivalent prediction accuracy, the criterion with the smallest lag length is preferred, thus is considered to have a superior performance. In the following paragraphs, the five criteria will be analyzed in comparison with each other to determine the criterion producing the smallest lag length.

Base on Moon, Perron and Wang (2007), both the modified FPE1 and the modified FPE2 are developed from the FPE; hence, the two new criteria are considered to be efficient criteria. As the fractional differencing number d is employed in the two new criteria; for $0 < \hat{d} < 0.5$, it follows $\left(1+\left(\frac{k}{T}\right)^{1-2\hat{d}}\right) > \left(1+\frac{k}{T}\right)$, hence, the penalty term of the modified FPE1 is greater than that of the FPE. As a result, the lag length selected by using the modified FPE1 is smaller than the one selected by using the FPE. Base on section 4.2, the lag order selected via the modified FPE2 is smaller than the one selected via the modified FPE1. Consequently, the modified FPE2 will always select a smaller lag length than the FPE too.

In regards to the comparison of the FPE and the AIC, it can be demonstrated that the model order selected using the FPE criterion never exceeds the one using the AIC. For an $AR(k)$ process, the lag length selected using the FPE and the AIC are the same when the number of observation is large enough but still finite. This is because the two criteria are asymptotically equivalent with this condition (Jones (1974)).

BIC is known as the asymptotically consistent criterion (Haughton (1988) (1989)), and has the same first term as the AIC. However, instead of the penalty term $\frac{2k}{T}$, it uses $\frac{k}{T}\ln(T)$. Hence, for any $T > \exp(2)$, the penalty term of BIC is larger than the one of AIC. Thus the lag length selected via BIC is smaller than the lag length selected via AIC.

When comparing the modified FPE1 or FPE2 to the BIC, although the modified FPE1 and FPE2 depend on the estimated fractional differencing number *d* and the number of observations, it is difficult to compare their penalty terms to the BIC's. Thus, it can not be established which one of these criteria can select the smallest lags.

In conclusion, the lag lengths selected from the above-mentioned criteria have the following relations:

1. $\text{FPE2} \leq \text{FPE1} < \text{FPE} \leq \text{AIC}$; and
2. $\text{BIC} < \text{AIC}$

Hence, the modified FPE2 criterion or BIC criterion can be the one that determines the smallest lag length for an AR approximation model.

5 Estimation of *d*

To estimate the parameters of a long memory process, a number of methods have been developed. In general, they can be classified into two groups: the parametric methods and the semi-parametric methods. Specifically, the parametric methods comprise the Gaussian maximum likelihood (Dahlhaus (1989)), Whittle likelihood (Fox and Taqqu (1986) and so on. The semi-parametric methods include GPH estimator (Geweke and Porter-Hudak (1983)) and Gaussian semiparametric estimate (Künsch (1986)).

Specifically, Geweke and Porter-Hudak (1983) proposed a widely used semi-parameter method: periodogram regression (GPH) to estimate the degree of the fractional differencing, *d*. This method is relatively simple thus commonly used.

The GPH method is based on a spectral density function and with the following periodogram regression:

$$\ln I_n(\omega_j) = a - 2d \ln(2\sin\frac{\omega_j}{2}) + \mu_j \qquad (j = 1,2,\cdots, g(n)),$$

Where,

$I_n(\omega_j)$ is the periodogram of the series when $\omega_j = \frac{2\pi j}{n}$ $(j = 1,2,\ldots, g(n))$,

$I(\xi) = \frac{|\sum_{t=1}^{n} e^{it\xi}(x_t - \bar{x})|^2}{2\pi n}$, where $\bar{x}$ is the mean of x_t.

Select the appropriate number of frequency of regression, $g(n)$, make $\lim_{n\to\infty} g(n) = \infty$, when conditions of $\lim_{n\to\infty}\frac{g(n)}{n} = 0$ and $\lim_{n\to\infty}\frac{\ln(n)^2}{g(n)} = 0$ are satisfied, the estimation of *d*, by using the OLS method, is consistent and $(\hat{d} - d)/s(\hat{d}) \to N(0,1)$.

Robinson (1995a) proved the consistency when $0 < d < 0.5$. Hassler (1993a) proved the consistency and asymptotic normality when $e_{x,t}$ of equation (1) is a process with normal distribution. Hassler's study also indicates that the GPH estimation is effective when the data are not normally distributed and with presence of ARCH effects.

When using the GPH for estimating *d*, the selection of the appropriate number of frequency $g(n)$ is very critical. If $g(n)$ is too large, it will cause the estimated amount to be too sensitive to a short term memory. Conversely, if $g(n)$ is too small, it will lead to inaccurate estimation. Therefore, GPH suggests that the condition, $g(n) << n$, must be satisfied.

Given the great significance of the ultimate selection of $g(n)$ for the estimation of *d*, usually different number of frequency such as $g(n) = n^{0.5}$, $g(n) = n^{0.6}$ and $g(n) = n^{0.8}$ are tested. Then, the resulting values of *d* are compared.

6 Forecast Evaluation

The forecasting methodology can be separated into the in-sample and the out-of-sample approaches. Considering the fact that the forecasting performance depends on the predictive accuracy rather than the goodness-of-fit of the historical data model, Fildes and Makridakis (1988) demonstrated that the out-of-sample approach increases the forecasting capabilities of the model. Furthermore, selecting the optimal model via the in-sample approach can not perfectly forecast the out-of-sample data even the in-sample approach may decrease the forecast error. As a consequence, the out-of-sample approach is selected to compare the forecasting performances in this paper.

In order to employ the in-sample-data to estimate the parameters, the out-of-sample approach is operated by first dividing the sample into two parts, and using the first part to determine the suitable model. Secondly, the other part of the sample is used to measure the forecasting ability.

For model building, recursive windows and rolling windows approaches emerge as two common ways to estimate the parameters.

Under a recursive windows approach, the operation always starts with the initial date, and one additional observation will be added for the next operation until reaching the last observation of the data set. Conversely, the rolling windows approach keeps the length of the in-sample period fixed, the start and end dates successively increase by one observation for each operation.

Moreover, the two abovementioned approaches will generate different results under varied conditions. For example, when considering the parameter estimation error, in other words, the forecasting accuracy, the quantity of data employed in estimating parameters has to be seriously evaluated. In this regard, if the series are heterogeneous and with structural variety, it is better to use all available in-sample data to estimate the parameters which means that the recursive windows approach should be undertake. However, if the earliest available data was somehow generated by DGP method thus unrelated to the present, using all available in-sample data may cause bias in parameter estimation and the subsequent forecasting.

On the other hand, using the rolling windows approach can reduce the mean squared forecast error from a weakened level of heterogeneity obtained by employing a relatively smaller sample. However, this gain in reduced heterogeneity from the small sample will lead to an increase in the variance of the estimated parameters. This resulting high variance will then cause the mean squared forecast error to increase. Consequently, there is indeed a dilemma when finding the right size of the data while minimizing the level of mean squared error.

To harvest the benefits of the two abovementioned approaches, both recursive and rolling windows will be used in this paper.

In addition, in order to evaluate the forecasting accuracy, the mean squared error (MSE) and mean absolute error (MAE) are computed:

$$MSE = \frac{1}{T^{**}} \sum_{t=1}^{T^{**}} (\hat{x}_t - x_t)^2 \quad ,$$

$$MAE = \frac{1}{T^{**}} \sum_{t=1}^{T^{**}} \left| \hat{x}_t - x_t \right| \quad ,$$

Where T^{**} is the number of the out-of-sample forecasts.

The smaller the MSE or MAE, the better the model is; this in turn indicates the according criterion, which is used to determine the lag length of the model, produces the superior forecasting accuracy.

7 Empirical Application

7.1 Source of Data and Processing

To accurately determine the RV, the source of the data and its according characteristics are very important. The accuracy of the estimation of RV can possibly be negatively affected by the measurement errors and the microstructure noise.

In financial markets, information is an important factor influencing the movements of stock market prices. Generally speaking, the higher the collection frequency, the more information lost. Data with high frequency will cause varying degrees of deficiencies, which will in turn induce a biased and inconsistent measure.

On the other hand, the microstructure noise is sometimes caused by bid-ask bounce, asynchronous trading, market closing effects and so on. As the sampling frequency increases, the microstructure noise becomes less significant. In other words, variance reduction can be obtained by utilizing data with high frequency.

Therefore, there is a trade off between the bias and variance when choosing the data frequency. Hence it must be carefully chosen in order to find the appropriate balance. Various frequencies have been suggested and the typical frequency is with the intra-daily interval of five minute proposed by Andersen & Bollerslev (1997b) and Koopman, Jungbacker and Hol (2005).

In this paper, six series of RV data[2] derived from Standard & Poor's Depositary Receipts (SPDRs) high-frequency data are used. SPDRs are the largest Exchange Traded Funds (ETF) in the U.S. and set up to mimic the movements of S&P 500. The value of each unit reflects the movements in the index. The price of a unit in the trust is always the current value of the S&P 500 divided by 10. The six series are calculated by different methods described in Bandi and Russell (2005, 2006). The sample data is for the time

---

[2] I thank Professor Benoit Perron for providing the data.

span between January 2, 1998 and March 31, 2006, a total of 2053 observations for each of the six series.

7.2 The Descriptive Statistic of the Sample Data

Table 1 and 2 summarise the descriptions of the sample. From Table 1 and 2, it is observed that the mean of RV of the sample is positive; whereas the mean of the log RV of the sample is negative. The standard deviation of log RV is relatively small.

From Table 1 and 2 as well as Figure 1 and 2, there are strong evidences indicating that the RV series are not normally distributed. All the values of skewness[3] (4.66~6.00) exceed zero by a large amount which indicates the RV series are heavily skewed to the right; all the values of kurtosis (36.10~64.45) exceed 3 greatly which indicates that the RV series are with thick tails. In contrast, the log RV series have much smaller kurtosis (3.20~4.97) and skewness (0.03~0.29). For this reason, the log RV series are considered to asymptotically have a normal distribution.

The six series have very similar characteristics; all of their means and standard deviations do not differ greatly. From Table 2, it is observed that their main differences are the range value, skewness and kurtosis. Hence the six series can be separated into two groups based on these main differences. One group consists of series 1, 2 and 3 with smaller range value and kurtosis, and the other consists of series 4, 5 and 6 with smaller skewness.

[3] The “skewness” and “kurtosis” measures are used to describe a distribution. A normal distribution has skewness of zero and kurtosis of three. Skewness is a measure of the direction and degree of the asymmetry. If skewness is zero, the data are symmetric; if skewness is positive, the data are skewed to the right which means that the right tail is longer relative to the left tail; if skewness is negative; the data are skewed to the left, which means that the left tail is longer relative to the right tail. The greater the absolute value the greater the degree of skewness. Kurtosis is a measure of the thickness of the tails. If kurtosis is greater than 3, the data are thick in the tails; if kurtosis is less than 3, the data are thin in the tails.

7.3 The Existence of Long Memory Test

To detect and estimate the long memory process, one of the most commonly used methods to estimate the fractional differencing number *d* is the GPH method.

According to Hosking (1981), when $0 < d < 0.5$, the series is a stationary long memory process; when $-0.5 < d < 0$, the series is an anti-persistent memory process; when $0.5 < d < 1$, the series is non-stationary process as it has an infinite variance but reverting.

The GPH method is used in this study to test the long memory of six series of log RV. Because there is no standard way of choosing the frequency, the following formula is used $g(T) = T^{\varphi}$, $T = 500, 501, \ldots, 2053$; different φ values such as $\varphi = 0.5$, 0.6, 0.7 and 0.8 are selected to estimate the *d* value.

Table 3 presents the value of estimated $\hat{d}$ by using the recursive windows approach. The number of frequency is observed to have a great influence on the value of $\hat{d}$. Also, from $T = 500$, all of the estimators are within the following interval $\hat{d} \in (0,1)$. Furthermore, since $(\hat{d} - d)/s(\hat{d}) \to N(0,1)$, the Student's t-test statistics for the null hypothesis of $d = 0$ is used to test the significance of $\hat{d}$. From Table 3 with results from using the recursive windows approach, all the t statistics are greater than all the critical values at all the significance level. Hence, the null hypothesis that all the six series are not long memory process is rejected.

From Figure 3 and 4, using the recursive windows approach, with the increase of φ, the standard deviation of $\hat{d}$ experiences a small change and the proportion of $\hat{d} \in (0, 0.5)$ increases; when $\varphi = 0.8$, the proportion of $\hat{d} \in (0, 0.5)$ is at its largest.

Base on Mood, Perron and Wang (2007), to ensure a better use of the AR approximation process, d should be within the following range: $0 < d < 0.5$. Hence, for the two modified FPE criteria, the ceiling of $\hat{d}$ is restricted to be 0.49; that is, if the value of $\hat{d}$ are higher than 0.49, 0.49 is used instead of the estimated $\hat{d}$.

7.4 Empirical Prediction Approaches

Six series of data are used in out-of-sample h-step-ahead forecasting employing both the recursive windows approach and rolling windows approach. There are a total of 2053 observations for each series. Accordingly, the data are split into two parts for each series: the first part is the in-sample part which has T observations, and the other part is the out-of-sample part which has 2053-T observations.

Each of the six series is separated into 3 groups according to the different T values. Group one is with $T = 500$, group two is with $T = 1000$, and group three is with $T = 1500$. For each group, daily data are used as the base of the forecast horizons: $h = 1$ (a day), $h = 5$ (a week), $h = 22$ (a month), $h = 132$ (6 months), $h = 252$ (a year). Since the maximum number of lags depends on the length of the series, and in the following study, the maximum number is chosen to be equal to 24.

Recursive windows approach

To estimate the lag length, forecasting with recursive windows approach is introduced as the following: always start with the first observation and use an increasing data window. Thus, the total observations (T^*) increase for each time estimation period.

For example, in group one ($T = 500$), for $h = 1$, base on the first 500 observations of the samples, 1553 models are estimated. At the first time estimation, by using observations from 1 to 500 ($T^* = 500$), the parameters are estimated by using the number of lags p from 0 to 24. Modified FPE1, modified FPE2, FPE, AIC and BIC are subsequently computed by using the estimator $\hat{S}^2_{h,p}$. The lag length is selected if it is able to minimize the criterion. Then, the value of the 501st observation is forecasted.

At the second time estimation, employing observations from 1 to 501, the abovementioned method is used for ($T^* = 501$) to re-estimate the parameters and to construct a new model for forecasting the value of the 502nd observation.

Repeat the process, the model parameters are re-estimated each time; and an additional observation is included into the estimation period for 1553 times. At the end, observations from 1 to 2052 ($T^* = 2052$) are used to forecast the value of the 2053rd value.

For h with different numbers, the theory of forecasting is the same; only the total number of forecasts done is different. For $h = 1$, there are 1553 times; for $h = 5$ there are 1549 times; for $h = 22$ there are 1532 times; for $h = 132$ there are 1422 times and for $h = 252$ there are 1302 times. To compare the forecasting accuracy, the MSE and the MAE are used.

For the other two groups ($T = 1000$ and 1500), the same steps, as in group one, are taken. The differences are the forecast starting points and the total forecast numbers.

Rolling windows approach

When forecasting with the rolling windows approach, similar steps as of the recursive windows are taken, except the starting date and the length of the in-sample part. For a rolling window, with a fixed moving data window, the total observations (T^*) is always equal to T for each group forecasting.

For estimating the fractional differencing number, d, different numbers of frequency of regression are used: $\varphi = 0.5, 0.6, 0.7, 0.8$. Thus, $g(T) = T^{0.5}, T^{0.6}, T^{0.7}, T^{0.8}$.

7.5 Empirical Illustration

7.5.1 Optimal Lag(s) Estimation

With the help of the criteria, the optimal lags of each period can be estimated. From Table 4 and figure 4, it is observed that the variance number of optimal lags in the whole data is larger under the recursive windows approach with $h = 1$, $T = 500$ and $\varphi = 0.5$ respectively. For example, for the modified FPE1 criterion, the best order of the model in some data achieves to 21, however in some it is only 4. The results demonstrate that under varied optimal lags, it is reasonable to use an out-of-sample approach to forecast the model.

7.5.2 Forecasting Results

The MSE and MAE are considered to measure the forecasting accuracy of the model selection criteria. In addition, the forecasting accuracy is used to evaluate the forecasting performance.

To better interpret the forecasting results, the MSEs and MAEs of all five criteria are divided by the MSEs and MAEs of the modified FPE1 criterion to obtain different ratios (referred hereafter as the "ratio"). The criterion with the smallest ratios has the best forecasting accuracy, thus also best forecasting performance.

From Figure 2 and Table 2, it is observed that series 1, 2 and 3 are somewhat comparable whereas series 4, 5, and 6 have very similar characteristics too. Hence, for the ease of presentation, only the detailed forecasting results achieved by using both the recursive and rolling windows approaches for series 1 and series 5 are analyzed, and only the results from the recursive windows approach for series 1 and 5 are reported in detail.

Tables 5 to 26 report the ratios of the MSEs and MAEs of all five criteria relative to the ones of the modified FPE1 criterion and the mean lag order selected from the recursive windows approach.

From the forecasting results, it is observed that the FPE and AIC criteria give the same results for all the *h*-step-ahead forecasts when the size of the in-sample part $T \geq 500$, thus they will be denoted as F&A in the following analyses. This result confirms the previous theoretical analysis discussed in section 4.3.

According to the forecasting results (Table 5 to 26), the effect of the different forecast horizons *h*, $\varphi\, s$ [4], the size of the in-sample part and the information criteria are explained in the following analysis.

Recursive windows approach

When $h = 1$, the modified FPE1 has the smallest ratio for all φs tested when $T = 1500$ in series 1, as well as when $T = 1000$ with $\varphi = 0.5, 0.6$ and 0.7 and for all φs tested when $T = 1500$ in series 5. On the other hand, the F&A has the smallest ratio for the rest groups and φs.

When $h = 5$, the modified FPE1 has the smallest ratio when $T = 1000$ with $\varphi = 0.7$ in series 1 and $\varphi = 0.5, 0.6$ and 0.7 in series 5, as well as when $T = 1500$ for all φs tested in both of series 1 and 5. In contrast, the F&A has the smallest ratio for the rest groups and φs.

When $h = 22$, the modified FPE1 has the smallest ratio when $T = 500$ for $\varphi = 0.7$ and 0.8 in series 1; whereas, the modified FPE2 has the smallest ratio when $T = 500$ and 1000 for $\varphi = 0.6$ in series 1. On the other hand, the F&A has the smallest ratio for the rest groups and φs.

When $h = 132$, the modified FPE1 has the smallest ratio for all φs tested when $T = 1500$ in series 1, as well as when $T = 500$ with $\varphi = 0.8$ and when $T = 1000$ for all

---

[4] φ affects the number of frequency of regression and the estimated fractional differencing number ($\hat{d}$)

φs tested in series 5; whereas, the modified FPE2 has the smallest ratio for all φs tested when $T = 500$ and $T = 1000$ in series 1, as well as when $T = 500$ with $\varphi = 0.5, 0.6$ and 0.7 in series 5. In contrast, the F&A has the smallest ratio for all φs tested when $T = 1500$ in series 5.

When $h = 252$, the FPE1 has the smallest ratio for all the groups and φs tested in series 5. On the other hand, the modified FPE2 has the smallest ratio for all the groups and φs tested in series 1.

Rolling windows approach

When $h = 1$, the modified FPE1 has the smallest ratio for all φs tested when $T = 1000$ in series 1, and for series 5 when $T = 500$ with $\varphi = 0.5, 0.6, 0.7$. The modified FPE2 has the smallest ratio when $T = 500$ with $\varphi = 0.6, 0.7, 0.8$ in series 1 and when $T = 1500$ with $\varphi = 0.8$ in series 5. On the other hand, the F&A has the smallest ratio for all φs tested when $T = 1500$ in series 1 and when $T = 500$ with $\varphi = 0.8$ in series 5. The BIC has the smallest ratio for all φs tested when $T = 500$ and 1000, as well as when $T = 1500$ with $\varphi = 0.5, 0.6, 0.7$ in series 5 and when $T = 500$ with $\varphi = 0.5$ in series 1.

When $h = 5$, the modified FPE1 has the smallest ratio when $T = 1000$ with $\varphi = 0.7$ and $T = 1500$ with $\varphi = 0.8$ in series 1. The modified FPE2 has the smallest ratio when $T = 500$ with $\varphi = 0.6, 0.7, 0.8$, $T = 1000$ with $\varphi = 0.5, 0.6, 0.8$ and almost all φs except $\varphi = 0.8$ when $T = 1500$ in series 1 and when $T = 1500$ with $\varphi = 0.5, 0.6, 0.7$ in series 5. In contrary, the F&A has the smallest ratio when $T = 1000$ with $\varphi = 0.7$, 0.8 and when $T = 1500$ with $\varphi = 0.8$ in series 5. The BIC has the smallest ratio when $T = 500$ for $\varphi = 0.5$ in series 1 and when $T = 500$ with all φs tested and when $T = 1000$ with $\varphi = 0.5, 0.6$ in series 5.

When $h = 22$, the modified FPE2 has the smallest ratio when $T = 500$ with $\varphi = 0.7, 0.8$ in series 5; whereas, the F&A has the smallest ratio when $T = 1000$ and 1500 with all φs tested in series 1 and 5. The BIC has the smallest ratio when $T = 500$ for all φs tested in series 1 and when $T = 500$ for $\varphi = 0.5, 0.6$ in series 5.

When $h = 132$, the modified FPE1 has the smallest ratio when $T = 1500$ for $\varphi = 0.5, 0.6, 0.7$ in series 1 and when $T = 500$ for $\varphi = 0.5, 0.6$ in series 5. The modified FPE2 has the smallest ratio when $T = 1000$ for all φs tested in series 1 and when $T = 500$ for $\varphi = 0.7, 0.8$ in series 5. On the other hand, the F&A has the smallest ratio when $T = 1500$ for $\varphi = 0.8$ in series 1 and when $T = 1000$ and 1500 for all φs tested in series 5. The BIC has the smallest ratio when $T = 500$ for all φs tested in series 1.

When $h = 252$, the modified FPE1 has the smallest ratio, when $T = 500$ with $\varphi = 0.5$, $T = 1500$ with $\varphi = 0.5, 0.6, 0.7$ in series 5. The modified FPE2 has the smallest ratio, when $T = 1000$ and 1500 for all φs tested in series 1 and when $T = 500$ with $\varphi = 0.7, 0.8$ and $T = 1000$ with $\varphi = 0.8$ in series 5. In contrast, the F&A has the smallest ratio when $T = 1500$ for $\varphi = 0.8$ in series 5. The BIC has the smallest ratio when $T = 500$ for all φs tested in series 1 and when $T = 500$ with $\varphi = 0.6$ and $T = 1000$ with $\varphi = 0.5, 0.6, 0.7$ in series 5.

Analysis of the summarized forecasting results

Based on the forecasting results, the conclusions are obtained as the following.

By using the recursive windows approach, in most cases, the modified FPE1 or the modified FPE2 produces the best forecasting accuracy comparing to other criteria when the forecast horizons $h \geq 132$ days. In addition, when $h \leq 5$ days and the size of in-sample part $T \geq 1000$, the modified FPE1 achieves the superior forecasting performance.

On the other hand, the FPE and AIC in general perform considerable better than the others when $5 < h \leq 132$ days.

By using the rolling windows approach, the modified FPE1 or FPE2 has the best forecasting performance in most cases when the forecast horizons $h \geq 252$ days and the size of in-sample part $T > 500$. When $h < 252$ days, the modified FPE1 or FPE2 has the best performance in several instances which depends on the in-sample size and the estimated fractional differencing number $\hat{d}$. In general, the forecasting gains are small in the short-range forecast.

In fact, it is difficult to identify which modified FPE criterion is better. Based on the formulas of the two modified criteria, the fractional differencing number, *d*, utilize in both modified FPE1 and FPE2, has an effect on the lag choice which in turn influences the structure of the forecasting model. Thus, the resulting forecasting accuracy certainly will be different as it depends on factors such as the characteristics of the series, the size and the variances of the in-sample part. Therefore, for a long-range[5] forecast, in most cases, for all the estimated fractional differencing number $\hat{d}$, the modified FPE1 or FPE2 has the best forecasting performance.

In most cases, when the modified FPE1 criterion has the best forecasting performance, it selects the smallest mean lag length comparing to AIC, BIC and FPE.

In addition, the mean lag length obtained from the modified FPE2 is the smallest when $h \geq 132$ days by using the recursive windows approach, and when $h \geq 252$ days by using the rolling windows approach. Particularly, when $h \geq 252$ days, the mean lag length obtained from the modified FPE2 is close to zero when using the recursive windows approach with the in-sample size $T \geq 500$, and when using the rolling windows approach with $T \geq 1500$. The most parsimonious model is therefore achieved in this case. In this regard, for long-range forecasts, the model can be estimated by using only the

[5] With the forecast horizons $h \geq 132$ days by using recursive windows, or $h \geq 252$ days by using rolling windows

simple regression model instead of the AR model when employing the modified FPE2 criterion.

Consequently, the modified FPE1 or FPE2 has a superior forecasting performance for the forecasting of RV in the long term, since it outperforms the other criteria in terms of the forecasting accuracy and the smallest mean lag length.

In summary, independent of the selected approach (recursive windows or the rolling windows approach), one of the modified FPE criteria will have the best performance for forecasting the RV in the long-range forecasts.

8 Conclusion

This paper has analyzed two new modified FPE criteria to determine the lag length of the AR approximation model for forecasting the RV. I use the the SPDRs high-frequency RV data, analyze the model characteristics of RV which resembles a long memory processes, and apply the AR approximation model using *h*-step-ahead out-of-sample forecasting. Both the recursive and rolling windows approaches are employed.

To evaluate the two new criteria, the forecasting performance of the modified FPE1 and FPE2 are compared to the classical information criteria such as the AIC, BIC and FPE criteria. The forecasting accuracy is examined to evaluate the forecasting performance of the RV by using different information criteria. To evaluate the accuracy, the MSE and MAE values are calculated. For ease of comparison, the ratios of MSEs and MAEs relative to the modified FPE1 are used in this paper. Hence, the criterion with the smallest ratio has the superior forecasting accuracy. Specifically, the information criterion producing a high level of forecasting performance is superior to the other criteria for forecasting the RV.

The comparison of the forecasting results indicates that: for long-range forecast using both of the recursive and rolling windows approaches, both the theoretical results and the empirical practices show that the modified FPE1 or the modified FPE2 will result in superior performance, in terms of not only the forecasting accuracy but also the simplicity, comparing to the classical information criteria. On the other hand, for a short-range forecast, the forecasting gain for both of the modified FPE1 and FPE2 are small, their usefulness is more doubtful.

As a result, for long-range forecast of the RV, the modified FPE1 or FPE2 criterion is recommended.

References

[1] Akaike, H., “Power spectrum estimation through autoregressive model fitting”, Annals of Institute of Statistics Mathematics 21, 1969, 407-419.

[2] Akaike, H., “Information theory and an extension of the maximum likelihood principle”, In B.N. Petrov and F. Csake (eds.), Second International Symposium on Information Theory. Budapest: Akademiai Kiado, 1973, 267-281.

[3] Andersen, T.G. and T. Bollerslev, “Intraday periodicity and volatility persistence in financial markets”, Journal of Empirical Finance, 4, 1997b, 115-158.

[4] Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, “The Distribution of Realized Stock Return Volatility”, Journal of Financial Economics, 61, 2001, 43-76.

[5] Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, “The Distribution of Realized Exchange Rate Volatility”, Journal of American Statistical Association, 2001a, 96, 42-55.

[6] Andersen, T., T. Bollerslev, F. X. Diebold, and P. Labys, “Modeling and Forecasting Realized Volatility”, Econometrica, 71, 2003, 579-626.

[7] Areal, N. M., and S. J. Taylor, “The Realized Volatility of FTSE-100 Futures Prices”, Journal of Futures Markets, 22 (7), 2002, 627–648.

[8] Bandi, F.M. and J. R. Russell, “ Microstructure Noise, Realized Variance, and Optimal Sampling”, Graduate School of Business, University of Chicago, 2005, working paper.

[9] Bandi, F.M. and J. R. Russell, “Separating Microstructure Noise from Volatility”, Journal of Financial Economics 79, 2006, 655-692.

[10] Barndorff-Nielsen, O.E. and N. Shephard, “Non-Gaussian OU based models and some of their uses in financial economics (with discussion)”, Journal of the Royal Statistical Society, Series B 63, 2001, 167–241.

[11] Barndorff-Nielsen, O.E. and N. Shephard, “Econometric analysis of realised volatility and its use in estimating Stochastic Volatility models”, Journal of the Royal Statistical Society, Series B 64, 2002, 253–280.

[12] Beran, J., Bhansali, R.J. and Ocker, D., “On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes”, Biometrika, Vol. 85, 1998, 921-934.

[13] Berk, K.N., “Consistent autoregressive spectral estimates”, The Annals of Statistics 2, 1974, 489-502.

[14] Cheung, Y., “Long Memory in Foreign-Exchange Rates”, Journal of Business and Economic Statistics, Vol. 11, 1990, 93-101.

[15] Crato, N. & B.K. Ray, “Model selection and forecasting for long range dependent processes”, Journal of Forecasting 15, 1996, 107-125.

[16] Dahlhaus, R., “Efficient Parameter Estimation for Self-similar Processes”, The Annals of Statistics, 17, 1989, 1749-1766.

[17] Fildes, R. and Makridakis, S., “Forecasting and loss functions”, International Journal of Forecasting, 4, 1988, 545-50.

[18] Franses, P. H. & M. Ooms, “A periodic long-memory model for quarterly UK inflation”, International Journal of Forecasting 13, 1997, 117-126.

[19] Fox, R. and Taqqu, M. S., “Large Samples Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series”, Annals of Statistics, 14, 1986, 517-532.

[20] Geweke, J. F., and S. Porter-Hudak, “The Estimation and Application of Long Memory Time Series Models”, Journal of Time Series Analysis, 4(4), 1983, 221–238.

[21] Hassler, U., “Regression of spectral estimators with fractionally integrated time Series”, Journal of Time Series Analysis, 14, 1993a, 369-380.

[22] Hosking, J.R.M., “Fractional differencing”, Biometrika 68, 1981, 165-176.

[23] Hosking, J.R.M., “Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series”, Journal of Econometrics 73, 1996, 261-284.

[24] Jones, R. H. “Identification and Autoregressive Spectrum Estimation”, IEEE Transactions on Automatic Control, vol. 19, 1974, 894-897.

[25] Koopman, S.J., B. Jungbacker and E. Hol , “Forecasting Daily Variability of the S&P100 Stock Index using Historical, Realized and Implied Volatility Measures”, Journal of Empirical Finance, 12, 2005, 445-475.

[26] Merton, R. C., “On Estimating the Expected Return on the Market; An Exploratory Investigation”, Journal of Financial Economics, 8, 1980, 323–361.

[27] Moon, H. R., Benoit Perron and S.-H. Wang, "Forecast of Long Memory Time Series Using AR Approximation", Mimeo, 2007.

[28] Oomen, R. C., "Statistical Models for High Frequency Security Prices", Manuscript Warwick Business School, 2003.

[29] Oomen, R. C., "Modelling Realized Volatility When Returns are Serially Correlated", Social Science Research Center, Berlin, Tech. Rep. SP II, 2004.

[30] Quenouille, M.H., "Approximate tests of correlation in time series", Journal of the Royal Statistical Society, Series B. 11, 1949, 18-44.

[31] Ray, B. K., "Modeling long-memory processes for optimal long-range prediction", Journal of Time Series Analysis, 14, 1993a, 511-525.

[32] Ray, B. K., "Long range forecasting of IBM product revenues using a seasonal fractionally differenced ARMA model", International Journal of Forecasting 9, 1993b, 255-269.

[33] Robinson, P.M., "Log-periodogram Regression of Time Series with Long Range Dependenc", Annals of Statistics, 23, 1995a, 1048-1072.

[34] Schmidt, C. M. and Tschernig, R., "Identification of fractional ARIMA models in the presence of long memory", Discussion Paper, University of Munich, 1993.

[35] Schwarz, G., "Estimating the dimension of a model", Annals of Statistics, 6, 1978, 461-464.

[36] Sowell, F., "Modeling long run behavior with the fractional ARIMA model", Journal of Monetary Economics, Vol. 29, 1992b, 277-302.

[37] Taylor, S.J. and X. Xu, "The incremental volatility information in one million foreign exchange quotations", Journal of Empirical Finance 4, 1997, 317-340.

[38] Wooldridge, Jeffrey M. "Introductory Econometrics: A Modern Approach", Cincinnati, OH: South-Western, College Publishing, 2006.

Figures

Figure 1 - Original RV

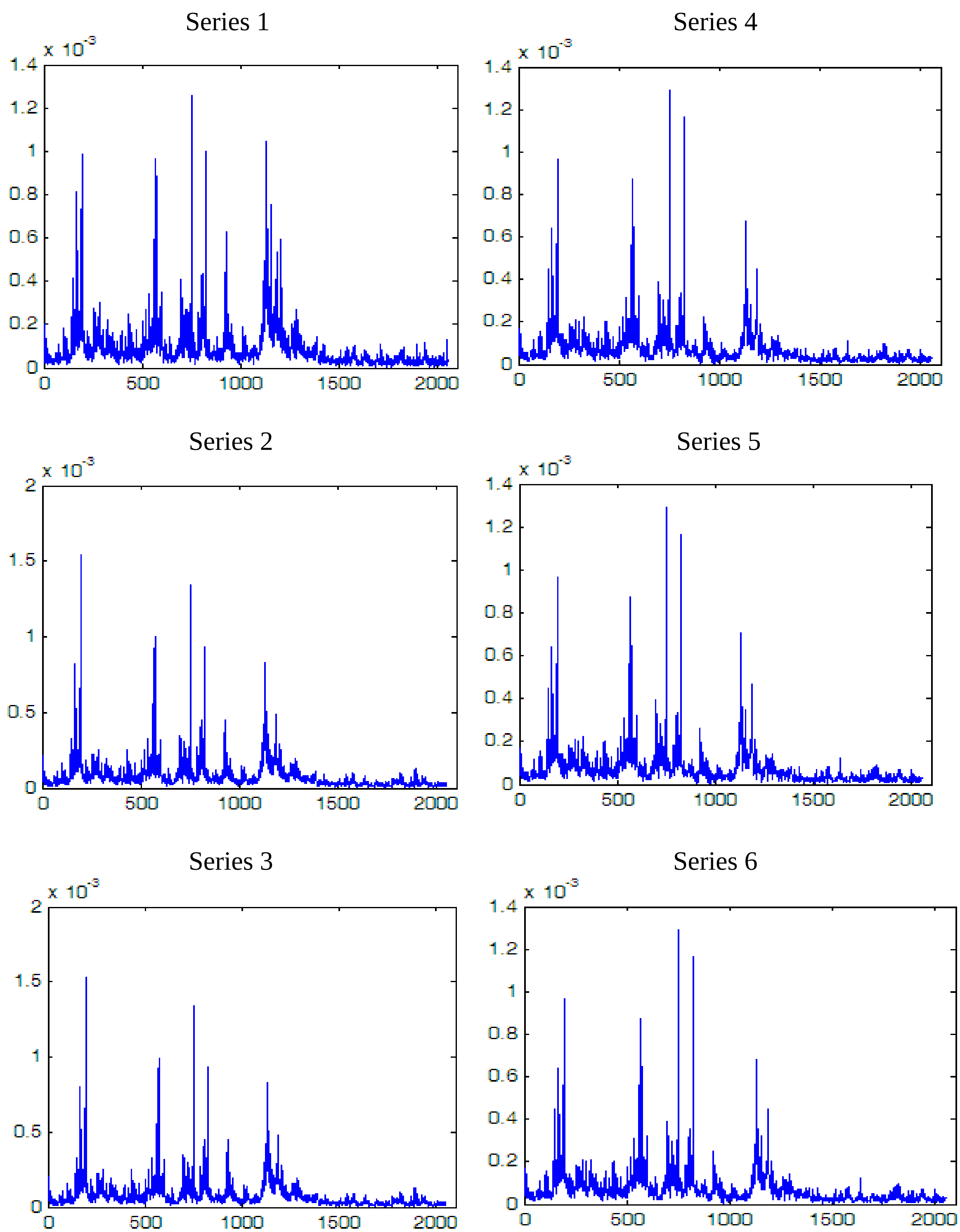
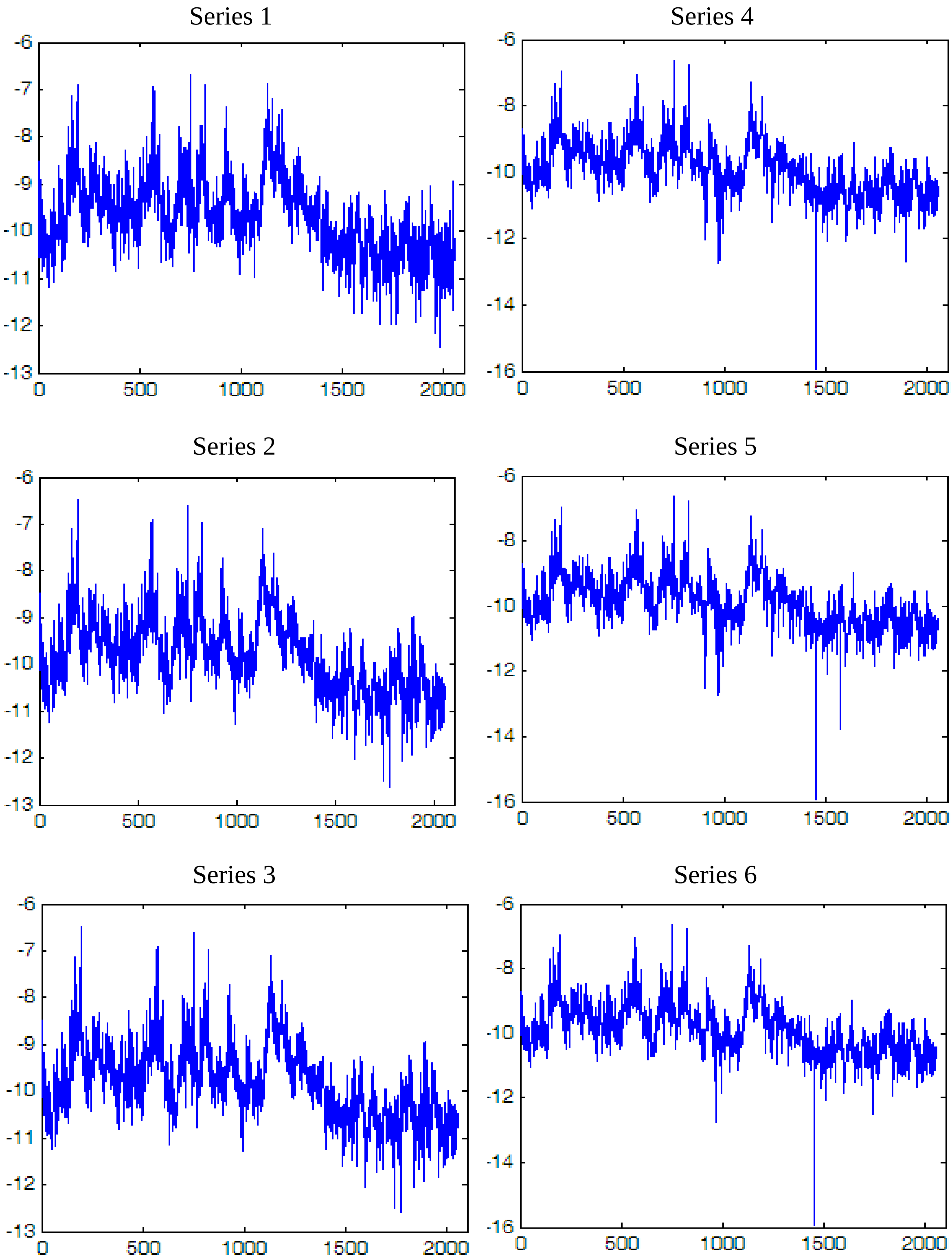


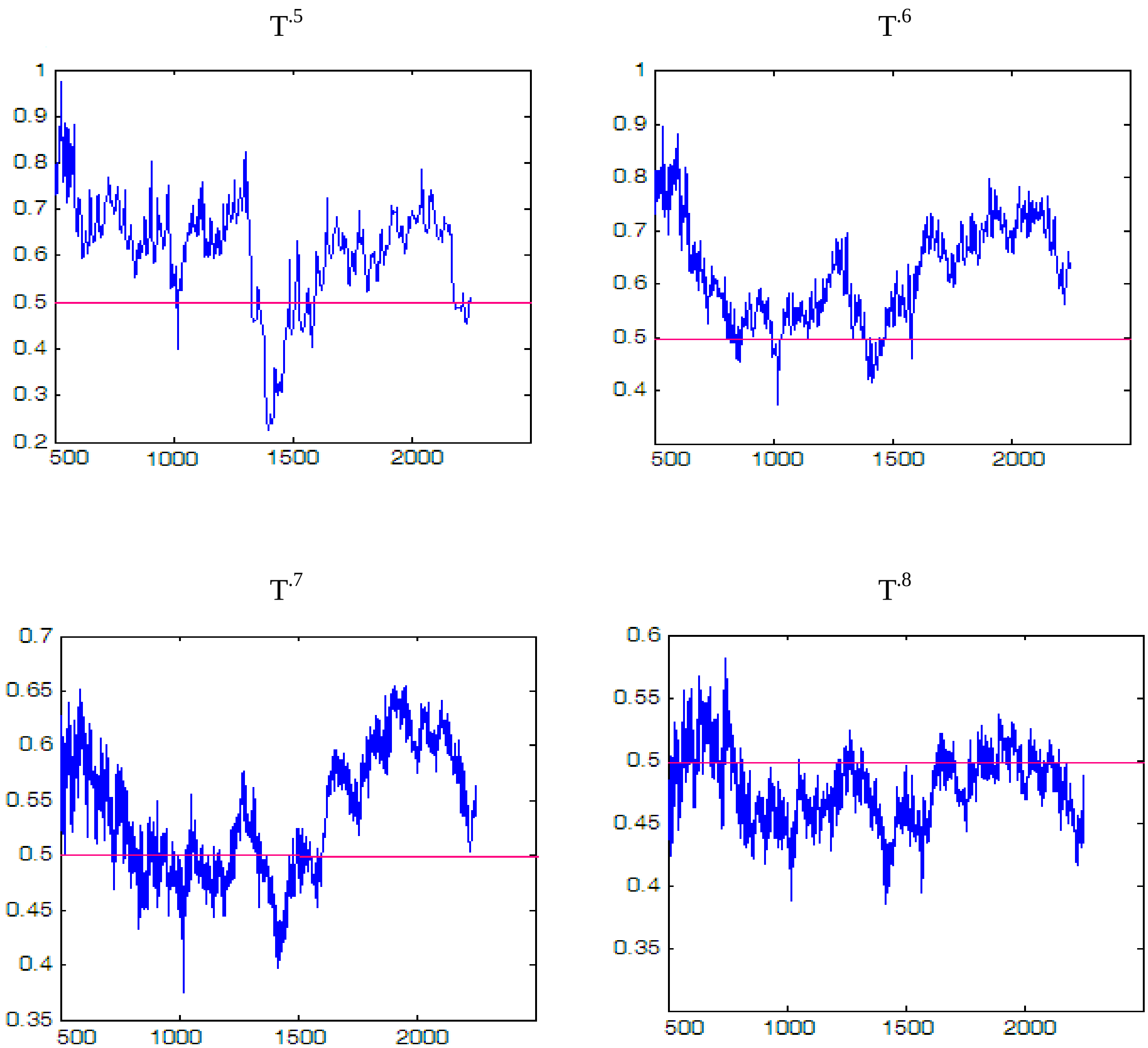
Figure 2 - Log RV



(X axis: the number of observations; Y axis: the value of Log RV)

Figure 3 - Estimators of “*d* ”

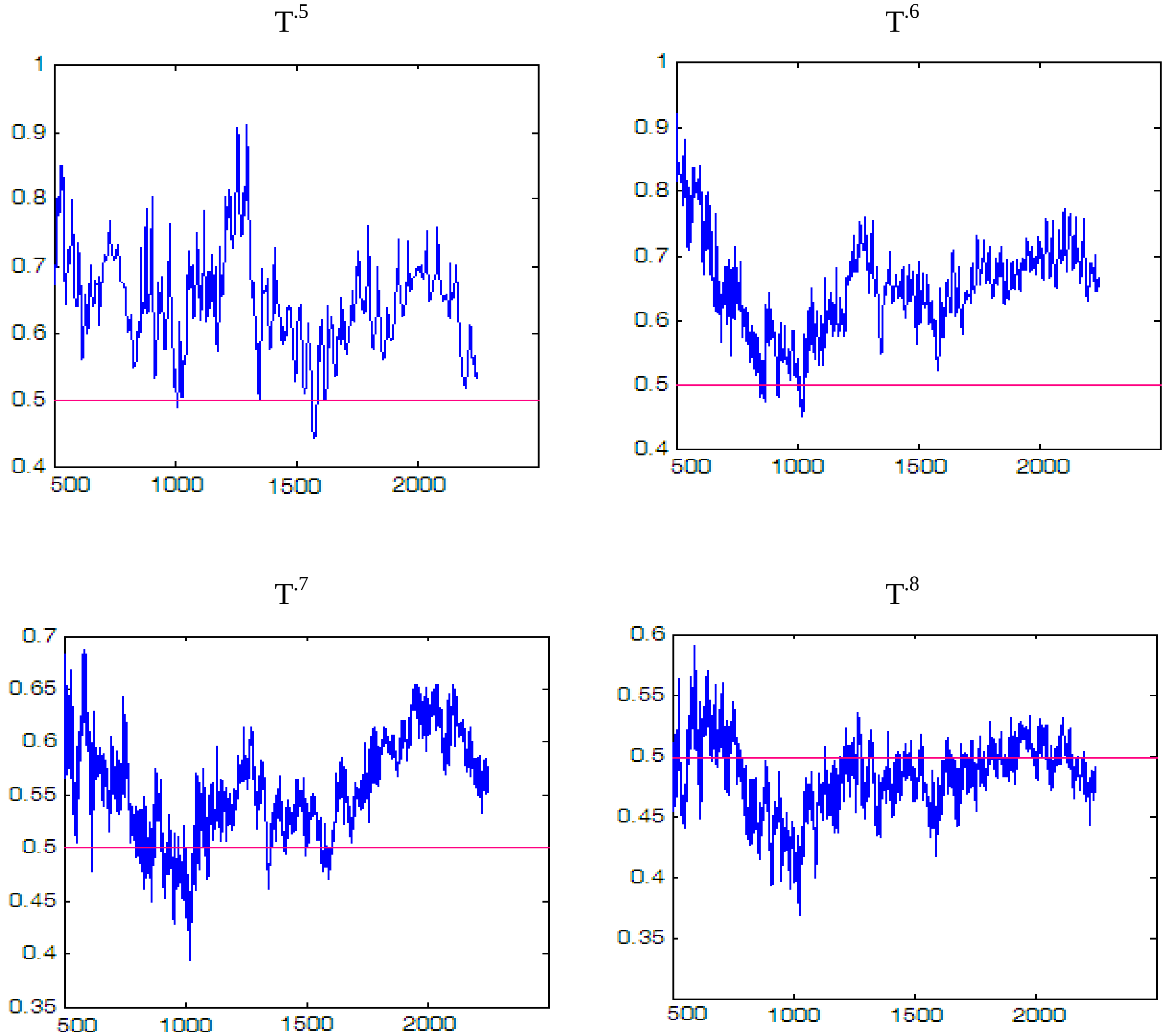
(Series 1, Recursive windows approach)



(X axis: the number of observations; Y axis: the value of estimated “*d* ”)

Figure 4 - Estimators of “*d* ”

(Series 5, Recursive windows approach)

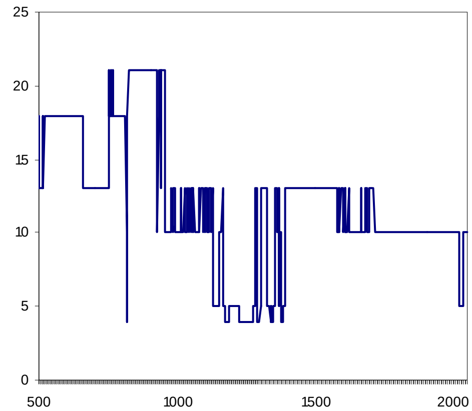


(X axis: the number of observations; Y axis: the value of estimated “*d* ”)

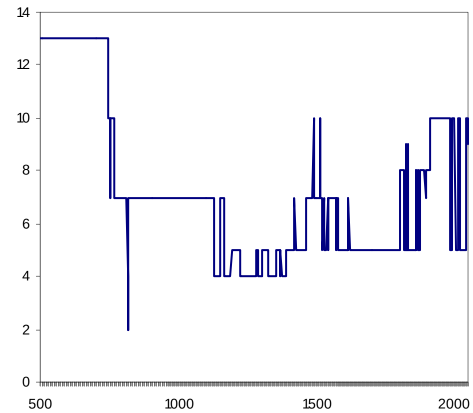
Figure 5 - Optimal Lag(s)

(Series 1, h=1, T=500, $g(T)=T^{0.5}$, Recursive windows approach)

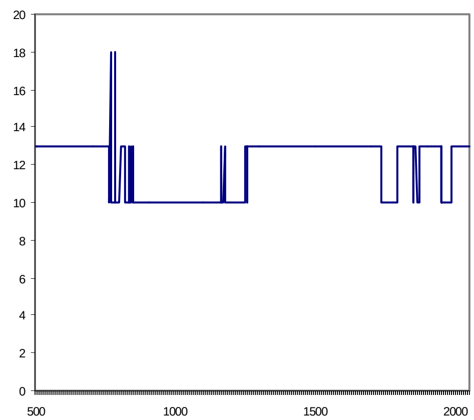
FPE_M1



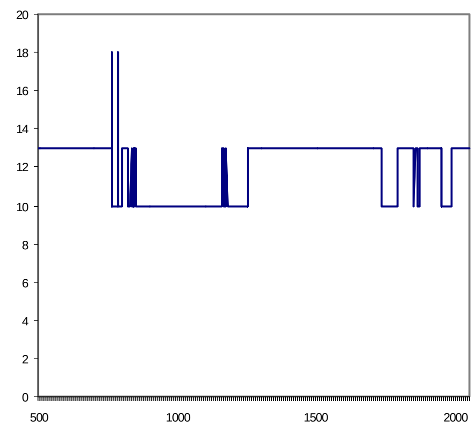
FPE_M2



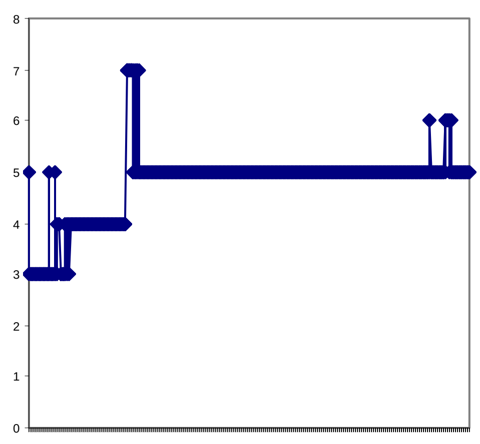
FPE



AIC



BIC



(X axis: the number of observations; Y axis: the lag length)

Tables

Table 1 - Original RV

	y1	y2	y3	y4	y5	y6
min	3.848e-006	3.274e-006	3.41e-006	1.236e-007	1.236e-007	1.236e-007
max	0.001258	0.001539	0.001529	0.001292	0.001292	0.001292
mean	8.938e-005	7.919e-005	7.829e-005	7.048e-005	7.121e-005	7.098e-005
median	6.001e-005	5.287e-005	5.217e-005	4.848e-005	4.868e-005	4.85e-005
mode	3.848e-006	2.653e-005	2.398e-005	7.018e-005	7.083e-005	7.072e-005
std	0.000102	9.524e-005	9.456e-005	8.119e-005	8.135e-005	8.108e-005
range	0.001254	0.001535	0.001526	0.001291	0.001291	0.001291
skewness	4.6631	6.0010	6.0491	5.9806	5.9822	6.0119
kurtosis	36.0963	63.6935	64.4499	63.7455	63.4539	64.0936

Table 2 - Log RV

	x1	x2	x3	x4	x5	x6
min	-12.47	-12.63	-12.59	-15.91	-15.91	-15.91
max	-6.678	-6.477	-6.483	-6.652	-6.652	-6.652
mean	-9.687	-9.815	-9.826	-9.906	-9.888	-9.887
median	-9.721	-9.848	-9.861	-9.934	-9.93	-9.934
mode	-12.47	-10.54	-10.64	-9.565	-9.555	-9.557
std	0.8233	0.8248	0.8241	0.8058	0.7955	0.7846
range	5.79	6.153	6.106	9.254	9.254	9.254
skewness	0.2388	0.2891	0.2940	0.0631	0.0339	0.1238
kurtosis	3.2030	3.1654	3.1795	4.6279	4.9724	4.8250

Table 3 - Estimator of "*d*" Value

(Series 1 and 5, Recursive windows approach)

Series 1

$\hat{d}$ \ g(T)	$T^{0.5}$	$T^{0.6}$	$T^{0.7}$	$T^{0.8}$
min	0.2244	0.3724	0.3746	0.3854
max	0.9727	0.8986	0.6678	0.5818
mean	0.6205	0.6184	0.5371	0.4772
std	0.1126	0.0924	0.0569	0.0292
t tste	5.5106	6.6926	9.4394	16.342

Series 5

min	0.4413	0.4503	0.3941	0.3688
max	0.9116	0.7738	0.6659	0.5599
mean	0.6387	0.6344	0.5510	0.4790
std	0.07352	0.06235	0.04749	0.02829
t test	8.6874	10.1748	11.6024	16.9318

Table 4 - The Optimal Lag Length of the Model

(Series 1, h=1, T=500, Recursive windows approach)

	Modified FPE1	Modified FPE2	FPE	AIC	BIC
Min	4	2	10	10	3
Max	21	13	18	18	7
Mean	12.1	7.295	11.91	11.91	4.766
std	4.578	2.946	1.46	1.46	0.6952

Table 5 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 1: T=500, $g(T)=T^{0.5}$)

h / Criteria	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(12.1017)		(10.0844)		(5.9742)		(5.1565)		(0.5892)	
$FPE^M 2$	0.9977	0.9986	1.0085	1.0046	1.0003	1.0007	0.9734	0.9890	0.9739	0.9881
	(7.2949)		(7.2872)		(4.1983)		(0.7064)		(0.0135)	
FPE	0.9952	0.9978	0.9978	0.9999	0.9946	0.9975	1.0089	0.9994	1.0266	1.0122
	(11.9131)		(10.2537)		(7.0026)		(7.5171)		(2.1249)	
AIC	0.9952	0.9978	0.9978	0.9999	0.9946	0.9975	1.0089	0.9994	1.0266	1.0122
	(11.9131)		(10.2537)		(7.0026)		(7.5171)		(2.1249)	
BIC	0.9997	1.0003	1.0114	1.0067	1.0005	1.0001	0.9772	0.9888	0.9909	0.9952
	(4.7663)		(6.5010)		(4.1539)		(0.9008)		(0.2428)	

(Parentheses represent the mean lag order)

Table 6 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 1: T=500, $g(T)=T^{0.6}$)

h / Criteria	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(12.6504)		(10.3207)		(6.6240)		(6.3226)		(0.5847)	
$FPE^M 2$	1.0000	0.9979	1.0051	1.0012	0.9929	0.9934	0.9701	0.9884	0.9742	0.9881
	(7.4082)		(7.2595)		(4.3451)		(0.7424)		(0.0135)	
FPE	0.9952	0.9959	0.9970	0.9988	0.9939	0.9973	1.0026	0.9975	1.0271	1.0126
	(11.9131)		(10.2537)		(7.0026)		(7.5171)		(2.1249)	
AIC	0.9952	0.9959	0.9970	0.9988	0.9939	0.9973	1.0026	0.9975	1.0271	1.0126
	(11.9131)		(10.2537)		(7.0026)		(7.5171)		(2.1249)	
BIC	0.9997	0.9983	1.0105	1.0056	0.9997	0.9999	0.9710	0.9869	0.9914	0.9956
	(4.7663)		(6.5010)		(4.1539)		(0.9008)		(0.2428)	

(Parentheses represent the mean lag order)

Table 7 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 1: T=500, $g(T)=T^{0.7}$)

h / Criteria	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^{M}1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(11.4688)		(9.9511)		(5.7914)		(4.3310)		(0.5396)	
$FPE^{M}2$	0.9990	0.9972	1.0055	1.001	1.0038	1.0017	0.9670	0.9857	0.9732	0.9872
	(6.9446)		(7.1333)		(3.9408)		(0.6265)		(0.0064)	
FPE	0.9969	0.9988	0.9977	0.9989	1.0004	1.0017	1.0006	0.9950	1.0263	1.0118
	(11.9131)		(10.2537)		(7.0026)		(7.5171)		(2.1249)	
AIC	0.9969	0.9988	0.9977	0.9989	1.0004	1.0017	1.0006	0.9950	1.0263	1.0118
	(11.9131)		(10.2537)		(7.0026)		(7.5171)		(2.1249)	
BIC	1.0014	1.0013	1.0112	1.0057	1.0062	1.0044	0.9690	0.9845	0.9906	0.9949
	(4.7663)		(6.5010)		(4.1539)		(0.9008)		(0.2428)	

(Parentheses represent the mean lag order)

Table 8 - MSE and MAE of Forecast Errors Relative to FPE

(Series 1: T=500, $g(T)=T^{0.8}$)

h / Criteria	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^{M}1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(8.8178)		(8.5293)		(4.3168)		(1.8680)		(0.3432)	
$FPE^{M}2$	1.0068	1.0042	1.0135	1.0035	1.0169	1.0122	0.9819	0.9911	0.9855	0.9939
	(5.2730)		(6.3071)		(2.9472)		(0.3967)		(0.0006)	
FPE	0.9961	1.0001	0.9998	0.9995	1.0001	1.0032	1.0228	1.0046	1.0366	1.0170
	(11.9131)		(10.2537)		(7.0026)		(7.5171)		(2.1249)	
AIC	0.9961	1.0001	0.9998	0.9995	1.0001	1.0032	1.0228	1.0046	1.0366	1.0126
	(11.9131)		(10.2537)		(7.0026)		(7.5171)		(2.1249)	
BIC	1.0006	1.0026	1.0134	1.0063	1.0060	1.0058	0.9906	0.9940	1.0005	1.0000
	(4.7663)		(6.5010)		(4.1539)		(0.9008)		(0.2428)	

(Parentheses represent the mean lag order)

Table 9 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 1: T=1000, $g(T)=T^{0.5}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(9.8348)		(9.5081)		(6.2612)		(2.6410)		(0.6638)	
$FPE^M 2$	1.0088	1.0053	1.0093	1.0055	1.0023	1.0038	0.9717	0.9850	0.9687	0.9862
	(5.9810)		(7.5973)		(4.3495)		(0.3799)		(0.0085)	
FPE	0.9990	1.0007	0.9989	1.0010	0.9929	0.9964	1.0093	0.9971	1.0309	1.0147
	(11.9801)		(10.1425)		(7.9212)		(7.2802)		(2.8424)	
AIC	0.9990	1.0007	0.9989	1.0010	0.9929	0.9964	1.0093	0.9971	1.0309	1.0147
	(11.9801)		(10.1425)		(7.9212)		(7.2802)		(2.8424)	
BIC	1.0063	1.0072	1.0137	1.0083	0.9987	0.9998	0.9768	0.9853	0.9895	0.9946
	(5.0237)		(6.9373)		(4.6059)		(0.8784)		(0.2270)	

(Parentheses represent the mean lag order)

Table 10 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 1: T=1000, $g(T)=T^{0.6}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(10.9772)		(9.9525)		(7.2896)		(4.6800)		(0.6638)	
$FPE^M 2$	1.0078	1.0026	1.0059	1.0013	0.9918	0.9941	0.9648	0.9800	0.9687	0.9846
	(6.3495)		(7.7284)		(4.6876)		(0.4596)		(0.0085)	
FPE	0.9986	0.9981	0.9988	0.9996	0.9914	0.9955	0.9983	0.9901	1.0312	1.0149
	(11.9801)		(10.1425)		(7.9212)		(7.2802)		(2.8424)	
AIC	0.9986	0.9981	0.9988	0.9996	0.9914	0.9955	0.9983	0.9901	1.0312	1.0149
	(11.9801)		(10.1425)		(7.9212)		(7.2802)		(2.8424)	
BIC	1.0059	1.0046	1.0136	1.0069	0.9973	0.9990	0.9661	0.9784	0.9897	0.9947
	(5.0237)		(6.9373)		(4.6059)		(0.8784)		(0.2270)	

(Parentheses represent the mean lag order)

Table 11 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 1: T=1000, $g(T)=T^{0.7}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(10.4539)		(9.7673)		(6.4520)		(3.3115)		(0.6638)	
$FPE^M 2$	1.0074	1.0033	1.0066	1.0027	1.0077	1.0044	0.9641	0.9913	0.9687	0.9846
	(6.1852)		(7.6733)		(4.4302)		(0.3704)		(0.0085)	
FPE	0.9975	0.9987	1.0002	1.0014	0.9983	0.9997	0.9987	0.9913	1.0312	1.0149
	(11.9801)		(10.1425)		(7.9212)		(7.2802)		(2.8424)	
AIC	0.9975	0.9987	1.0002	1.0014	0.9983	0.9997	0.9987	0.9901	1.0312	1.0149
	(11.9801)		(10.1425)		(7.9212)		(7.2802)		(2.8424)	
BIC	1.0049	1.0052	1.0151	1.0088	1.0042	1.0032	0.9665	0.9796	0.9897	0.9947
	(5.0237)		(6.9373)		(4.6059)		(0.8784)		(0.2270)	

(Parentheses represent the mean lag order)

Table 12 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 1: T=1000, $g(T)=T^{0.8}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(8.2469)		(8.7502)		(4.9706)		(1.3922)		(0.4587)	
$FPE^M 2$	1.0066	1.0069	1.0120	1.0042	1.0223	1.0134	0.9832	0.9915	0.9826	0.9923
	(5.1368)		(6.8870)		(3.5280)		(0.2441)		(0.0009)	
FPE	0.9945	0.9988	0.9982	0.9987	0.9946	0.9977	1.0267	1.0061	1.0426	1.0208
	(11.9801)		(10.1425)		(7.9212)		(7.2802)		(2.8424)	
AIC	0.9945	0.9988	0.9982	0.9987	0.9946	0.9977	1.0267	1.0061	1.0426	1.0208
	(11.9801)		(10.1425)		(7.9212)		(7.2802)		(2.8424)	
BIC	1.0019	1.0053	1.0131	1.0060	1.0005	1.0012	0.9936	0.9942	1.0007	1.0005
	(5.0237)		(6.9373)		(4.6059)		(0.8784)		(0.2270)	

(Parentheses represent the mean lag order)

Table 13 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 1: T=1500, g(T)= $T^{0.5}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(10.5769)		(10.8626)		(6.8987)		(0.6184)		(0.8445)	
$FPE^M 2$	1.0087	1.0045	1.0074	1.0029	1.0124	1.0078	1.0096	1.0086	0.9715	0.9861
	(6.5118)		(8.8499)		(5.7378)		(0.3472)		(0.0163)	
FPE	1.0019	1.0012	1.0002	1.0001	0.9890	0.9937	1.0157	1.0035	1.0392	1.0190
	(12.4141)		(10.9403)		(9.1700)		(2.6546)		(4.0958)	
AIC	1.0019	1.0012	1.0002	1.0001	0.9890	0.9937	1.0157	1.0035	1.0392	1.0190
	(12.4141)		(10.9403)		(9.1700)		(2.6546)		(4.0958)	
BIC	1.0089	1.0103	1.0116	1.0075	1.0174	1.0094	1.0002	0.9991	0.9844	0.9919
	(5.0452)		(7.7848)		(5.4611)		(0.6781)		(0.3219)	

(Parentheses represent the mean lag order)

Table 14 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 1: T=1500, g(T)= $T^{0.6}$ or $T^{0.7}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(10.6944)		(10.9222)		(6.9783)		(0.6184)		(0.8445)	
$FPE^M 2$	1.0064	1.0041	1.0057	1.0016	1.0122	1.0067	1.0095	1.0079	0.9714	0.9854
	(6.8373)		(8.9367)		(5.7631)		(0.3472)		(0.0163)	
FPE	1.0008	1.0010	1.0012	1.0008	0.9900	0.9942	1.0162	1.0038	1.0397	1.0193
	(12.4141)		(10.9403)		(9.1700)		(2.6546)		(4.0958)	
AIC	1.0008	1.0010	1.0012	1.0008	0.9900	0.9942	1.0162	1.0038	1.0397	1.0193
	(12.4141)		(10.9403)		(9.1700)		(2.6546)		(4.0958)	
BIC	1.0079	1.0101	1.0126	1.0083	1.0184	1.0099	1.0007	0.9994	0.9849	0.9922
	(5.0452)		(7.7848)		(5.4611)		(0.6781)		(0.3219)	

(Parentheses represent the mean lag order)

Table 15 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 1: T=1500, g(T)=$T^{0.8}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(8.9060)		(9.9675)		(6.1989)		(0.5407)		(0.5642)	
$FPE^M 2$	1.0071	1.0101	1.0141	1.0046	1.0221	1.0108	1.0176	1.0106	0.9948	0.9984
	(5.4539)		(7.8553)		(5.0344)		(0.3454)		(0.0018)	
FPE	1.0003	1.0010	1.0025	0.9998	0.9852	0.9903	1.0195	1.0032	1.0584	1.0584
	(12.4141)		(10.9403)		(9.1700)		(2.6546)		(4.0958)	
AIC	1.0003	1.0010	1.0025	0.9998	0.9852	0.9903	1.0195	1.0032	1.0584	1.0584
	(12.4141)		(10.9403)		(9.1700)		(2.6546)		(4.0958)	
BIC	1.0073	1.0101	1.0139	1.0072	1.0134	1.0059	1.0039	0.9988	1.0026	1.0015
	(5.0452)		(7.7848)		(5.4611)		(0.6781)		(0.3219)	

(Parentheses represent the mean lag order)

Table 16 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 5: T=500, g(T)=$T^{0.5}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(12.1880)		(9.5010)		(8.7553)		(1.7624)		(0.1726)	
$FPE^M 2$	1.0070	1.0026	0.9987	1.0002	1.0150	1.0091	0.9895	0.9969	1.0051	1.0036
	(8.0135)		(6.9665)		(3.7579)		(0.6594)		(0.0670)	
FPE	0.9978	0.9992	0.9977	0.9977	0.9963	0.9958	1.0059	0.9991	1.0182	1.0087
	(11.4179)		(9.9433)		(7.8989)		(2.9028)		(0.5750)	
AIC	0.9978	0.9992	0.9977	0.9977	0.9963	0.9958	1.0059	0.9991	1.0182	1.0087
	(11.4179)		(9.9433)		(7.8989)		(2.9028)		(0.5750)	
BIC	1.0000	0.9960	1.0081	1.0063	1.0145	1.0101	0.9977	0.9994	0.9987	1.0001
	(4.9994)		(5.9285)		(3.5003)		(0.9008)		(0.1127)	

(Parentheses represent the mean lag order)

Table 17 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 5: T=500, $g(T)=T^{0.6}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(12.1487)		(9.4797)		(8.4746)		(1.7379)		(0.1668)	
$FPE^M 2$	1.0074	1.0039	1.0003	1.0003	1.0131	1.0075	0.9891	0.9968	1.0063	1.0041
	(7.9878)		(6.9549)		(3.7476)		(0.6439)		(0.0670)	
FPE	0.9992	1.0007	0.9979	0.9979	0.9964	0.9962	1.0050	0.9983	1.0194	1.0092
	(11.4179)		(9.9433)		(7.8989)		(2.9028)		(0.5750)	
AIC	0.9992	1.0007	0.9979	0.9979	0.9964	0.9962	1.0050	0.9983	1.0194	1.0092
	(11.4179)		(9.9433)		(7.8989)		(2.9028)		(0.5750)	
BIC	1.0014	0.9975	1.0083	1.0065	1.0145	1.0105	0.9968	0.9986	0.9998	1.0006
	(4.9994)		(5.9285)		(3.5003)		(0.8719)		(0.1127)	

(Parentheses represent the mean lag order)

Table 18 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 5: T=500, $g(T)=T^{0.7}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(11.6587)		(9.2917)		(7.6072)		(1.6098)		(0.1545)	
$FPE^M 2$	1.0061	1.0020	1.0031	1.0010	1.0144	1.0078	0.9912	0.9978	1.0064	1.0043
	(7.6735)		(6.9086)		(3.6265)		(0.6182)		(0.0670)	
FPE	0.9986	0.9989	0.9974	0.9974	0.9973	0.9975	1.0085	1.0000	1.0196	1.0094
	(11.4179)		(9.9433)		(7.8989)		(2.9028)		(0.5750)	
AIC	0.9986	0.9989	0.9974	0.9974	0.9973	0.9975	1.0085	1.0000	1.0196	1.0094
	(11.4179)		(9.9433)		(7.8989)		(2.9028)		(0.5750)	
BIC	1.0008	0.9957	1.0078	1.0060	1.0155	1.0118	1.0003	1.0003	1.0000	1.0008
	(4.9994)		(5.9285)		(3.5003)		(0.8719)		(0.1127)	

(Parentheses represent the mean lag order)

Table 19 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 5: T=500, $g(T)=T^{0.8}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(9.0670)		(8.0019)		(4.7656)		(0.8944)		(0.1056)	
$FPE^M 2$	1.0134	1.0057	1.0045	1.0019	1.0211	1.0106	1.0045	1.0033	1.0081	1.0051
	(5.6787)		(6.3123)		(2.7978)		(0.4076)		(0.0290)	
FPE	0.9991	0.9989	0.9970	0.9987	0.9965	0.9974	1.0262	1.0073	1.0211	1.0105
	(11.4179)		(9.9433)		(7.8989)		(2.9028)		(0.5750)	
AIC	0.9991	0.9989	0.9970	0.9987	0.9965	0.9974	1.0262	1.0073	1.0211	1.0105
	(11.4179)		(9.9433)		(7.8989)		(2.9028)		(0.5750)	
BIC	1.0013	0.9957	1.0074	1.0073	1.0147	1.0117	1.0178	1.0076	1.0015	1.0019
	(4.9994)		(5.9285)		(3.5003)		(0.8719)		(0.1127)	

(Parentheses represent the mean lag order)

Table 20 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 5: T=1000, $g(T)=T^{0.5}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(11.1510)		(8.9060)		(6.8509)		(0.5717)		(0.0617)	
$FPE^M 2$	1.0132	1.0054	1.0084	1.0082	1.0236	1.0137	1.0033	1.0050	1.0096	1.0062
	(7.8490)		(7.4558)		(4.0674)		(0.2953)		(0.0342)	
FPE	1.0011	1.0006	1.0017	0.9996	0.9898	0.9923	1.0188	1.0062	1.0223	1.0110
	(11.1918)		(10.3286)		(8.2830)		(2.4786)		(0.5461)	
AIC	1.0011	1.0006	1.0017	0.9996	0.9898	0.9923	1.0188	1.0062	1.0223	1.0110
	(11.1918)		(10.3286)		(8.2830)		(0.7189)		(0.5461)	
BIC	1.0080	1.0033	1.0165	1.0150	1.0236	1.0131	1.0102	1.0062	0.9990	1.0004
	(5.0009)		(6.3048)		(3.9516)		(0.9008)		0.0703	

(Parentheses represent the mean lag order)

Table 21 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 5: T=1000, $g(T)=T^{0.6}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(11.2460)		(8.9601)		(6.8661)		(0.5745)		(0.0617)	
$FPE^M 2$	1.0121	1.0050	1.0083	1.0081	1.0226	1.0128	1.0028	1.0048	1.0096	1.0062
	(7.9573)		(7.4558)		(4.1016)		(0.2953)		(0.0342)	
FPE	1.0011	1.0009	1.0017	0.9995	0.9909	0.9927	1.0183	1.0059	1.0223	1.0110
	(11.1918)		(10.3286)		(8.2830)		(2.4786)		(0.5461)	
AIC	1.0011	1.0009	1.0017	0.9995	0.9909	0.9927	1.0183	1.0059	1.0223	1.0110
	(11.1918)		(10.3286)		(8.2830)		(2.4786)		(0.5461)	
BIC	1.0080	1.0036	1.0164	1.0148	1.0247	1.0135	1.0097	1.0060	0.9990	1.0004
	(5.0009)		(6.3048)		(3.9516)		(0.7189)		(0.0703)	

(Parentheses represent the mean lag order)

Table 22 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 5: T=1000, $g(T)=T^{0.7}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(11.1899)		(8.8946)		(6.8613)		(0.56600)		(0.0617)	
$FPE^M 2$	1.0130	1.0056	1.0071	1.0075	1.0228	1.0126	1.0047	1.0056	1.0096	1.0062
	(7.8186)		(7.4558)		(4.0712)		(0.2953)		(0.0342)	
FPE	1.0011	1.0005	1.0005	0.9989	0.9909	0.9929	1.0202	1.0068	1.0223	1.0110
	(11.1918)		(10.3286)		(8.2830)		(2.4786)		(0.5461)	
AIC	1.0011	1.0005	1.0005	0.9989	0.9909	0.9929	1.0202	1.0068	1.0223	1.0110
	(11.1918)		(10.3286)		(8.2830)		(2.4786)		(0.5461)	
BIC	1.0081	1.0031	1.0152	1.0142	1.0248	1.0137	1.0116	1.0068	0.9990	1.0004
	(5.0009)		(6.3048)		(3.9516)		(0.7189)		(0.0703)	

(Parentheses represent the mean lag order)

Table 23 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 5: T=1000, $g(T)=T^{0.8}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(9.7360)		(8.1567)		(6.0019)		(0.4767)		(0.0617)	
$FPE^M 2$	1.0105	1.0052	1.0032	1.0039	1.0275	1.0151	1.0120	1.0087	1.0096	1.0062
	(6.0855)		(7.0513)		(3.4520)		(0.2830)		(0.0342)	
FPE	0.9965	0.9975	0.9974	0.9957	0.9829	0.9895	1.0263	1.0089	1.0222	1.0110
	(11.1918)		(10.3286)		(8.2830)		(2.4786)		(0.5461)	
AIC	0.9965	0.9975	0.9974	0.9957	0.9829	0.9895	1.0263	1.0089	1.0222	1.0110
	(11.1918)		(10.3286)		(8.2830)		(0.7189)		(0.5461)	
BIC	1.0034	1.0002	1.0121	1.0110	1.0165	1.0102	1.0176	1.0089	0.9989	1.0004
	(5.0009)		(6.3048)		(3.9516)		(0.9008)		(0.0703)	

(Parentheses represent the mean lag order)

Table 24 - MSE and MAE of Forecast Errors Relative to FPE_M1

(Series 5: T=1500, $g(T)=T^{0.5}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(10.0000)		(8.9349)		(9.8427)		(0.8825)		(0.1175)	
$FPE^M 2$	1.0121	1.0067	1.0030	1.0026	1.0366	1.0196	1.0285	1.0176	1.0227	1.0128
	(7.4485)		(8.7722)		(5.1139)		(0.5624)		(0.0651)	
FPE	1.0004	1.0004	1.0039	1.0001	0.9913	0.9942	0.9759	0.9875	1.0031	1.0017
	(10.0488)		(11.5298)		(11.2297)		(1.5244)		(0.6347)	
AIC	1.0004	1.0001	1.0039	1.0001	0.9913	0.9942	0.9759	0.9875	1.0031	1.0017
	(10.0488)		(11.5298)		(11.2297)		(1.5244)		(0.6347)	
BIC	1.0089	1.0066	1.0190	1.0150	1.0420	1.0213	1.0013	1.0015	0.9977	1.0009
	(5.0018)		(6.5805)		(4.8156)		(0.8969)		(0.1338)	

(Parentheses represent the mean lag order)

Table 25 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 5: T=1500, g(T)=$T^{0.6}$, $T^{0.7}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(10.0000)		(8.9349)		(9.8427)		(0.8825)		(0.1175)	
$FPE^M 2$	1.0121	1.0067	1.0030	1.0026	1.0366	1.0196	1.0285	1.0176	1.0227	1.0128
	(7.4485)		(8.7722)		(5.1139)		(0.5624)		(0.0651)	
FPE	1.0004	1.0001	1.0039	1.0001	0.9913	0.9942	0.9759	0.9875	1.0031	1.0017
	(10.0488)		(11.5298)		(11.2297)		(1.5244)		(0.6347)	
AIC	1.0004	1.0001	1.0039	1.0001	0.9913	0.9942	0.9759	0.9875	1.0031	1.0017
	(10.0488)		(11.5298)		(11.2297)		(1.5244)		(0.6347)	
BIC	1.0089	1.0066	1.0190	1.0150	1.0420	1.0213	1.0013	1.0015	0.9977	1.0009
	(5.0018)		(6.5805)		(4.8156)		(0.8969)		(0.1338)	

(Parentheses represent the mean lag order)

Table 26 - MSE and MAE of Forecast Errors Relative to FPE_M1
(Series 5: T=1500, g(T)=$T^{0.8}$)

Criteria \ h	1		5		22		132		252	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
$FPE^M 1$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	(9.6474)		(8.8535)		(8.7902)		(0.8680)		(0.1175)	
$FPE^M 2$	1.0147	1.0107	1.0018	1.0019	1.0318	1.0181	1.0290	1.0186	1.0243	1.0143
	(6.2966)		(8.0018)		(4.6944)		(0.5389)		(0.0651)	
FPE	1.0005	1.0089	1.0042	1.0000	0.9823	0.9901	0.9737	0.9866	1.0028	1.0017
	(10.0488)		(11.5298)		(11.2297)		(1.5244)		(0.6347)	
AIC	1.0005	1.0089	1.0042	1.0000	0.9823	0.9901	0.9737	0.9866	1.0028	1.0017
	(10.0488)		(11.5298)		(11.2297)		(1.5244)		(0.6347)	
BIC	1.0089	1.0080	1.0193	1.0149	1.0325	1.0171	0.9990	1.0006	0.9974	1.0008
	(5.0018)		(6.5805)		(4.8156)		(0.8969)		(0.1338)	

(Parentheses represent the mean lag order)