Modeling and Forecasting Stock Return Volatility Using a Random Level Shift Model*

Yang K. Lu† | Pierre Perron‡

Boston University | **Boston University**

July 28, 2008; Revised August 4, 2009.

Abstract

We consider the estimation of a random level shift model for which the series of interest is the sum of a short memory process and a jump or level shift component. For the latter component, we specify the commonly used simple mixture model such that the component is the cumulative sum of a process which is 0 with some probability $(1-\alpha)$ and is a random variable with probability α. Our estimation method transforms such a model into a linear state space with mixture of normal innovations, so that an extension of Kalman filter algorithm can be applied. We apply this random level shift model to the logarithm of absolute returns for the S&P 500, AMEX, Dow Jones and NASDAQ stock market return indices. Our point estimates imply few level shifts for all series. But once these are taken into account, there is little evidence of serial correlation in the remaining noise and, hence, no evidence of long-memory. Once the estimated shifts are introduced to a standard GARCH model applied to the returns series, any evidence of GARCH effects disappears. We also produce rolling out-of-sample forecasts of squared returns. In most cases, our simple random level shifts model clearly outperforms a standard GARCH(1,1) model and, in many cases, it also provides better forecasts than a fractionally integrated GARCH model.

JEL Classification Number: C22.

Keywords: structural change, forecasting, GARCH models, long-memory.

---

*Perron acknowledges financial support for this work from the National Science Foundation under Grant SES-0649350. We are grateful to Zhongjun Qu and Adrien Verdelhan for helpful comments, and to Adam McCloskey for detailed comments on a previous draft that improved the presentation.

†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (yanglu@bu.edu).

‡Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

1 Introduction

Recently, there has been an upsurge of interest in the possibility of confusing long-memory with structural change in levels. This idea extends that exposited by Perron (1989, 1990) who showed that structural change and unit roots are easily confused: when a stationary process is contaminated by structural change, the estimate of the sum of its autoregressive coefficients is biased towards 1 and tests of the null hypothesis of a unit root are biased toward non-rejection. This phenomenon has been shown to apply to the long-memory context as well. That is, when a stationary short-memory process is contaminated by structural change in levels, the estimate of the long-memory parameter d is biased away from 0 and the autocovariance function of the process exhibits a slow rate of decay. Relevant references on this issue include Diebold and Inoue (2001), Engle and Smith (1999), Gourieroux and Jasiak (2001), Granger and Ding (1996), Granger and Hyung (2004), Lobato and Savin (1998), Mikosch and Stărică (2004), Parke (1999) and Teverosovky and Taqqu (1997).

The literature on modeling and forecasting stock return volatility is voluminous. Two approaches that have proven useful are the GARCH and stochastic volatility (SV) models[1]. In their standard forms, the ensuing volatility processes are stationary and weakly dependent with autocorrelations that decrease exponentially. This contrasts with the empirical findings obtained using various proxies for volatility (e.g., daily absolute returns) which indicate autocorrelations that decay very slowly at long lags. In light of this, several long-memory models have been proposed. For example, Baillie, Bollerslev and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996) considered fractionally integrated GARCH and EGARCH models, while Breidt, Crato and De Lima (1998) and Harvey (1998) proposed long memory SV (LSV) models where the log of volatility is modelled as a fractionally integrated process.

More recently, attempts have been made to distinguish between stationary noise plus level shift and long-memory models; see, in particular, Granger and Hyung (2004). They documented the fact that, when breaks determined via some pre tests are accounted for, the evidence for long-memory is weaker. This evidence is, however, inconclusive since structural change tests are severely biased in the presence of long-memory and log periodogram estimates of the memory parameter are biased downward when sample-selected breaks are introduced. This is an overfitting problem that Granger and Hyung (2004, p. 416) clearly recognized. Stărică and Granger (2005) presented evidence that log-absolute returns of the S&P 500 index is an *i.i.d.* series affected by occasional shift in the unconditional variance

[1] For extensive reviews and collected works, see Engle (1995) and Shephard (2005).

and show that this specification has better forecasting performance than the more traditional GARCH(1,1) model and its fractionally integrated counterpart. Mikosch and Stărică (2005) considered the autocorrelation function of the absolute returns of the S&P 500 index for the period 1953-1977. They documented the fact that for the full period, it resembles that of a long-memory process. But, interestingly, if one omits the last fours years of data, the autocorrelation function is very different and looks like one associated with a short-memory process. They explain this finding by arguing that the volatility of the S&P 500 returns has increased over the period 1973-1977. Morana and Beltratti (2004) also argue that breaks in the level of volatility partially explain the long-memory features of some exchange rate series. Perron and Qu (2007) analyzed the time and spectral domain properties of a stationary short memory process affected by random level shifts. Perron and Qu (2008) showed that, when applied to daily S&P 500 absolute returns, their square roots and log absolute returns over the period 1928-2002, the level shift model explains both the shape of the autocorrelations and the path of log periodogram estimates as a function of the number of frequency ordinates used. Qu and Perron (2008) estimated a stochastic volatility model with level shifts using a Bayesian approach using daily data on returns from the S&P 500 and NASDAQ indices over the period 1980.1-2005.12. They showed that the level shifts account for most of the variation in volatility, that their model provides a better in-sample fit than alternative models and that its forecasting performance is better for the NASDAQ and just as good for the S&P 500 as standard short or long-memory models without level shifts.

Our approach extends the work of Stărică and Granger (2005) by directly estimating a structural model. We adopt a specification for which the series of interest is the sum of a short-memory process and a jump or level shift component. For the latter, we specify a simple mixture model such that the component is the cumulative sum of a process which is 0 with some probability $(1-\alpha)$ and is a random variable with probability α. To estimate such a model, we transform it into a linear state space form with innovations having a mixture of two normal distributions and adopt an algorithm similar to the one used by Perron and Wada (2009) and Wada and Perron (2007). We restrict the variance of one of the two normal distributions to be zero, allowing us to achieve a simple but efficient algorithm.

We apply this random level shift model to the logarithm of absolute returns for the following stock market return indices: S&P 500 (1962/07/03 to 2004/03/25; 10504 observations), AMEX (1962/07/03 to 2006/12/31; 11201 observations), Dow Jones (1957/03/04 to 2002/10/30; 11534 observations) and NASDAQ (1972/12/15 to 2006/12/31; 8592 observations). Our point estimates imply few level shifts for all series. But once these are taken into account, there is little evidence of serial correlation in the remaining noise and, hence, no evidence of long-memory. Furthermore, once the estimated shifts are introduced to a standard GARCH model, any evidence of GARCH effects disappears. We also produce recursive out-of-sample forecasts of squared returns. In most cases, our simple random level shifts model clearly outperforms a standard GARCH(1,1) model and, in many cases, it also provides better forecasts than a fractionally integrated GARCH model.

A few comments on our modeling and forecasting strategy are in order. A main goal is to provide forecasts of volatility proxied by daily squared returns. We however, apply our level shift model to log-absolute returns since they do not suffer from a non-negativity constraint as do, say, absolute or squared returns. There is also no loss relative to using squared returns in identifying level shift since log-absolute returns are a monotonic transformation. Since we wish to identify the probability of shifts and their locations, the fact that log-absolute returns are quite noisy is not problematic since our methods are robust to the presence of noise. Another reason is the fact that for many asset returns, a log-absolute transformation yields a series that is closer to being normally distributed (see, e.g., Andersen, Bollerslev, Diebold and Labys, 2001). A second comment of interest is the fact we use daily returns as opposed to realized volatility series constructed from intra-daily high frequency data which has recently become popular. While realized volatility series are a less noisy measure of volatility than daily squared returns, their use in our context is problematic. First, such series are typically available for a short time span. Given the fact that level shifts will turn out to be relatively rare, it is imperative to have a long span of data available to obtain reliable estimates of the probability of occurrence of level shifts. Secondly, such series are available only for specific assets, as opposed to market indices. In our framework, the intent of the level shift model is to have a framework which allows for special events affecting overall markets. Using data on a specific asset would confound such market-wide events with idiosyncratic ones associated with the particular asset used. Finally, we wish to re-evaluate the adequacy of GARCH models applied to daily returns when taking into account the possibility of level shifts. Hence, it is important to have estimates of these level shifts for squared daily returns which are equivalent to those obtained using log-absolute returns.

The structure of the paper is as follows. Section 2 presents the model and the specifications adopted. Section 3 discusses the estimation procedure while the estimation results and various diagnostics are presented in Section 4. The forecasting comparisons between our model and GARCH and FIGARCH models are reported in Section 5. Section 6 offers brief conclusions.

2 The model

The model we apply to log-absolute returns is given by

$$y_t = a + \tau_t + c_t, \tag{1}$$

where a is a constant, τ_t is the random level shift components and c_t is a short-memory process, included to model the remaining noise. The level shift component is specified by

$$\tau_t = \tau_{t-1} + \delta_t,$$

where

$$\delta_t = \pi_t \eta_t.$$

Here, π_t is a binomial variable that takes value 1 with probability α and value 0 with probability $(1 - \alpha)$. If it takes value 1, then a random level shift η_t occurs, specified by $\eta_t \sim i.i.d.\ N(0, \sigma_\eta^2)$.

In its most general form, the short-memory process c_t is defined by $c_t = C(L) e_t$, with $e_t \sim i.i.d.\ N(0, \sigma_e^2)$ and $E|e_t|^r < \infty$ for some $r > 2$. The polynomial $C(L)$ satisfies $C(L) = \sum_{i=0}^{\infty} c_i L^i$, $\sum_{i=0}^{\infty} i |c_i| < \infty$ and $C(1) \neq 0$. These conditions allow us to approximate the process by a finite order autoregression, provided its order is chosen suitably. As we shall see, once the level shifts are accounted for, very little autocorrelation remains in the process so that an AR(1) specification of the short-memory component is sufficient. Even then, the autoregressive parameter estimate will be small and insignificant. Hence, in what follows, we shall for simplicity assume that

$$c_t = \phi c_{t-1} + e_t,$$

understanding that our method of estimation can easily be extended to cover more general processes with straightforward modifications. We also assume that the components π_t, η_t and c_t are mutually independent. The normality assumption for e_t is needed to construct the likelihood function, though the parameter estimates remain consistent without it.

It is easy to confuse this model with the popular Markov regime switching model (e.g., see Hamilton 1989) since τ_t can be interpreted as a latent regime variable. The two models share the feature of allowing the time series to follow different processes over different subsamples. But there is a key distinction that allows our model to have more flexibility. In the standard Markov regime switching model, only a finite number of possible regimes, usually two or some other small value, are allowed (see, for example, Hamilton 1989, Gray 1996 and Filardo and Gordon 1998). In contrast, our model does not restrict the magnitude of the level shifts since they are draws from a normal distribution. Hence, any number of regimes is possible. It is also important to note that the probability that π_t be 0 or 1 is independent of past realizations, unlike the Markov switching type models. Here, the different regimes are associated with different magnitudes of the shifts. Our goal is to develop a framework which allows for special events. Hence, the probability that a shift occurs should be independent of past shifts. Accordingly, it is appropriate to make the probability of being in a given regime as independent of past realizations.

In order to estimate the model, we shall embed it in a state space framework involving errors that have a mixture of two normal distributions. The level shift component τ_t can be specified as a random walk process with innovations distributed according to a mixture of two normally distributed processes:

$$\begin{aligned} \tau_t &= \tau_{t-1} + \delta_t, \\ \delta_t &= \pi_t \eta_{1t} + (1 - \pi_t)\, \eta_{2t}, \end{aligned}$$

where $\eta_{it} \sim i.i.d.\ N(0, \sigma^2_{\eta i})$ and π_t is a Bernoulli random variable that takes value one with probability α and value 0 with probability $1 - \alpha$. By specifying $\sigma^2_{\eta 1} = \sigma^2_{\eta}$ and $\sigma^2_{\eta 2} = 0$, we recover our level shift model.

We next specify the model in terms of first-differences of the data:

$$\Delta y_t = \tau_t - \tau_{t-1} + c_t - c_{t-1} = \delta_t + c_t - c_{t-1},$$

where

$$\delta_t = \pi_t \eta_{1t} + (1 - \pi_t)\, \eta_{2t}.$$

We then have the following state space form:

$$\begin{aligned} \Delta y_t &= c_t - c_{t-1} + \delta_t, \\ c_t &= \phi c_{t-1} + e_t, \end{aligned}$$

or more generally,

$$\begin{aligned} \Delta y_t &= H X_t + \delta_t, \\ X_t &= F X_{t-1} + U_t, \end{aligned}$$

where, in the case of an $AR(p)$ process,

$$X_t = [c_t, c_{t-1}, ..., c_{t-p-1}]',$$

$$F=\begin{bmatrix}\phi_1 & \phi_2 & \cdots & \phi_p\\ 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{bmatrix},$$

$H=[1,-1,0,...,0]$ and U_t is a p-dimensional normally distributed random vector with mean zero and covariance matrix

$$Q=\begin{pmatrix}\sigma_e^2 & 0_{1\times(p-1)}\\ 0_{(p-1)\times 1} & 0_{(p-1)\times(p-1)}\end{pmatrix}.$$

The important difference between this model and a standard state space model is that the distribution of δ_t is a mixture of two normal distributions with variances σ_η^2 and 0, occurring with probabilities α and $1-\alpha$, respectively. Note that it is also easy to extend our model to have the short-memory component c_t follow an ARMA process as long as c_t can be represented in a state space form.

3 The estimation method

The model described in the previous section is a special case of the class of models considered by Wada and Perron (2007). They modeled the trend-cycle decomposition of a macroeconomic time series allowing for mixtures of normal distributions for the shocks affecting the level, slope and cyclical components. Here, we only have shocks affecting the level of the series and we simply need to impose the restriction that the variance of one component of the mixture of normal distributions is zero. In what follows, we provide a brief description of the method, more details can be found in Wada and Perron (2007).

The first thing to note is that despite their fundamental differences, our model and the Markov regime switching models share similarities in estimation methodology. The basic ingredient for estimation is the augmentation of the states by the realizations of the mixture at time t so that the Kalman filter can be used to generate the likelihood function, conditional on the realizations of the states. The latent states are eliminated from the final likelihood expression by summing over all possible state realizations. As we shall show, our model takes a structure that is similar to a version of the Markov regime switching model.

Let $Y_t = (\Delta y_1, ..., \Delta y_t)$ be the vector of data available up to time t and denote the vector of parameters by $\theta = [\sigma_\eta^2, \alpha, \sigma_e^2, \phi_1, ..., \phi_q]$. To illustrate the similarities we adopt the notations in Hamilton (1994), where $\mathbf{1}$ represents a (4×1) vector of ones, the symbol $\odot$ denotes element-by-element multiplication, $\hat{\xi}_{t|t-1} = vec(\widetilde{\xi}_{t|t-1})$ with the $(i,j)^{th}$ element of $\widetilde{\xi}_{t|t-1}$ being $\Pr(s_{t-1}=i, s_t=j|Y_{t-1};\theta)$ and $\boldsymbol{\omega}_t = vec(\widetilde{\boldsymbol{\omega}}_t)$ with the $(i,j)^{th}$ element of $\widetilde{\boldsymbol{\omega}}_t$ being $f(\Delta y_t|s_{t-1}=i, s_t=j, Y_{t-1};\theta)$ for $i,j\in\{1,2\}$. Here $s_t=1$ (resp., 0) when $\pi_t=1$ (resp., 0), i.e., a level shift occurs (resp., does not occur). The log likelihood function is

$$\ln(L) = \sum_{t=1}^{T} \ln f(\Delta y_t|Y_{t-1};\theta), \tag{2}$$

where

$$\begin{aligned} f(\Delta y_t|Y_{t-1};\theta) &= \sum_{i=1}^{2}\sum_{j=1}^{2} f(\Delta y_t|s_{t-1}=i, s_t=j, Y_{t-1};\theta)\Pr(s_{t-1}=i, s_t=j|Y_{t-1};\theta) \\ &\equiv \mathbf{1}'\left(\hat{\xi}_{t|t-1}\odot\boldsymbol{\omega}_t\right). \end{aligned} \tag{3}$$

We first focus on the evolution of $\hat{\xi}_{t|t-1}$. Applying rules for conditional probabilities, Bayes' rule and the independence of s_t with past realizations, we have

$$\begin{aligned} \widetilde{\xi}^{ij}_{t|t-1} &\equiv \Pr(s_{t-1}=i, s_t=j|Y_{t-1};\theta) \\ &= \Pr(s_t=j)\sum_{k=1}^{2}\Pr(s_{t-2}=k, s_{t-1}=i|Y_{t-1};\theta), \end{aligned}$$

and

$$\begin{aligned} \widetilde{\xi}^{ki}_{t-1|t-1} &\equiv \Pr(s_{t-2}=k, s_{t-1}=i|Y_{t-1};\theta) \\ &= \frac{f(\Delta y_{t-1}|s_{t-2}=k, s_{t-1}=i, Y_{t-2};\theta)\Pr(s_{t-2}=k, s_{t-1}=i|Y_{t-2};\theta)}{f(\Delta y_{t-1}|Y_{t-2};\theta)}. \end{aligned}$$

Therefore, the evolution of $\hat{\xi}_{t|t-1}$ is given by:

$$\begin{bmatrix} \widetilde{\xi}^{11}_{t+1|t} \\ \widetilde{\xi}^{21}_{t+1|t} \\ \widetilde{\xi}^{12}_{t+1|t} \\ \widetilde{\xi}^{22}_{t+1|t} \end{bmatrix} = \begin{bmatrix} \alpha\left(\widetilde{\xi}^{11}_{t|t}+\widetilde{\xi}^{21}_{t|t}\right) \\ \alpha\left(\widetilde{\xi}^{12}_{t|t}+\widetilde{\xi}^{22}_{t|t}\right) \\ (1-\alpha)\left(\widetilde{\xi}^{11}_{t|t}+\widetilde{\xi}^{21}_{t|t}\right) \\ (1-\alpha)\left(\widetilde{\xi}^{12}_{t|t}+\widetilde{\xi}^{22}_{t|t}\right) \end{bmatrix} = \begin{bmatrix} \alpha & \alpha & 0 & 0 \\ 0 & 0 & \alpha & \alpha \\ (1-\alpha) & (1-\alpha) & 0 & 0 \\ 0 & 0 & (1-\alpha) & (1-\alpha) \end{bmatrix} \begin{bmatrix} \widetilde{\xi}^{11}_{t|t} \\ \widetilde{\xi}^{21}_{t|t} \\ \widetilde{\xi}^{12}_{t|t} \\ \widetilde{\xi}^{22}_{t|t} \end{bmatrix}.$$

or more compactly by:

$$\hat{\xi}_{t+1|t} = \Pi\hat{\xi}_{t|t}, \tag{4}$$

with

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot \boldsymbol{\omega}_t}{\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \boldsymbol{\omega}_t)}.$$

The conditional likelihood for Δy_t is the following Normal density:

$$\widetilde{\boldsymbol{\omega}}_t^{ij} = f\left(\Delta y_t | s_{t-1} = i, s_t = j, Y_{t-1}; \theta\right) = \frac{1}{\sqrt{2\pi}} \left| f_t^{ij} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{\upsilon_t^{ij\prime} \left(f_t^{ij} \right)^{-1} \upsilon_t^{ij}}{2} \right\}, \tag{5}$$

where $\upsilon_t^{ij} = \Delta y_t - \Delta y_{t|t-1}^i$ is the prediction error and $f_t^{ij} = E(\upsilon_t^{ij} \upsilon_t^{ij\prime})$ is the prediction error variance. Note that $\Delta y_{t|t-1}^i = E[\Delta y_t | s_{t-1} = i, Y_{t-1}; \theta]$ does not depend on the state j at time t because we are conditioning on time $t-1$ information. However, Δy_t does depend on $s_t = j$ so that the prediction error and its variance depend on both i and j. The best forecast for the state variable and its associated variance conditional on past information and $s_{t-1} = i$ are

$$\begin{aligned} X_{t|t-1}^i &= F X_{t-1|t-1}^i, \\ P_{t|t-1}^i &= F P_{t-1|t-1}^i F' + Q. \end{aligned} \tag{6}$$

We have the measurement equation $\Delta y_t = H X_t + \delta_t$, where the measurement error δ_t has mean zero and a variance which can take two possible values: $R_1 = \sigma_\eta^2$, with probability α, or $R_2 = 0$, with probability $1 - \alpha$. Hence, the prediction error is $\upsilon_t^{ij} = \Delta y_t - H X_{t|t-1}^i$ with associated variance $f_t^{ij} = H P_{t|t-1}^i H' + R_j$. Applying standard updating formulae, we have given $s_t = j$ and $s_{t-1} = i$,

$$\begin{aligned} X_{t|t}^{ij} &= X_{t|t-1}^i + P_{t|t-1}^i H' \left(H P_{t|t-1}^i H' + R_j \right)^{-1} \left(\Delta y_t - H X_{t|t-1}^i \right), \\ P_{t|t}^{ij} &= P_{t|t-1}^i - P_{t|t-1}^i H' \left(H P_{t|t-1}^i H' + R_j \right)^{-1} H P_{t|t-1}^i. \end{aligned} \tag{7}$$

To reduce the dimension of the estimation problem, we adopt the re-collapsing procedure suggested by Harrison and Stevens (1976), given by

$$\begin{aligned} X_{t|t}^j &= \frac{\sum_{i=1}^2 \Pr\left(s_{t-1} = i, s_t = j | Y_t; \theta\right) X_{t|t}^{ij}}{\Pr\left(s_t = j | Y_t; \theta\right)} = \frac{\sum_{i=1}^2 \widetilde{\xi}_{t|t}^{ij} X_{t|t}^{ij}}{\sum_{i=1}^2 \widetilde{\xi}_{t|t}^{ij}}, \\ P_{t|t}^j &= \frac{\sum_{i=1}^2 \Pr\left(s_{t-1} = i, s_t = j | Y_t; \theta\right) \left[P_{t|t}^{ij} + \left(X_{t|t}^j - X_{t|t}^{ij} \right) \left(X_{t|t}^j - X_{t|t}^{ij} \right)' \right]}{\Pr\left(s_t = j | Y_t; \theta\right)} \\ &= \frac{\sum_{i=1}^2 \widetilde{\xi}_{t|t}^{ij} \left[P_{t|t}^{ij} + \left(X_{t|t}^j - X_{t|t}^{ij} \right) \left(X_{t|t}^j - X_{t|t}^{ij} \right)' \right]}{\sum_{i=1}^2 \widetilde{\xi}_{t|t}^{ij}}. \end{aligned} \tag{8}$$

By doing so, we make $\boldsymbol{\omega}_t^{ij}$ unaffected by the history of states before time $t-1$.

If we define $S_t \equiv (s_t, s_{t-1})$, we then have four possible states corresponding to $S_t = 1$ when $(s_t = 1, s_{t-1} = 1)$, $S_t = 2$ when $(s_t = 1, s_{t-1} = 2)$, $S_t = 3$ when $(s_t = 2, s_{t-1} = 1)$ and $S_t = 4$ when $(s_t = 2, s_{t-1} = 2)$ with the transition matrix Π as defined in (4). The vector of conditional densities $\boldsymbol{\omega}_t = (\boldsymbol{\omega}_t^1, ..., \boldsymbol{\omega}_t^4)$ thus have a more compact representation given by:

$$\boldsymbol{\omega}_t^\ell = f\left(\Delta y_t | S_t = \ell, Y_{t-1}; \theta\right) = \frac{1}{\sqrt{2\pi}} \left|f_t^\ell\right|^{-\frac{1}{2}} \exp\left\{-\frac{\upsilon_t^{\ell\prime}\left(f_t^\ell\right)^{-1}\upsilon_t^\ell}{2}\right\} \tag{9}$$

where υ_t^ℓ and f_t^ℓ are as defined in (5) with the values of s_t and s_{t-1} corresponding to $S_t = \ell$. This definition of $\boldsymbol{\omega}_t$, together with $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t+1|t}$, the collection of conditional probabilities $\Pr\left(S_t = \ell | Y_t; \theta\right)$ for $\ell = 1, ..., 4$ and its the one-period ahead forecasts, evolving as in (4), give us the same structure as a version of the Markov regime switching model used by Hamilton (1994). However, there are two extra complexities here. Firstly, the mean and variance in the conditional density function are nonlinear functions of the fundamental parameters θ and past realizations $\{\Delta y_{t-j}; j \geq 1\}$. This non-linearity and time dependence complicate the maximization of the log-likelihood function since we cannot separate out some elements of θ in the first order conditions. Accordingly, the standard EM algorithm does not apply. Secondly, the conditional probability of being in a given regime $\hat{\xi}_{t|t}$ is not separable from the conditional densities $\boldsymbol{\omega}_t$ since $\widetilde{\xi}_{t|t}^{ij}$ enters in its construction (see (8)).

4 Empirical results for returns on stock market indices

In this section, we apply our model and estimation method to the returns of four major market indices: the S&P 500, AMEX, Dow Jones and NASDAQ. The daily returns are computed by first differencing the logarithm of the index price series $r_t = \ln(P_t) - \ln(P_{t-1})$. The data coverage is from 1962/07/03 to 2004/03/25 for the S&P 500 (10504 observations), from 1962/07/03 to 2006/12/31 for the AMEX (11201 observations), from 1957/03/04 to 2002/10/30 for the Dow Jones (11534 observations) and from 1972/12/15 to 2006/12/31 for the NASDAQ (8592 observations). For reasons stated in the introduction, we model log absolute returns. When returns are zero or close to it, the log absolute value transformation implies extreme negative values. Using our method, these outliers would be attributed to the level shift component and thus bias the probability of shifts upward. To avoid this problem, we bound absolute returns away from zero by adding a small constant, i.e., we use $\log(|r_t| + 0.001)$, a technique introduced to the stochastic volatility literature by Fuller (1996). The results are robust to alternative specifications, for example using another value for this so-called offset parameter, deleting the zero observations, or replacing them by a small value.

To specify the short-memory component c_t, we start by setting $c_t = e_t$, following Stărică and Granger (2005) who report that the short-memory component in such series is just white noise. As a robustness check, we also report the estimates for the specification that c_t follows an AR(1) process, $c_t = \phi c_{t-1} + e_t$. Since all components of the state vector are stationary, we can initialize the state vector and its covariance matrix by their unconditional expected values, i.e., $X_{0|0} = (0, 0)'$ and

$$P_{0|0} = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

We obtain estimates by directly maximizing the likelihood function (2). In order to avoid the problem of local maxima, we re-estimate with different initial values of θ_0 and pick the estimates associated with the largest likelihood value upon convergence.

4.1 Estimation results

The parameter estimates are presented in Table 1 for both the cases in which the short-memory component c_t is specified to be white noise and an AR(1) process. To assess the relative importance of the two components to the total variation of the series, we also report the standard deviation of the original series $\log |r_t|$. The estimates reported are the standard deviation of level shift component σ_η, the probability of a shift α, the standard deviation of the stationary component σ_e and the autoregressive coefficient ϕ when considering the AR(1) specification for c_t.

There results exhibit noteworthy features. First, when considering the AR(1) specification for c_t, the estimate of ϕ is very small and close to zero, except for the AMEX series for which it is the largest but still only 0.063. In all cases, adopting either an AR(1) or white noise specification for c_t yields very similar results (there are some slight differences for the AMEX series but the subsequent results to be discussed here are unaffected by the specification of c_t). Thus in subsequent sections we shall only consider results based on the white noise specification for the short-memory component. These results are in agreement with those of Stărică and Granger (2005).

The second noteworthy feature is the fact that the probability of level shifts is very small in all cases considered. Indeed, the point estimates of α imply the following number of shifts for each index: 15 for the S&P 500, 28 for the AMEX, 12 for the Dow Jones and 7 for the

NASDAQ. However, the shifts are important given that their standard deviation is of the same order as the standard deviation of the series. Nevertheless, given that the shifts occur so infrequently, the noise component accounts for the bulk of total variation.

4.2 The effect of level shifts on long-memory and conditional heteroskedasticity

Given the features documented above, it would be interesting to determine whether the level shifts are important to dictating the overall behavior of the series. We shall address this issue by investigating whether the shifts can explain a) the well-documented feature of long-memory and b) the presence of conditional heteroskedasticity.

To do so we need estimates of the dates at which the level shifts occur as well as the means within each segment. Ideally, one would use a smoothed estimate of the level shift component τ_t. However, in this context of multiple abrupt changes, conventional smoothers perform very poorly as we shall see. Hence, we also resort to using the following strategy. We use the point estimate of α to obtain a point estimate of the number of changes, as reported above. We then apply the method of Bai and Perron (2003) to obtain the estimates of the break dates that globally minimize the following sum of squared residuals:

$$\sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} [y_t - \mu_i]^2,$$

where m is the number of breaks, T_i $(i = 1, ..., m)$ are the break dates with $T_0 = 0$ and $T_{m+1} = T$ and μ_i $(i = 1, ..., m + 1)$ are the means within each regime which can easily be estimated once the break dates are. This method is efficient and can easily handle the large number of observations we are using, even with as many as 28 breaks (see Bai and Perron, 2003 for more details). Note that since our model allows for consecutive level shifts, we set the minimal length of a segment to just one observation. The fitted level shift component obtained using this method is presented in Figure 1 along with the original series $\log |r_t|$ and the smoothed estimate of the level shift component obtained using the smoothing method described in Wada and Perron (2007). As can be seen, the smoothed estimates are quite erratic, though they generally follows the overall changing mean of the series as depicted using the method of Bai and Perron (2003). This highlights the advantage of using this latter method since it delivers estimates of the level shift component that are consistent with the model postulated. A look at the graph indicates that the shifts agree with the major changes in the scale of $\log |r_t|$ and, as expected, there is a brief but important shift in October 1987 for all series.

The first issue we address is whether the level shift can account for the long-memory feature of the series $\log|r_t|$. To do so, we plot the autocorrelation functions (up to 300 lags) of the original series $\log|r_t|$ and its short-memory component c_t, obtained by subtracting the fitted trend from $\log|r_t|$. We do so using the trend obtained using the smoothed estimate of the level shift component and that obtained using the algorithm of Bai and Perron (2003). The results are presented in Figure 2. One can see that for all cases, the log-absolute returns clearly display an autocorrelation function that resembles that of a long-memory process: it decays very slowly and the values remain important even at lag 300. On the other hand, once the level shifts are accounted for, the picture is completely different and there is barely any evidence of remaining correlation, consistent with the estimate of ϕ obtained above. Given that we have ten thousands or so observations for each series, the autocorrelations at small lags may be deemed significant but they have no practical importance given their small values. Hence, for all practical purposes, we can view the short-memory component as being nearly white noise. These results are quite impressive. Despite the fact that the shifts are rare and account for only a small portion of the total variation of the series, they account for (nearly) all of the autocorrelation present in log-absolute returns.

It is almost universally accepted that stock returns exhibit conditional heteroskedasticity. For that reason, the GARCH(1,1) model, introduced by Bollerslev (1986) following the work of Engle (1982), has been used widely to model returns and is generally perceived as one of the leading candidates of forecasting models. Though Lamoureux and Lastrapes (1990) documented that structural changes in the level of variance can amplify the evidence of conditional heteroskedasticity, no strong evidence has been previously presented on assessing whether accounting for such changes can completely eliminate all evidence of conditional heteroskedasticity. We shall present evidence to this effect.

To account for the fat-tailed distribution of returns, the most popular model is the GARCH(1,1) model with Student-t errors given by, for the demeaned returns process $\widetilde{r}_t$,

$$\begin{aligned} \widetilde{r}_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \mu + \beta_r \widetilde{r}_{t-1}^2 + \beta_\sigma \sigma_{t-1}^2, \end{aligned} \tag{10}$$

where ε_t is i.i.d Student-t distributed with mean 0 and variance 1. The parameters of interest are β_r and β_σ, which measure the extent of conditional heteroskedasticity present in the data. As is standard, we let the degrees of freedom of the Student-t distribution be an unknown parameter, estimated jointly with the others. We estimate such a model for the four returns indices using maximum likelihood. We also consider the case that accounts for level shifts at the dates documented in the fitted level shift component, as presented of Figure 1. This implies the following extended specification, sometimes referred to as a components GARCH model:

$$
\begin{aligned}
\widetilde{r}_t &= \sigma_t \varepsilon_t, \\
n_t &= \mu + \rho\left(n_{t-1} - \mu\right) + \phi\left(\widetilde{r}_{t-1}^2 - \sigma_{t-1}^2\right) + \sum_{i=2}^{m+1} D_{i,t}\gamma_i, \\
\left(\sigma_t^2 - n_t\right) &= \beta_r\left(\widetilde{r}_{t-1}^2 - n_{t-1}\right) + \beta_\sigma\left(\sigma_{t-1}^2 - n_{t-1}\right),
\end{aligned} \tag{11}
$$

where $D_{i,t} = 1$ if t is in regime i, i.e., $t \in \{T_{i-1}+1, ..., T_i\}$, and 0 otherwise, with T_i $(i = 1, ..., m)$ being the break dates documented in Figure 1 (again $T_0 = 0$ and $T_{m+1} = T$). The coefficients γ_i, which index the magnitude of the shifts, are treated as unknown and are estimated with the remaining parameters, while the number of breaks is obtained from the point estimate of α. The estimates of β_r and β_σ obtained from the standard GARCH(1,1) model (10) and the extended model (11) that allows for level shifts are presented in Table 2. The results are quite informative. Using the standard GARCH(1,1) model, both estimates are highly significant for all series. In particular, the value of β_σ is quite high, as often documented, with values ranging between 0.866 and 0.935. We also estimated the components GARCH model without level shifts and the results are similar for the estimates of β_r and β_σ. The estimates of ρ are very close to one. This is due to the fact that the model attempts to capture features akin to long memory so that a value of ρ close to one implies highly persistent shocks.

The picture is very different when we allow for level shifts. None of the estimates of β_σ are significant, with small values ranging from -0.045 to 0.151. The estimate of β_r is insignificant for the S&P 500 and Dow Jones series. It is significant with p-values close to 0.01 for the AMEX and NASDAQ series, though the point estimates are much smaller (0.05 for both) so that their economic importance is minor. Also of interest is the fact that with level shifts the estimates of ρ are now well below one, between 0.85 and 0.95, implying that shocks have a half life between between 4 and 13 days. Hence, introducing few level shifts imply a markedly different interpretation of the data. These results indicate that level shifts in log-absolute or squared returns account for nearly all of the documented evidence of conditional heteroskedasticity in these stock returns series. They also imply that shock to volatility have very little persistence.

We also assessed the sensitivity of the results using the smoothed estimate of the trend function. This is done by replacing the term $\sum_{i=2}^{m+1} D_{i,t}\gamma_i$ by the smoothed estimate of the level shift component. The results are qualitatively similar, though it must be noted that, in this case, the estimates of the parameters are quite sensitive to minor changes in the sample used.

To summarize the results so far, the level shifts model with white noise errors appears to provide an accurate description of the data. The level shift component is an important feature that explains both the long-memory and conditional heteroskedasticity features generally perceived as stylized facts. Our analysis so far is retrospective and one may rightfully raise the issue of data overfitting as an explanation of the results, especially when multiple structural changes are present. It remains therefore to see whether our level shift model provides reasonable forecasts in comparison to traditional models.

5 Forecasting

We now consider the performance of the level shift model with white noise errors for log-absolute returns in forecasting volatility proxied by squared returns relative to two popular models: the GARCH(1,1) model and its fractionally integrated counterpart, the FI-GARCH(1,1) model. Given that the smoothed estimates of the level shift components are erratic and not in accord with the postulated model, we shall henceforth not use them. The design of the forecasting experiment follows Stărică and Granger (2005). We start forecasting at observations 2,000. We re-estimate the models every 20 days, at which point forecasts of up to 200 days are constructed. The proxy for volatility being the realized squared returns are quite noisy, and to reduce the effect of sampling variability, we follow Stărică and Granger (2005) in the construction of a metric to gauge relative performance. Let $\hat{\sigma}^2_{t,p}$ be a p-step ahead forecast of σ^2_{t+p}, the variance of returns r_t at time $t+p$, proxied by the squared demeaned returns. Let n be the number of forecasts produced, then the estimated MSE is constructed as

$$MSE(p) = \frac{1}{n}\sum_{t=1}^{n}\left(\bar{r}^2_{t,p} - \sigma^2_{t,p}\right)^2, \tag{12}$$

where $\sigma^2_{t,p} = \sum_{k=1}^{p}\hat{\sigma}^2_{t+k}$ and $\bar{r}^2_{t,p} = \sum_{k=1}^{p} r^2_{t+k}$ is the realized volatility over the interval $[t+1, t+p]$. This estimate of the MSE is preferable to the simpler version $\sum_{t=1}^{n}(r^2_{t+p} - \hat{\sigma}^2_{t+p})^2$, because the latter uses a poor measure of realized return volatility (see Anderson and Bollerslev, 1998). Through averaging, some of the idiosyncratic noise in the high-frequency component of squared returns is canceled out. Throughout, the relative forecasting performance of the two models is evaluated by the ratio of their MSEs as defined above.

5.1 Construction of the forecasts

We now describe how the forecasts are constructed for each method. We start with the level shift model applied to log-absolute returns for which we assume, following the evidence documented above, a short-memory component that is simply white noise. Since our model and estimates pertain to log-absolute returns, we need an appropriate transformation that yields forecasts of the variance of returns. Recall that our model specifies the following process for log absolute returns [2]:

$$log\left(|r_t| + C\right) = \tau_t + c_t,$$

where C is a positive constant used to bound returns away from zero. This yields the following model for returns:

$$|r_t| + C = h_t^{1/2}\widetilde{\varepsilon}_t,$$

where $h_t = e^{2\tau_t}E(e^{2c_t})$ and $\widetilde{\varepsilon}_t = e^{c_t}/[E\left(e^{2c_t}\right)]^{1/2}$ so that $E\widetilde{\varepsilon}_t^2 = 1$ and $\widetilde{\varepsilon}_t$ is independent of h_t. The sequence $\widetilde{\varepsilon}_t$ is also i.i.d., given that $c_t = e_t$. Hence, $2c_t \sim i.i.d.\ N(0, 4\sigma_e^2)$ and $E(e^{2c_t}) = e^{0+4\sigma_e^2/2} = e^{2\sigma_e^2}$. Since the level shifts are rare and there is considerable uncertainty about their timing and magnitudes, we ignore them when forecasting. We then have,

$$E_t\left(|r_{t+k}| + C\right)^2 = E_t h_{t+k} = \exp(2\tau_t + 2\sigma_e^2),$$

so that the k-period ahead forecast of the squared returns r_{t+k}^2 is

$$E_t r_{t+k}^2 = \exp(2\tau_t + 2\sigma_e^2) - 2CE_t\left|r_{t+k}\right| - C^2,$$

where

$$E_t\left|r_{t+k}\right| = E_t \exp\left(\tau_t + c_{t+k}\right) - C = \exp\left(\tau_t + 0.5\sigma_e^2\right) - C.$$

We now consider forecasting with the Student-t GARCH(1,1) model, which can be written as:

$$\widetilde{r}_t = r_t - \mu = \sigma_t\varepsilon_t, \tag{13}$$

$$\sigma_t^2 = \alpha_1 + \alpha_2\widetilde{r}_{t-1}^2 + \alpha_3\sigma_{t-1}^2, \tag{14}$$

where the innovation ε_t is i.i.d Student-t distributed with mean 0 and variance 1 and $\widetilde{r}_t$ are demaned returns. We start with the following transformation:

$$\widetilde{r}_t^2 = \alpha_1 + \left(\alpha_2 + \alpha_3\right)\widetilde{r}_{t-1}^2 + \omega_t - \alpha_3\omega_{t-1}, \tag{15}$$

[2] Note that the constant term present in (1) does not enter in the estimation procedure. The unconditional mean of $\log(|r_t| + C)$ can be captures by constructing τ_t appropriately.

where $\omega_t = \widetilde{r}_t^2 - \sigma_t^2$ is the forecast error associated with the forecast of $\widetilde{r}_t^2$ based on its lagged values (to see this, note that $E_{t-1}(\widetilde{r}_t^2) = E_{t-1}(\sigma_t^2 \varepsilon_t^2) = \sigma_t^2 var(\varepsilon_t) = \sigma_t^2$). Thus, ω_t is a white noise process that is fundamental to $\widetilde{r}_t^2$. The unconditional mean of $\widetilde{r}_t^2$ can then be computed easily from (15), assuming $\alpha_2 + \alpha_3 < 1$, by

$$E\left(\widetilde{r}_t^2\right) = \sigma^2 = \alpha_1 / \left(1 - \alpha_2 - \alpha_3\right).$$

The recursive form for the squared demeaned returns is

$$\widetilde{r}_t^2 - \sigma^2 = \left(\alpha_2 + \alpha_3\right)\left(\widetilde{r}_{t-1}^2 - \sigma^2\right) + \omega_t - \alpha_3 \omega_{t-1}$$

so that, for $k > 1$, the k-period ahead forecast of the squared demeaned returns is

$$E_t \widetilde{r}_{t+k}^2 = \sigma^2 + \left(\alpha_2 + \alpha_3\right)^{k-1}\left(E_t \widetilde{r}_{t+1}^2 - \sigma^2\right) = \sigma^2 + \left(\alpha_2 + \alpha_3\right)^{k-1}\left(\sigma_{t+1}^2 - \sigma^2\right),$$

where $\sigma_{t+1}^2 = \alpha_1 + \alpha_2 \widetilde{r}_t^2 + \alpha_3 \sigma_t^2$. To recover the forecasts for the squared returns, we make the following adjustment, using the fact that the time variation in the conditional mean of returns is quantitatively negligible:

$$E_t r_{t+k}^2 = E_t \widetilde{r}_{t+k}^2 + \left(E_t r_{t+k}\right)^2 \approx E_t \widetilde{r}_{t+k}^2 + \mu^2.$$

We finally consider the FIGARCH(1,1) model, given by

$$\begin{aligned} \widetilde{r}_t &= r_t - \mu = \sigma_t \varepsilon_t, \\ \left(1 - \alpha_3 L\right)\sigma_t^2 &= \alpha_1 + [1 - \alpha_3 L - \left(1 - \alpha_2 L\right)\left(1 - L\right)^d]\widetilde{r}_t^2, \end{aligned}$$

where the innovation ε_t is i.i.d standard normal and $\left(1 - L\right)^d = \sum_{j=0}^{\infty} \pi_j L^j$ with

$$\pi_j = \frac{\Gamma\left(j + d\right)}{\Gamma\left(j + 1\right)\Gamma\left(d\right)} = \prod\nolimits_{0 < k \leqslant j} \frac{k - 1 - d}{k}.$$

To facilitate forecasting, we transform the conditional variance equation into an infinite order ARCH representation:

$$\sigma_t^2 = \alpha_1 / \left(1 - \alpha_3\right) + \lambda\left(L\right)\widetilde{r}_t^2,$$

where $\lambda_1 = \alpha_2 - \alpha_3 + d$ and $\lambda_k = \alpha_3 \lambda_{k-1} + [(k - 1 - d)/k - \alpha_2]\pi_{k-1}$, for $k \geqslant 2$. Then, the recursive forecasts can be constructed from

$$E_t \widetilde{r}_{t+k}^2 = E_t \sigma_{t+k}^2 = \alpha_1 / \left(1 - \alpha_3\right) + \sum_{i=1}^{t+k-1} \lambda_i E_t \widetilde{r}_{t+k-i}^2,$$

where $E_t \widetilde{r}_{t+k-i}^2 = \widetilde{r}_{t+k-i}^2$ for $i \geqslant k$. To recover the forecasts of the squared returns, we adjust $E_t \widetilde{r}_{t+k}^2$ by adding back the squared mean, so that $E_t r_{t+k}^2 \approx E_t \widetilde{r}_{t+k}^2 + \mu^2$.

5.2 Forecasting Comparisons

We start with an unfair but instructive comparison. For the GARCH(1,1) and FIGARCH(1,1) models, we re-estimate every 20 observations. However, for the random level shifts model, we use the fitted means obtained from the full sample, i.e., the fitted level shift process depicted in Figure 1 obtained using the Bai and Perron (2003) algorithm and the number of breaks implied by the full sample estimate of α, the probability of a level shift at each period. The results are presented in Figure 3 (level shifts versus GARCH) and Figure 4 (level shifts versus FIGARCH). They show that the random level shift model has much better forecasting performance at all horizons and for all series, except for very short horizons when considering the AMEX and NASDAQ indices, in which case the GARCH and FIGARCH have a slight advantage. As stated, this is an unfair comparison but it indicates that if we have a precise estimate of the mean of log-absolute returns at a given date, we can obtain much better forecasts from the level shift model than the other models.

The issue then becomes how to obtain a good estimate of the current mean of a regime at a given date, at which the forecasts are made, without using information after that date. This turns out to be a delicate issue. The use of the filtered estimates of τ_t, the level shift component, obtained via the Kalman filter algorithm are too volatile to be useful. An obvious approach then is to repeat what we did for the full sample every 20 observations and forecast using the estimate of the mean for the last regime. The problem here is that if one allows for the possibility of a change at each date, as in the theoretical model, the fitted values often indicate that a change occurred at the end of the sample when, ex-post, no such long lasting change have occurred. In light of these issues, we resort to using a backwards CUSUM procedure, as in Pesaran and Timmerman (1999). At each forecasting period, we use the CUSUM test of Brown, Durbin and Evans (1975). We determine the cutoff point to get the mean to forecast as the first time the CUSUM statistic crosses one of the critical lines, determined by the criterion that the probability of at least one of the last 1,000 cumulative sums of standardized recursive residuals crossing a line is 10%. Note that this procedure is consistent with the model postulated. Indeed, with the assumption that the level shifts are independent and identically distributed (and, hence, cannot be forecasted), the best forecast is the mean obtained using the data from the last regime and some portion of the data prior. The CUSUM is a procedure that effectively indicates the date at which a forecast failure occurs and is, accordingly, the best suited from a forecasting perspective (see, e.g., Pesaran and Timmerman, 1999).

The results are presented in Figure 5 (level shifts versus GARCH) and Figure 6 (level shifts versus FIGARCH). For the S&P 500 and Dow Jones indices, the level shift model forecasts better than either the GARCH(1,1) or the FIGARCH(1,1) model at nearly all horizons (except at a few isolated short horizons). For the AMEX and NASDAQ indices, the GARCH(1,1) model performs better only at very short horizons, while the random level shift model is better for longer forecasts. Compared to the FIGARCH(1,1) model, the random level shift model performs better in the case of the AMEX index, except at a few isolated short horizons. For the NASDAQ index, the level shift model performs better at horizons greater than 100 days, but the FIGARCH model performs better at short horizons.

Following Anderson et al. (2003) and Stărică and Granger (2005), in the tradition of Mincer and Zarnowitz (1969), we also evaluate the models' relative forecasting accuracy by regressing proxies for volatility on a constant and the forecast values provided by the various models. We consider regressions of the form

$$r_{t+p}^2 - r_t^2 = b_0 + b_1(f_{t,p}^{LS} - r_t^2) + b_2(f_{t,p}^i - r_t^2) + u_t, \tag{16}$$

where $f_{t,p}^{LS}$ denotes the p-step ahead forecast of r_{t+p}^2 from the level shift model and $f_{t,p}^i$ denotes the p-step ahead forecast from the GARCH model ($i = GA$) or the FIGARCH model ($i = FI$). See the discussion in Stărică and Granger (2005) for the choice of this particular regression. The goal is to see first if the forecasts from the level shift model are uncorrelated with the forecast errors from either of the two competitors. This is done by testing the null hypothesis $H_0^A : (b_0, b_1, b_2) = (0, 0, 1)$ using a standard Wald test with an asymptotic chi-square distribution. The results favor the level shift model if the null hypothesis can be rejected. Conversely, one can test if the forecast errors from the level shift model are uncorrelated with the forecasts from either of the two competitors. This is done by testing the null hypothesis $H_0^B : (b_0, b_1, b_2) = (0, 1, 0)$ using again a standard Wald test. The results favor the level shift model if the null hypothesis cannot be rejected. An outcome for which the p-value for the test pertaining to H_0^A is less (resp., greater) than 5% and that for the test pertaining to H_0^B is greater (respc., less) than 5% is viewed as evidence in favor (resp., against) the level shift model. The cases for which both p-values are either greater or smaller than 5% are viewed as providing no evidence in favor of either model.

The results are presented in Table 3 for various horizons between 20 and 200 days. Consider first the comparison between the level shift and GARCH models. For the S&P 500 and AMEX indices, the evidence is in favor of the level shift model at all horizons. For the Dow Jones index, the evidence is also in favor of the level shift model for horizons up to 100 days while for longer horizons the tests are inconclusive (both tests have p-values less than 5%). For the NASDAQ index, the level shift model is deemed superior for horizons 20, 40, 60, 80, 140 and 160 while the GARCH model performs better at horizons 100, 120 and 180 (for a horizon of 200 days, the evidence is inconclusive). Hence, the overall results strongly indicate that the forecasting performance of the level shift model is superior to that of the GARCH(1,1) model. When compared to the FIGARCH model, the level shift model generally performs better at shorter horizons: up to 100 days for the S&P 500 and Dow Jones indices, up to 60 days for the NASDAQ index and at horizon 20 for the AMEX index. The FIGARCH occasionally performs better: at horizon 200 for the S&P 500 and AMEX indices and at horizons 80, 100, 120 and 180 for the NASDAQ index. For all other cases, the evidence does not favor one model over the other, using the criterion adopted.

Overall, the results are encouraging and show that the random level shift model is a serious contender as a forecasting model and that gains in forecast accuracy can be substantial. These results are important since not only does the level shift model provide a better description of the data in sample, it does so without sacrificing forecasting performance, often increasing it. Hence, level shifts appear to be genuinely present in the data and not simply a modeling convenience that allows for a better in-sample fit. This contrasts with the common perception that structural change models are not useful for forecasting.

6 Conclusion

In this paper, we first proposed a simple estimation method for a model of a series composed of a random level shift and a short-memory component. The model, applied to log-absolute returns of the S&P 500, AMEX, Dow Jones and NASDAQ indices, yields impressive results. The level shift model with white noise errors appears to provide an accurate description of the data. The level shift component is an important feature that explains the presence of both the long-memory and conditional heteroskedasticity features that are generally perceived as stylized facts. When accounting for level shifts, the evidence in favor of long-memory and conditional heteroskedasticity disappears. The model can also provide important improvements in forecasting volatility when using squared returns as a proxy. These are especially noticeable when the mean of the regime in effect at the time of forecasting is well estimated. Precise estimates of this mean are, however, difficult to obtain in real time. We have nevertheless shown that in many cases, forecasting improvements can still be obtained from our model by using a backward recursive CUSUM test to determine the length of the last regime.

References

Andersen, T.G., and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review* 39, 885-905.

Anderson, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association* 96, 42-55.

Anderson, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-626.

Bai, J., and P. Perron (2003), "Computation and Analysis of Multiple Structural Change Models," *Journal of Applied Econometrics* 18*, 1-22*.

Baillie, R.T., T. Bollerslev, and H.O. Mikkelsen (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 73, 3–30.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31, 307-327.

Bollerslev, T., and H.O. Mikkelsen (1996), "Modeling and Pricing Long Memory in Stock Market Volatility," *Journal of Econometrics* 73, 151-184.

Breidt, J.F., N. Crato, and P.J.F. de Lima (1998), "On the Detection and Estimation of Long-memory in Stochastic Volatility," *Journal of Econometrics* 83, 325–348.

Brown, R.L., J. Durbin, and J.M. Evans (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society* B 37, 149-163.

Diebold, F., and A. Inoue (2001), "Long Memory and Regime Switching," *Journal of Econometrics* 105, 131-159.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation," *Econometrica* 50, 987-1008.

Engle, R.F. (1995), *ARCH: Selected Readings*. Oxford University Press.

Engle, R.F., and A.D. Smith (1999), "Stochastic Permanent Breaks," *Review of Economics and Statistics* 81, 533-574.

Filardo, A. J., and S. F. Gordon (1998), "Business Cycle Durations," *Journal of Econometrics* 85, 99-123.

Fuller, W. A. (1996): *Introduction to Time Series* (2nd ed.), New York: John Wiley.

Gourieroux, C., and J. Jasiak (2001), "Memory and Infrequent Breaks," *Economics Letters* 70, 29-41.

Granger, C.W.J., and Z. Ding (1996), "Varieties of Long Memory Models," *Journal of Econometrics* 73, 61-77.

Granger, C.W.J., and N. Hyung (2004), "Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns," *Journal of Empirical Finance* 11, 399-421.

Gray, S. (1996), "Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process," *Journal of Financial Economics* 42, 27-62.

Hamilton, J. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business-Cycle," *Econometrica* 57, 357-384.

Harrison, P. J., and C. F. Stevens (1976), "Bayesian Forecasting," *Journal of the Royal Statistical Society,* Series B 38, 205-247.

Harvey, A.C. (1998), "Long Memory in Stochastic Volatility," in *Forecasting Volatility in Financial Markets,* ed. by J. Knight and S. Satchell. Oxford: Butterworth-Heineman, 307-320.

Lamoureux, C.G., and W.D. Lastrapes (1990), "Persistence in Variance, Structural Change, and the GARCH Model," Journal of Business & Economic Statistics 8, 225-234.

Lobato, I.N., and N.E. Savin (1998), "Real and Spurious Long-memory Properties of Stock-market Data," *Journal of Business and Economics Statistics* 16, 261-268.

Mikosch, T., and C. Stărică (2004), "Changes of Structure in Financial Time Series and the GARCH Model," REVSTAT - Statistical Journal, 2, 42-73..

Mincer, J. and V. Zarnowitz (1969): "The Evaluation of Economic Forecasts," in Economic Forecasts and Expectations, ed. by J. Mincer. New York: National Bureau of Economic Research.

Morana, C., and A. Beltratti (2004), "Structural Change and Long-Range Dependence in Volatility of Exchange Rates: Either, Neither, or Both," *Journal of Empirical Finance* 11, 629-658.

Parke, W.R. (1999), "What is fractional integration?" *Review of Economics and Statistics* 81, 632-638.

Perron, P. (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica* 57, 1361-1401.

Perron, P. (1990), "Testing for a Unit Root in a Time Series Regression with a Changing Mean," *Journal of Business and Economic Statistics* 8, 153-162.

Perron, P. and Z. Qu (2007), "An Analytical Evaluation of the Log-periodogram Estimate in the Presence of Level Shifts," Unpublished Manuscript, Department of Economics, Boston University.

Perron, P., and Z. Qu (2008), "Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices", Unpublished Manuscript, Department of Economics, Boston University.

Perron, P., and T. Wada (2009), "Let's Take a Break: Trends and Cycles in U.S. Real GDP," forthcoming in *Journal of Monetary Economics.*

Pesaran, M.H., and A. Timmerman (1999), " Model Instability and Choice of Observation Window," Discussion Paper 99-19, Department of Economics, University of California, San Diego.

Qu, Z., and P. Perron (2008), “A Stochastic Volatility Model with Random Level Shifts: Theory and Applications to S&P 500 and NASDAQ Indices,” Unpublished Manuscript, Department of Economics, Boston University.

Shephard, N. (2005): *Stochastic Volatility: Selected Readings*, Edited volume, Oxford University Press.

Stărică, C., and C. Granger (2005), “Nonstationarities in Stock Returns,” *The Review of Economics and Statistics* 87, 503-522.

Teverovsky, V., and Taqqu, M. (1997), “Testing for Long-range Dependence in the Presence of Shifting Means or a Slowly Declining Trend, Using a Variance-type Estimator,” *Journal of Time Series Analysis* 18, 279-304.

Wada, T., and P. Perron (2007), “An Alternative Trend-Cycle Decomposition using a State Space Model with Mixtures of Normals: Specifications and Applications to International Data,” mimeo, Boston University.

Table 1: Maximum Likelihood Estimates.

	σ_η	α	σ_e	ϕ
S&P 500	0.75123	0.00145	0.73995	
(SD: 0.8021)	0.74290	0.00150	0.73950	-0.01040
AMEX	0.98696	0.00246	0.70346	
(SD: 0.7773)	1.08830	0.00167	0.70607	0.06321
Dow Jones	0.95720	0.00100	0.73947	
(SD: 0.7888)	0.95860	0.00103	0.73909	-0.00984
NASDAQ	1.45396	0.00077	0.74255	
(SD: 0.8528)	1.45589	0.00076	0.74258	0.00070

Table 2.a: S&P 500 and Dow Jones; Parameter Estimates; GARCH and CGARCH models

	S&P500				
	Coefficient	Estimate	Std. Error	t-Statistic	p-value
No level shifts in GARCH	β_r	0.066	0.005	14.38	0
	β_σ	0.932	0.005	209.35	0
No level shifts in CGARCH	β_r	0.047	0.007	7.14	0
	β_σ	0.926	0.010	89.07	0
	ρ	0.9998	0.0003	3149.13	0
	φ	0.027	0.005	5.36	0
Level shift in CGARCH (using the Bai-Perron algorithm)	β_r	0.009	0.011	0.786	0.432
	β_σ	0.151	1.186	0.127	0.899
	ρ	0.942	0.007	141.92	0
	φ	0.062	0.006	9.65	0
Level shifts in CGARCH (using the smoothed estimate)	β_r	-0.090	0.067	-1.33	0.183
	β_σ	0.494	0.288	1.72	0.088
	ρ	0.622	0.006	109.42	0
	φ	0.149	0.069	2.17	0.030
	Dow Jones				
	Coefficient	Estimate	Std. Error	t-Statistic	p-value
No level shifts in GARCH	β_r	0.059	0.004	14	0
	β_σ	0.935	0.004	212.39	0
No level shifts in CGARCH	β_r	0.041	0.007	5.889	0
	β_σ	0.931	0.010	91.23	0
	ρ	0.998	0.001	1044.74	0
	φ	0.023	0.006	3.709	0
Level shift in CGARCH (using the Bai-Perron algorithm)	β_r	0.015	0.010	1.48	0.139
	β_σ	0.034	0.629	0.053	0.958
	ρ	0.948	0.007	140.6	0
	φ	0.053	0.006	9.47	0
Level shifts in CGARCH (using the smoothed estimate)	β_r	-0.047	0.031	-1.51	0.129
	β_σ	0.460	0.332	1.38	0.167
	ρ	0.726	0.010	69.44	0
	φ	0.092	0.031	2.93	0.003

Table 2.b: AMEX and NASDAQ; Parameter Estimates; GARCH and CGARCH models

	AMEX				
	Coefficient	Estimate	Std. Error	t-Statistic	p-value
No level shifts in GARCH	β_r	0.127	0.007	17.15	0
	β_σ	0.867	0.007	124.77	0
No level shifts in CGARCH	β_r	0.123	0.010	12.01	0
	β_σ	0.797	0.018	43.49	0
	ρ	0.999	0.001	973.16	0
	φ	0.035	0.006	5.82	0
Level shift in CGARCH (using the Bai-Perron algorithm)	β_r	0.050	0.020	2.57	0.010
	β_σ	-0.045	0.291	-0.155	0.877
	ρ	0.854	0.018	48.39	0
	φ	0.114	0.015	7.54	0
Level shifts in CGARCH (using the smoothed estimate)	β_r	-0.000	0.021	-0.01	0.992
	β_σ	0.140	69.13	0.002	0.998
	ρ	0.719	0.005	158.6	0
	φ	0.147	0.019	7.53	0

	NASDAQ				
	Coefficient	Estimate	Std. Error	t-Statistic	p-value
No level shifts in GARCH	β_r	0.094	0.007	14.40	0
	β_σ	0.904	0.006	145.36	0
No level shifts in CGARCH	β_r	0.082	0.010	7.83	0
	β_σ	0.839	0.023	36.16	0
	ρ	0.9999	0.000	4964.46	0
	φ	0.039	0.005	7.39	0
Level shift in CGARCH (using the Bai-Perron algorithm)	β_r	0.049	0.017	2.91	0.004
	β_σ	0.068	0.304	0.222	0.824
	ρ	0.952	0.008	122.69	0
	φ	0.086	0.01	8.85	0
Level shifts in CGARCH (using the smoothed estimate)	β_r	-0.017	0.059	-0.29	0.771
	β_σ	0.787	0.643	1.22	0.221
	ρ	0.876	0.0004	2488.73	0
	φ	0.130	0.061	2.14	0.032

Table 3: Comparison of forecasting performance between the level shift, GARCH and FIGRACH models.

Horizon p (days)	p value of Wald statictics for				p value of Wald statictics for			
	H_0^A $f_{t,p}^{LS} \perp e_{t,p}^{GR}$	H_0^B $e_{t,p}^{LS} \perp f_{t,p}^{GR}$	H_0^A $f_{t,p}^{LS} \perp e_{t,p}^{FI}$	H_0^B $e_{t,p}^{LS} \perp f_{t,p}^{FI}$	H_0^A $f_{t,p}^{LS} \perp e_{t,p}^{GR}$	H_0^B $e_{t,p}^{LS} \perp f_{t,p}^{GR}$	H_0^A $f_{t,p}^{LS} \perp e_{t,p}^{FI}$	H_0^B $e_{t,p}^{LS} \perp f_{t,p}^{FI}$
	S&P 500				Dow Jones			
20	0.00	0.64	0.00	0.57	0.00	0.13	0.00	0.09
40	0.00	0.85	0.01	0.75	0.00	0.55	0.00	0.55
60	0.00	0.79	0.00	0.45	0.00	0.43	0.00	0.25
80	0.00	0.99	0.08	0.58	0.00	0.24	0.00	0.18
100	0.00	0.67	0.05	0.36	0.00	0.26	0.00	0.17
120	0.00	0.54	0.13	0.09	0.00	0.00	0.00	0.00
140	0.00	0.50	0.27	0.07	0.00	0.02	0.00	0.06
160	0.00	0.96	0.32	0.19	0.00	0.01	0.00	0.02
180	0.00	0.97	0.30	0.27	0.00	0.00	0.00	0.04
200	0.00	0.17	0.11	0.02	0.02	0.00	0.04	0.00
	AMEX				NASDAQ			
20	0.00	0.68	0.04	0.51	0.00	0.48	0.00	0.53
40	0.00	0.58	0.26	0.15	0.00	0.80	0.00	0.94
60	0.00	0.67	0.42	0.14	0.00	0.76	0.08	0.44
80	0.00	0.69	0.51	0.35	0.00	0.12	0.07	0.01
100	0.00	0.64	0.53	0.18	0.35	0.00	0.16	0.00
120	0.00	0.52	0.39	0.31	0.22	0.00	0.45	0.00
140	0.00	0.86	0.79	0.12	0.00	0.40	0.66	0.10
160	0.00	0.78	0.71	0.11	0.00	0.49	0.53	0.11
180	0.00	0.85	0.79	0.11	0.12	0.00	0.35	0.00
200	0.00	0.93	0.67	0.03	0.00	0.00	0.01	0.00

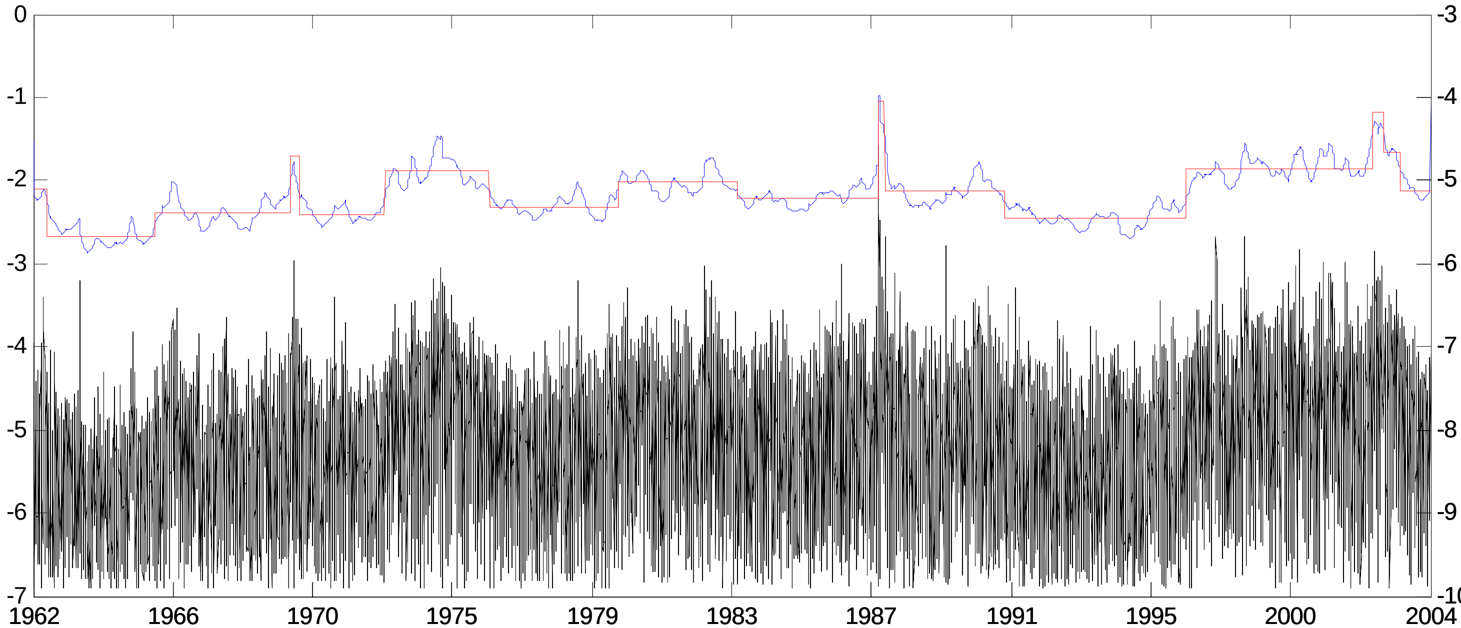

Figure 1.a: S&P 500; fitted level shift component (right axis) and series (left axis).

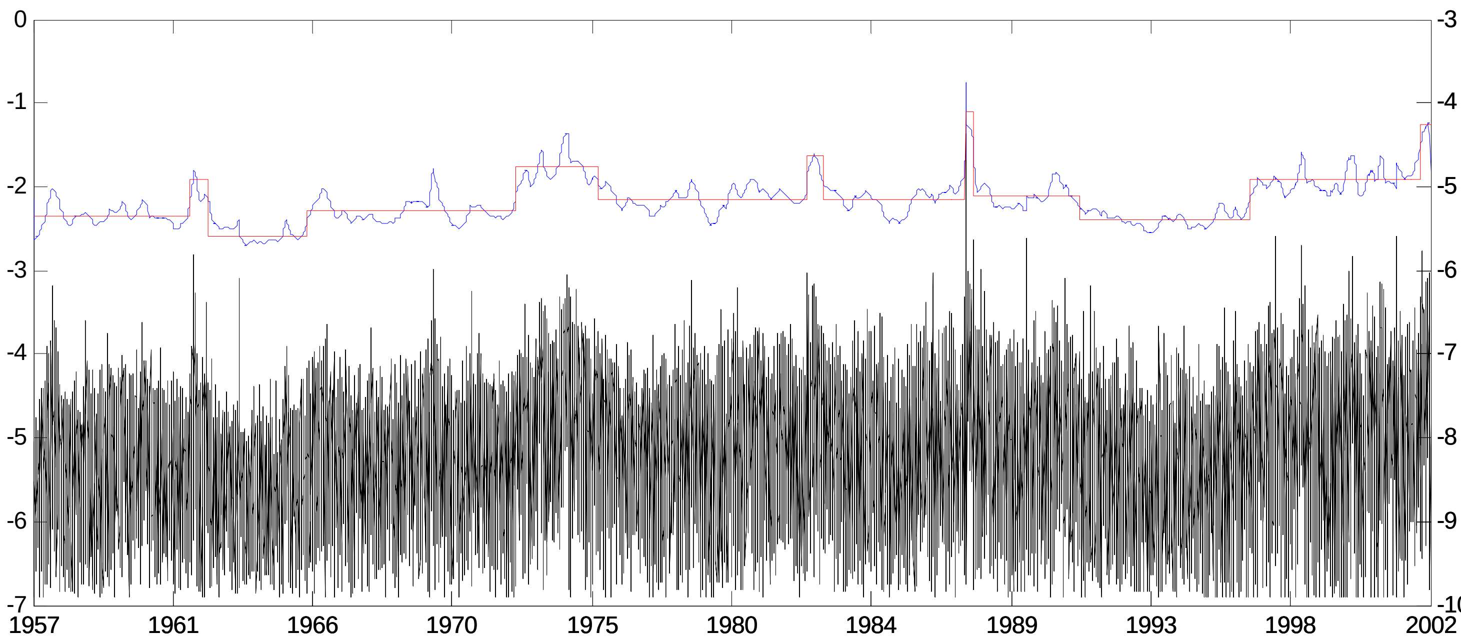

Figure 1.b: Dow Jones; fitted level shift component (right axis) and series (left axis).

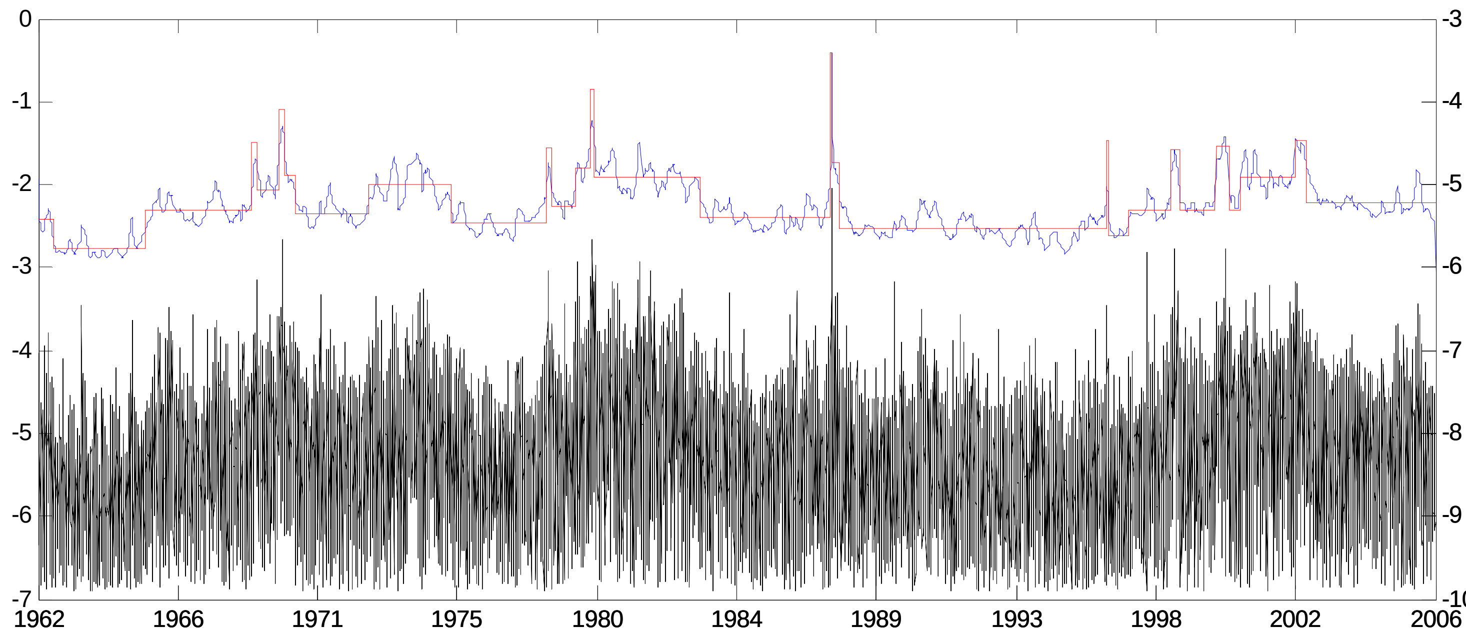

Figure 1.c: AMEX; fitted level shift component (right axis) and series (left axis).

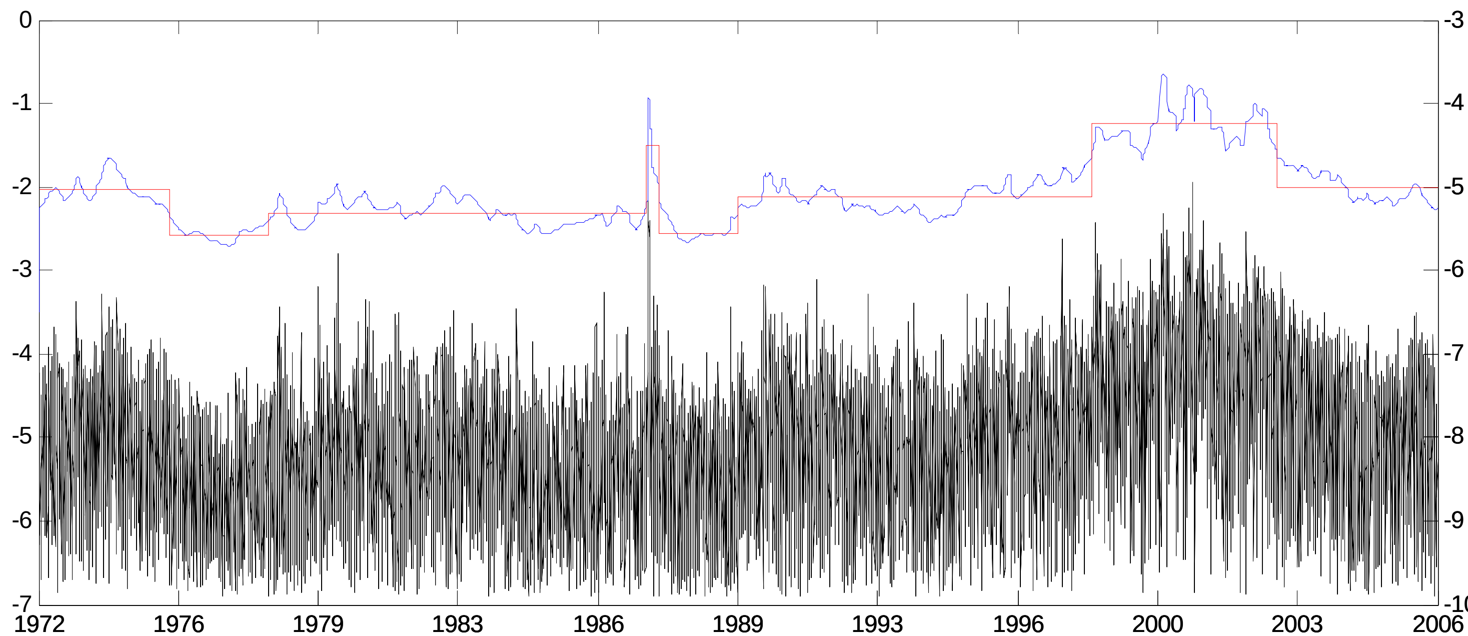

Figure 1.d: NASDAQ; fitted level shift component (right axis) and series (left axis).

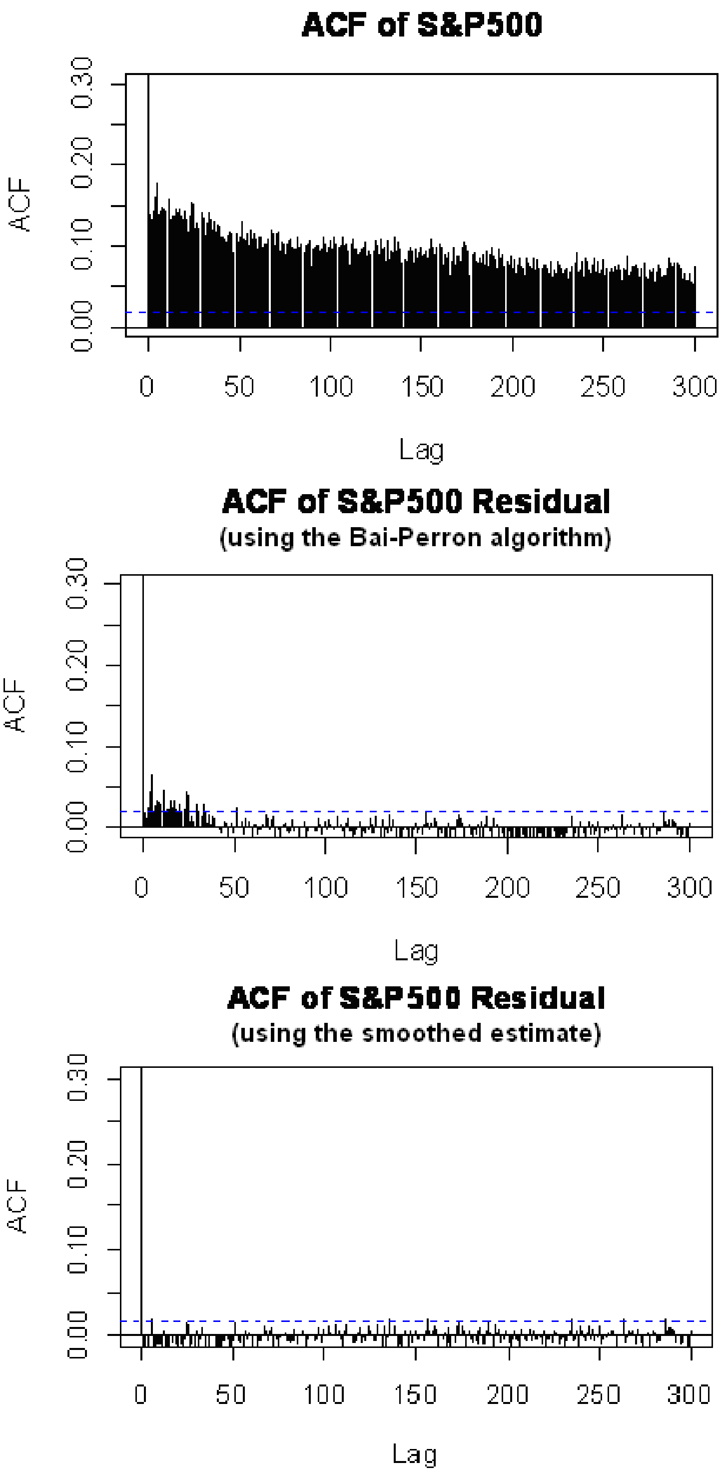

Figure 2.a: S&P 500; Autocorrelations.

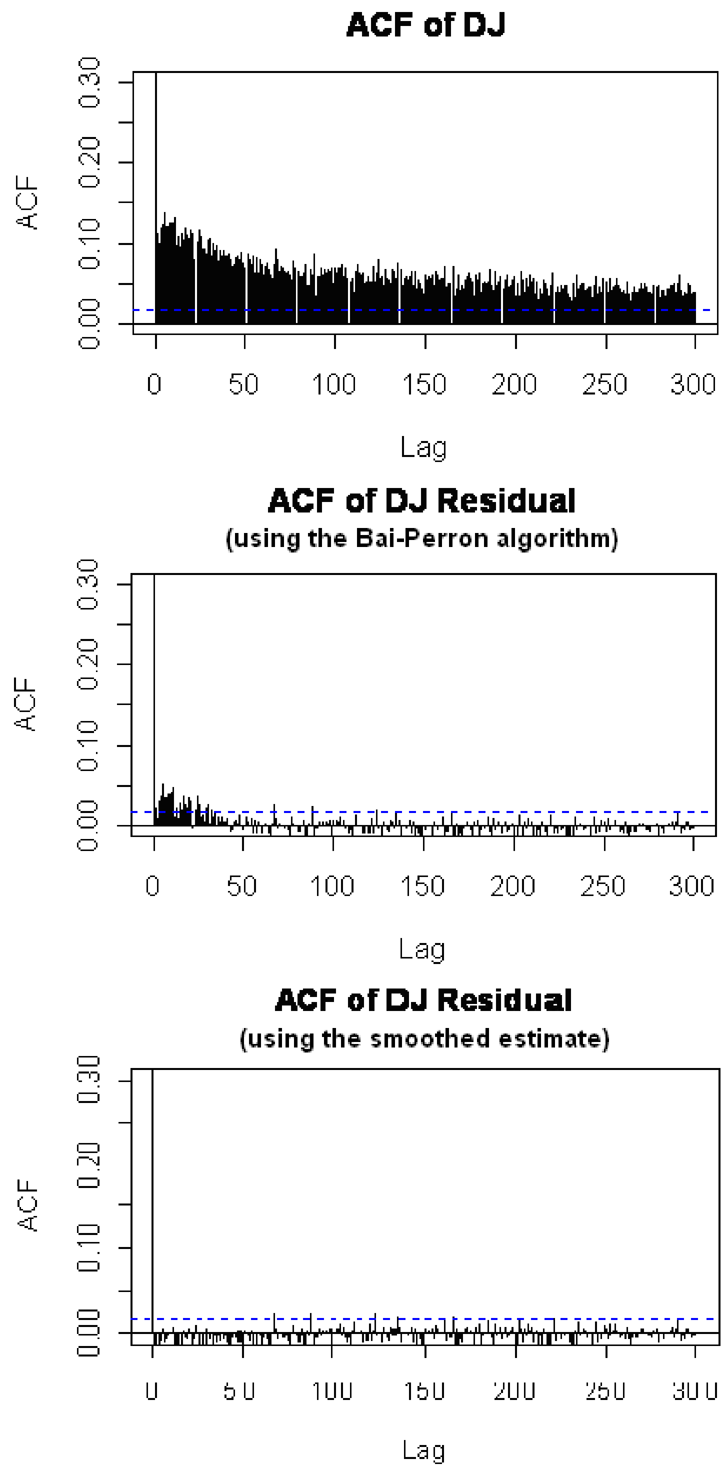

Figure 2.b: Dow Jones; Autocorrelations.

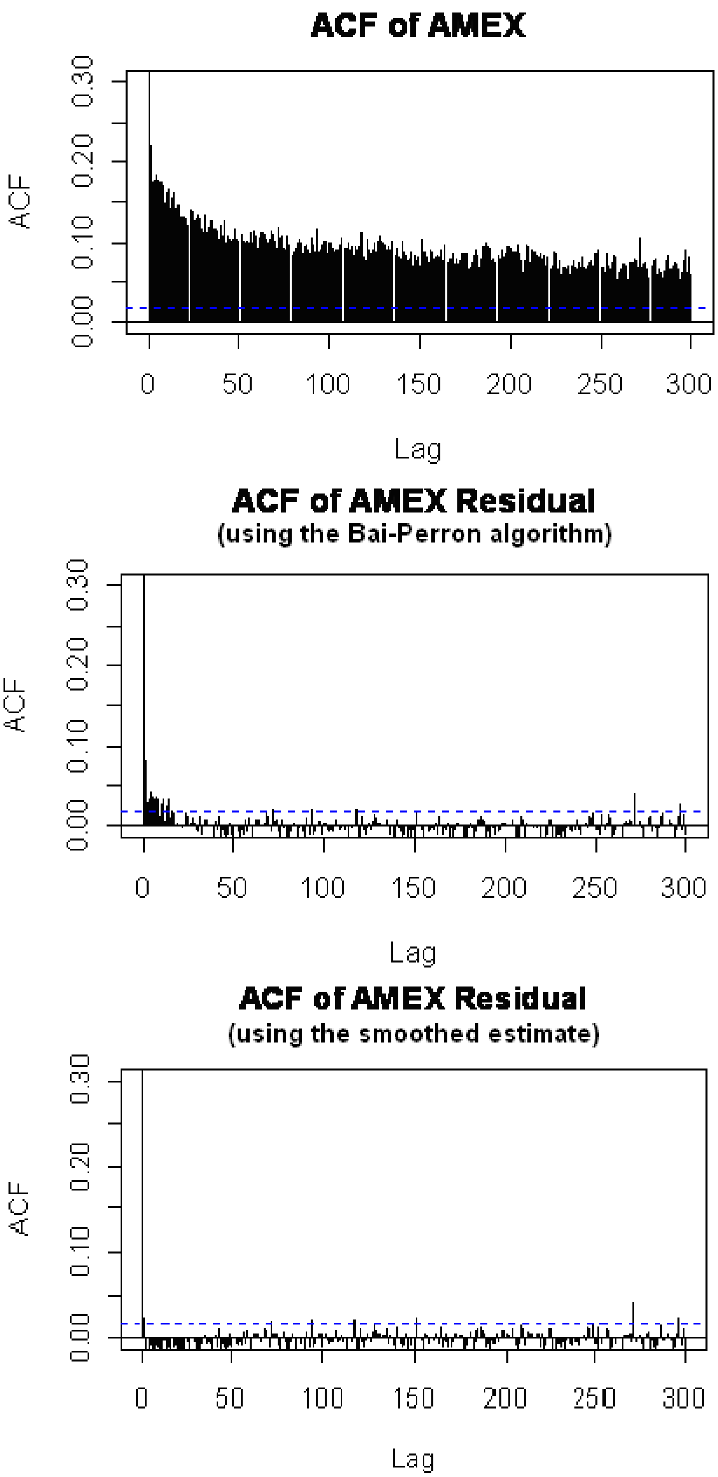

Figure 2.c: AMEX; Autocorrelations.

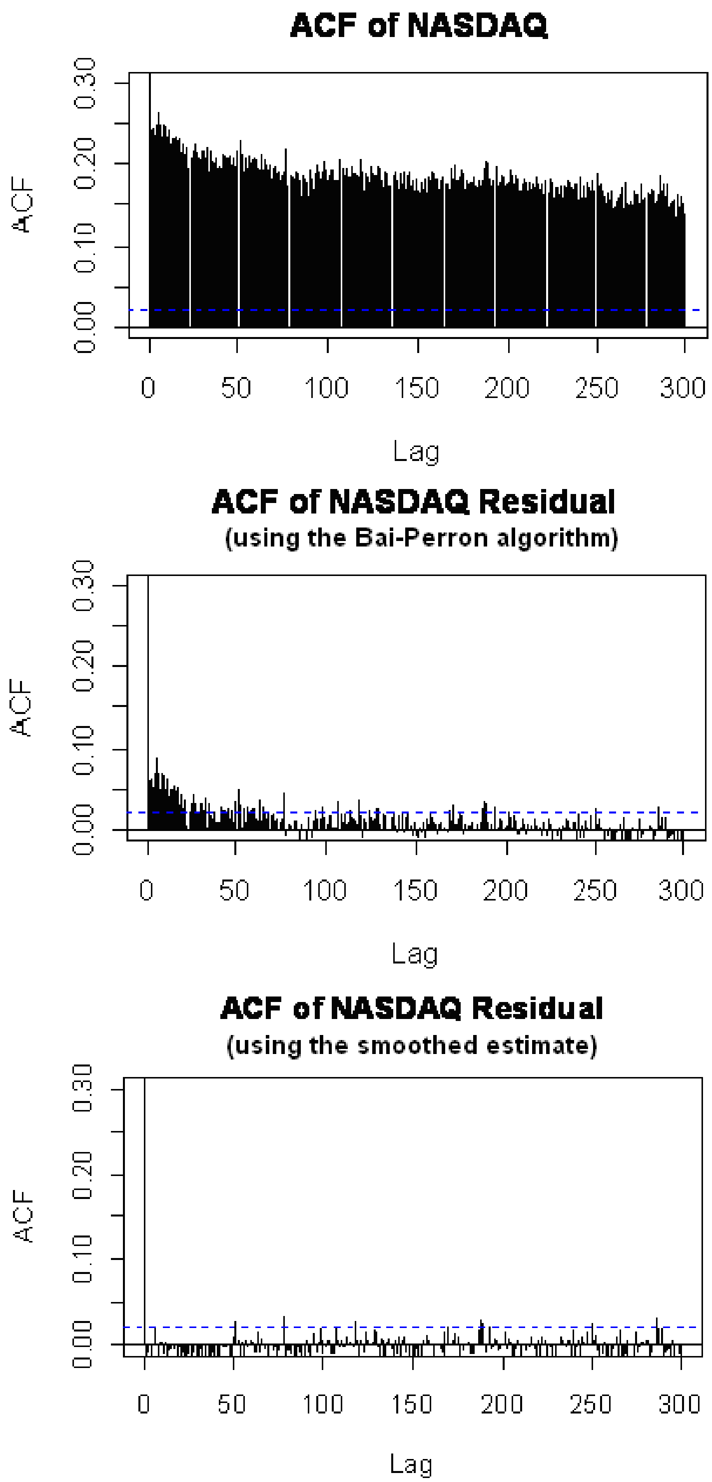

Figure 2.d: NASDAQ; Autocorrelations.

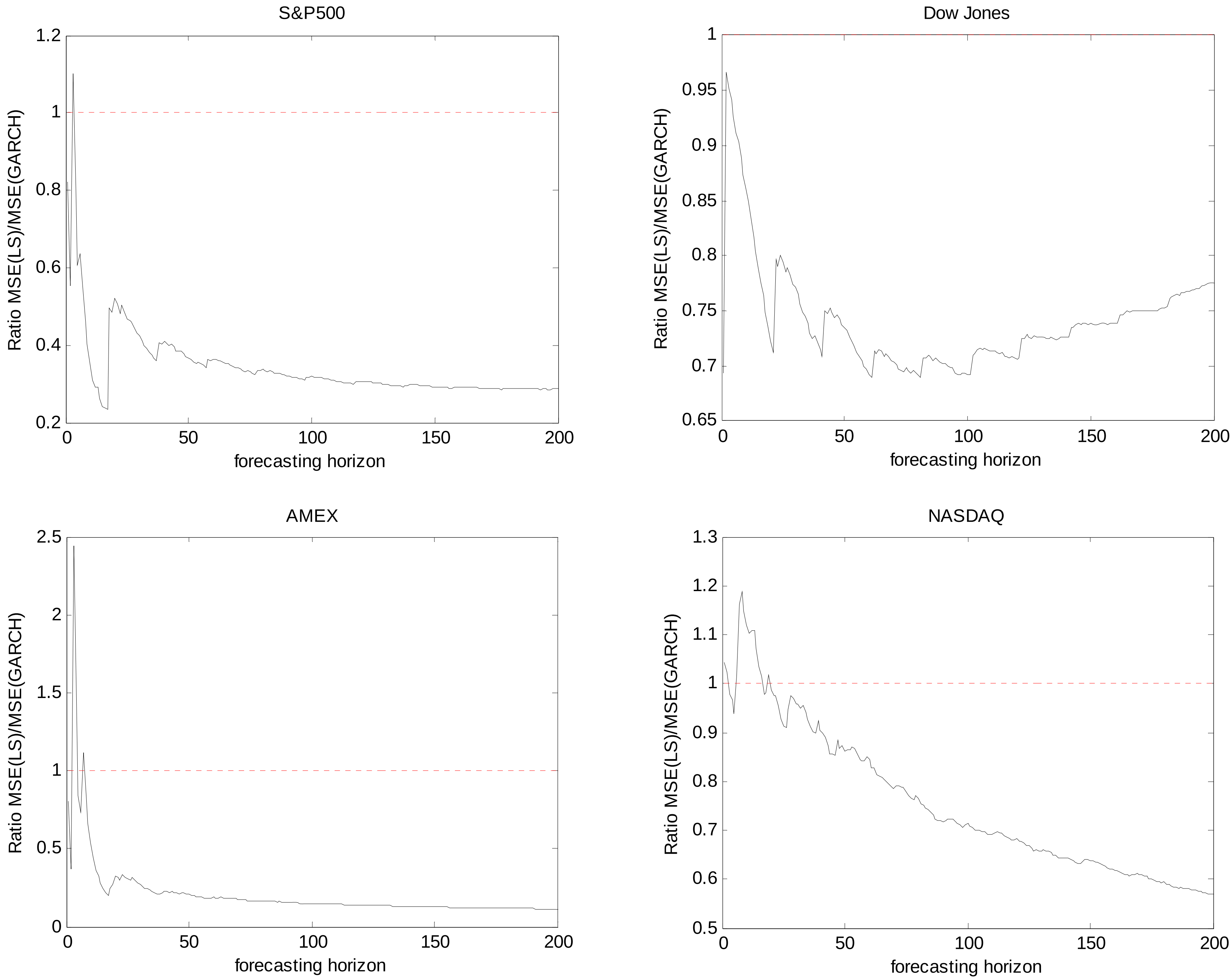

Figure 3: In-sample level shift versus out-of-sample GARCH forecasts

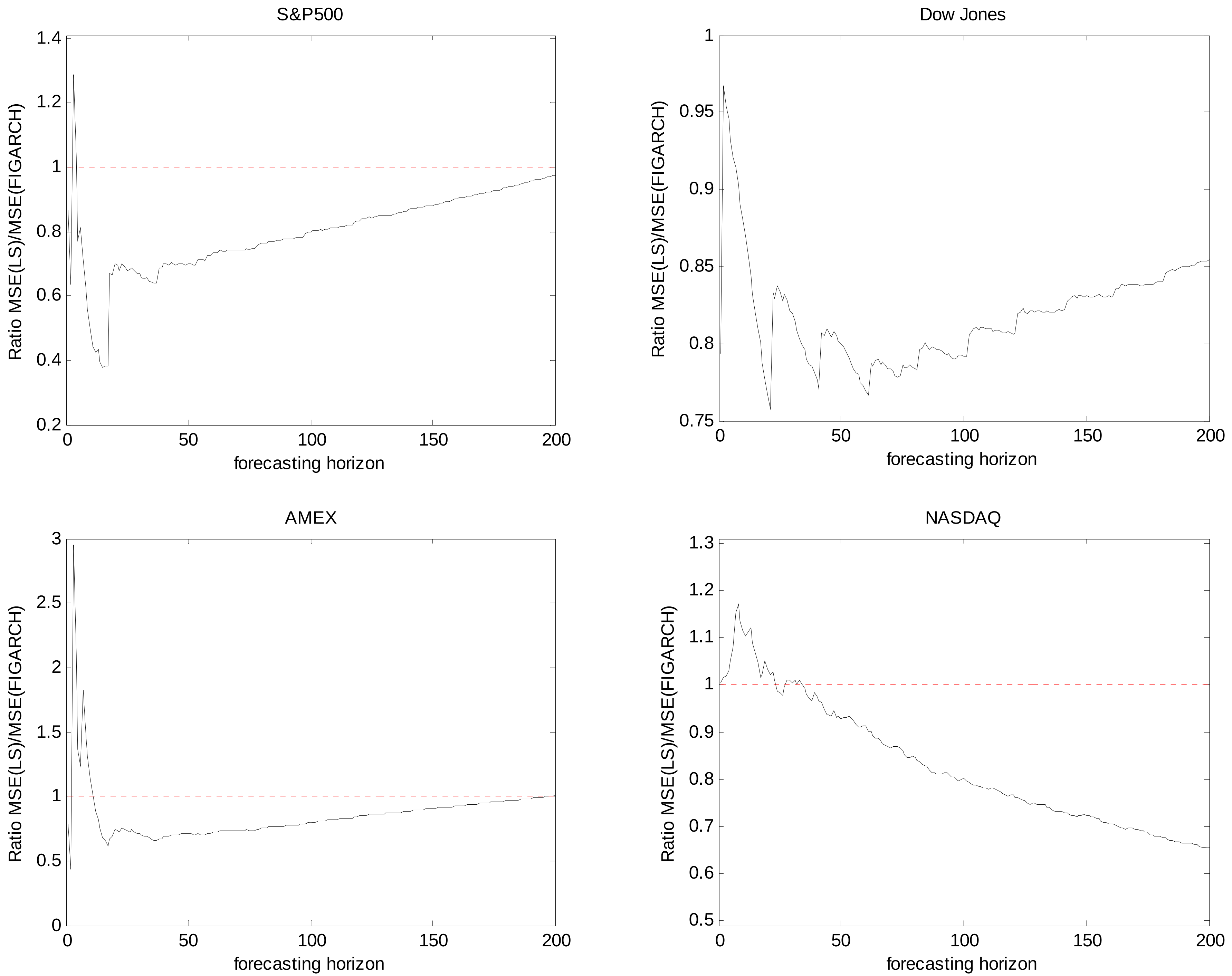

Figure 4: In-sample level shift versus out-of-sample FIGARCH forecasts

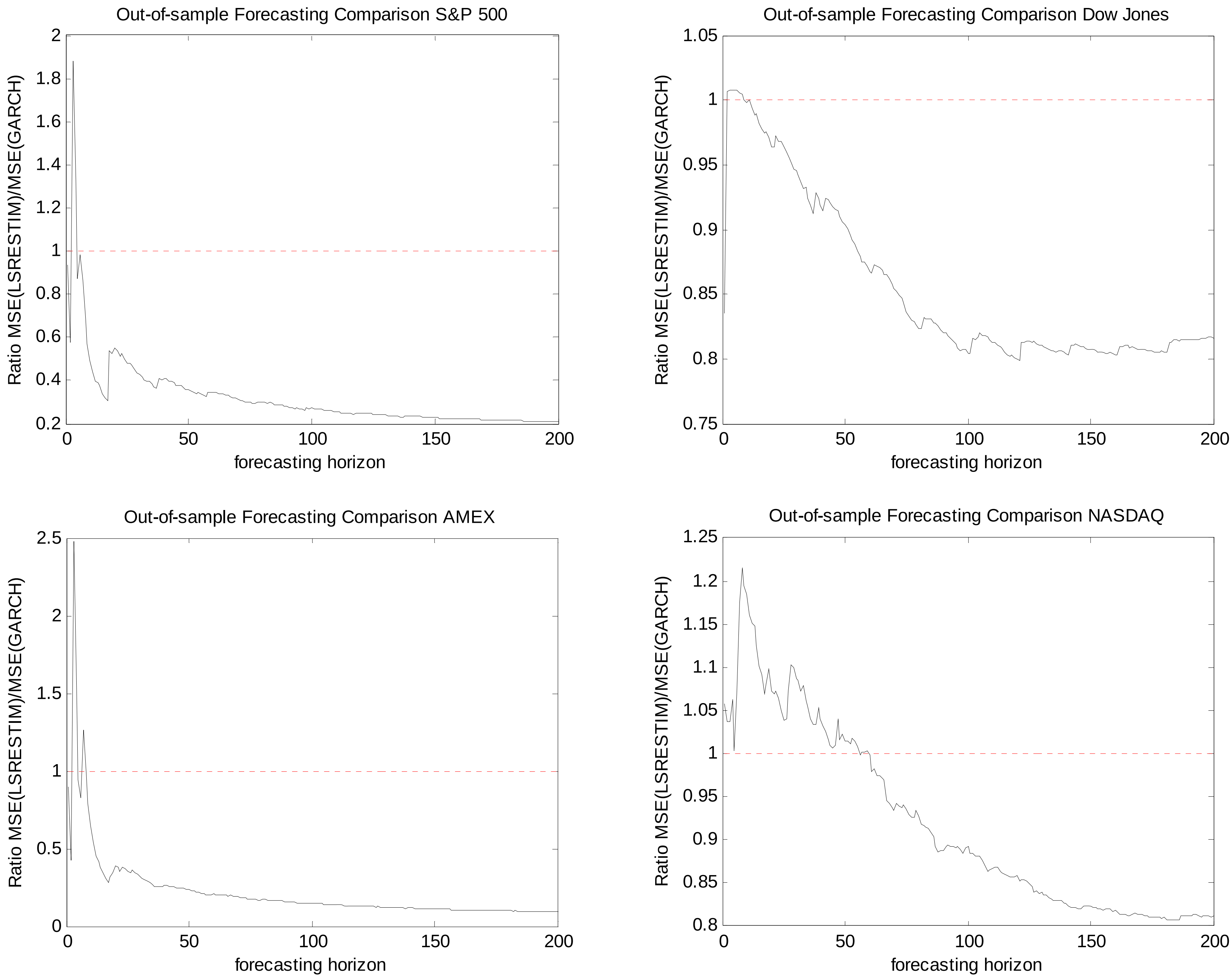

Figure 5: Out-of-sample level shift versus out-of-sample GARCH forecasts

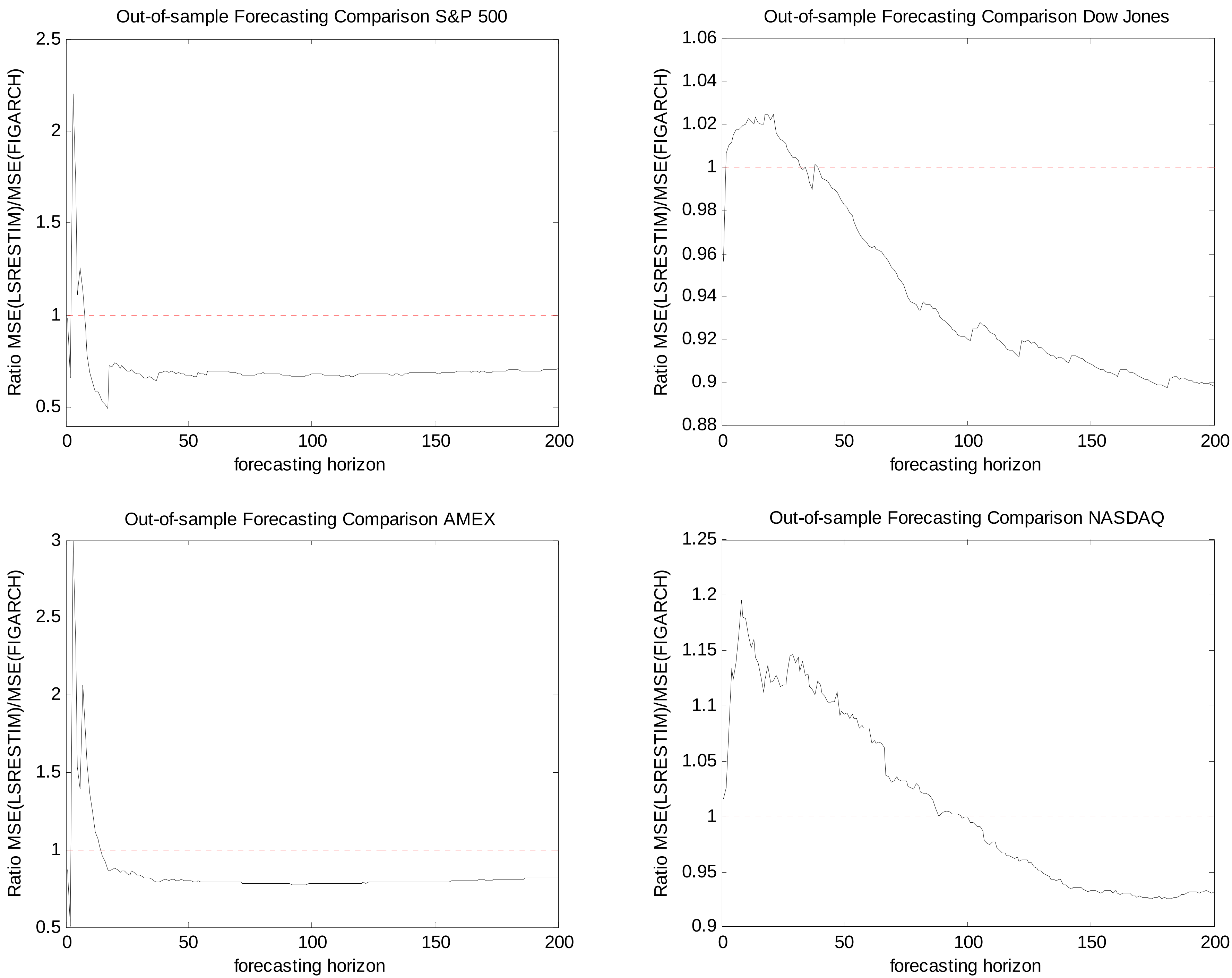

Figure 6: Out-of-sample level shift versus out-of-sample FIGARCH forecasts