A Stochastic Volatility Model with Random Level Shifts: Theory and Applications to S&P 500 and NASDAQ Return Indices*

Zhongjun Qu[†] **Boston University**

Pierre Perron[‡] **Boston University**

November 1, 2007; Revised June 10, 2008.

Abstract

Empirical findings related to the time series properties of stock returns volatility indicate autocorrelations that decay slowly at long lags. In light of this, several long-memory models have been proposed. However, the possibility of level shifts has been advanced as a possible explanation for the appearance of long-memory and there is growing evidence suggesting that it may be an important feature of stock returns volatility. Nevertheless, it remains a conjecture that a model incorporating random level shifts in variance can explain the data well and produce reasonable forecasts. We show that a very simple stochastic volatility model incorporating both a random level shift and a short-memory component indeed provides a better in-sample fit of the data and produces forecasts that are no worse, and sometimes better, than standard stationary short and long-memory models. We use a Bayesian method for inference and develop algorithms to obtain the posterior distributions of the parameters and the smoothed estimates of the two latent components. We apply the model to daily S&P 500 and NASDAQ returns over the period 1980.1-2005.12. Although the occurrence of a level shift is rare, about once every two years, the level shift component clearly contributes most to the total variation in the volatility process. The half-life of a typical shock from the short-memory component is very short, on average between 8 and 14 days. We also show that, unlike common stationary short or long-memory models, our model is able to replicate keys features of the data. For the NASDAQ series, it forecasts better than a standard stochastic volatility model, and for the S&P 500 index, it performs equally well.

*Pierre Perron acknowledges financial support from the National Science Foundation under Grant SES-0649350. We are grateful to Adam McCloskey for detailed comments on a previous draft. **JEL Classification**: C11, C12, C53, G12. **Keywords**: Bayesian estimation; Structural change; Forecasting; Long-memory; State-space models; Latent process.

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

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

1 Introduction

The literature on modeling and forecasting stock return volatility is voluminous. Two approaches that have proven useful are the GARCH and the 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 models where log-volatility is modeled as a fractionally integrated process. A related literature explores semiparametric estimation of the long-memory parameter. For example, see Arteche (2004) and Deo and Hurvich (2001).

Structural changes or level shifts in volatility has been advanced as a possible explanation to the long-memory phenomenon. It is now well known that when a stationary short-memory process is contaminated by level shifts, estimates of the memory parameter are biased away from zero and the autocovariance function exhibits a slow rate of decay, akin to a long-memory process (see Diebold and Inoue, 2001). The resulting phenomenon is often referred to as spurious long-memory. There is growing evidence suggesting that such structural changes may be important features of stock returns volatility. Using daily data on S&P 500 returns, Granger and Hyung (2004) documented the fact that when breaks (determined via some pre tests) are accounted for, the evidence for long-memory is weaker. Stărică and Granger (2005) presented evidence suggesting that this series can be well approximated by a sequence of identically and independently distributed shocks affected by occasional shifts in unconditional variance. They argued that such a specification exhibits better forecasting performance than long-memory or GARCH(1,1) models. In a recent study, Perron and Qu (2007) analyzed the time and spectral domain properties of a stationary short-memory process affected by random level shifts. They showed that the shape of the autocorrelation function is different from that obtained from a long-memory process when a wide range of lags is considered and that the limiting distribution of the log periodogram estimate of the memory parameter changes as the number of frequencies used varies. When applied to daily S&P 500 log absolute returns over the period 1928-2002, the level shift model explains both

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

the shape of the autocorrelations and the path of the log periodogram estimates as a function of the number of frequency ordinates used, see Perron and Qu (2008).

Despite these results, it remains a conjecture that a model incorporating random level shifts in variance can explain the data well and produce reasonable forecasts. The aim of this paper is to show that a very simple stochastic volatility model incorporating a random level shift and a short-memory component indeed provides a better in-sample fit of the data and produces forecasts that are no worse, and sometimes better, than standard stationary short and long-memory models.

We propose a stochastic volatility model with a stationary short-memory component and a level shift component modeled by a compound binomial process. More specifically, let x_t denote the mean corrected return process. The model analyzed is given by

$$x_t = \exp(h_t/2 + \mu_t/2)\varepsilon_t,$$

where h_t is a stationary AR(1) component, μ_t is a compound binomial process that models random level shifts and $\varepsilon_t \sim i.i.d.\ N(0,1)$. More specifically,

$$\mu_{t+1} = \mu_t + \delta_t \sigma_\eta \eta_t,$$

where $\eta_t \sim i.i.d.\ N(0,1)$ and δ_t is a sequence of independent Bernoulli random variables taking value 1 with unknown probability p. The first process h_t is common to standard SV models and is intended to capture short run dynamics. The second term is a departure from the standard models and captures persistent changes in level which, as we shall see, provides an explanation of the long-memory features present in the data. A salient feature of the model is that it allows shocks to affect the system in different ways. In particular, shocks to h_t die out quickly and shocks to μ_t have a long lasting effect, remaining in the system until the next level shift occurs. This feature will be useful for understanding the evolution of the volatility process, particularly the effect of rare but influential events.

We propose methods for inference and forecasting. For inference, two difficulties must be addressed. First, as common to all SV models, the processes that determine volatility are latent. This difficulty can be handled in several ways and we shall adopt a Bayesian approach following the work of Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998). Second, and more importantly, the new component μ_t is a compound binomial process making both the number and locations of shifts unknown. Since the value of μ_t at any given time depends on the whole path of shifts, this complicates inference. We address this difficulty by appropriate conditioning and data augmentation techniques. These will

allow us to develop algorithms to obtain the posterior distributions of all parameters and smoothed estimates of the processes h_t and μ_t. We also present algorithms for filtering, likelihood evaluation and forecasting. They use the technique of particle filtering and are relatively straightforward to implement.

We conduct simulations to assess the performance of our proposed method. To this end, we simulate short-memory processes with and without level shifts. We calibrate the parameters to empirical estimates to make the simulations practically relevant. Two issues are investigated: the convergence property of the algorithm and the ability of the model to distinguish between the two processes. The results suggest a convergence speed that is acceptable. Also, for the parameterizations considered, the model can detect level shifts when they occur and it does not tend to report spurious shifts.

The model is then applied to the volatility of daily S&P 500 and NASDAQ returns over the period 1980.1-2005.12. The choice of priors pertaining to the level shift component reflects the belief that shifts are rare and their magnitudes are large. The data are, however, informative enough to substantially alter the prior and make the probability of shifts even smaller. The priors pertaining to the stationary component are the same as those of Kim, Shephard and Chib (1998). For the S&P 500 series, we obtain the following results (they are qualitatively similar for the NASDAQ series). First, the occurrence of a level shift is rare. The posterior mean for the probability that a level shift occurs is 0.00187, which implies one shift every 535 days, on average. Importantly, although rare, the level shift component clearly contributes more to total variation in the volatility process than the short-memory component. Second, the model explains both the shape of the autocorrelation function and the path of the log periodogram estimates as a function of the number of frequency ordinates used, as documented by Perron and Qu (2007, 2008). Finally, the forecasting comparisons are encouraging. For the NASDAQ series, our model forecasts as well or better than a standard stochastic volatility model or a fractionally integrated GARCH model, and for the S&P 500 index, it performs equally well. This finding is in contrast to the common perception that structural change models are not useful for forecasting and suggests that important gains may be achieved by modeling changes appropriately.

Our paper contributes to the rapidly growing literature on modeling the strong persistence in the volatility of stock returns and other financial/macroeconomic variables. It fills a noticeable gap by providing a unified framework for modeling, inference and forecasting of volatility in the presence of structural change of random timing, magnitude and frequency. From a methodological perspective, the paper is related to the growing literature on Bayesian

change-point models, including, inter alia, Chib (1998), McCulloch and Tsay (1993), Pesaran, Pettenuzzo and Timmermann (2006) and Koop and Potter (2007). Our modeling of the shift process is closest to that of McCulloch and Tsay (1993). Nevertheless, there exists an important difference. Namely in this paper, structural changes occur through a latent component. This generates nontrivial difficulties and our analysis provides novel technical solutions to them.

The paper is organized as follows. Section 2 provides a brief review of issues related to the long-memory phenomenon in return volatility. Section 3 presents our stochastic volatility model with level shifts. Section 4 presents the Bayesian inference procedure. Section 5 discusses methods for filtering, likelihood evaluation and forecasting. Section 6 conducts simulations to assess the performance of our suggested algorithm when confronting series with and without level shifts. Section 7 presents the results from empirical applications related to the volatility of daily S&P 500 and NASDAQ returns. Section 8 offers brief concluding remarks and some technical derivations are contained in appendices.

2 Long memory in stock return volatility

Let z_t be a stationary time series with autocorrelation function $\gamma_z(\tau)$ at delay τ. A common definition of long-memory (see Beran, 1994), states that the autocorrelation function takes the form

$$\gamma_z(\tau) = g(\tau)\tau^{2d-1} \quad as \quad \tau \to \infty, \tag{1}$$

where $d > 0$ and $g(\tau)$ is a slowly varying function as $\tau \to \infty$ (i.e., for any real c, $g(c\tau)/g(\tau) \to 1$ as $\tau \to \infty$). For $0 < d < 1/2$, this implies that the autocorrelations decrease to zero at a slow hyperbolic rate which depends on the memory parameter d, in contrast to the fast geometric rate of decay that applies to a short-memory process.

Several papers have reported that transformations of stock returns, say r_t, of the form $|r_t|^\theta$ for some $\theta > 0$ have time series properties that resemble those of a stationary long-memory process (see e.g., Ding, Engle and Granger, 1993, Granger and Ding, 1995 and Lobato and Savin, 1998). Using daily data on S&P 500 daily returns for the period 1928-1991, the log periodogram estimates of d with the number of frequency ordinates used in the regression being $m = \sqrt{n}$ where n is the sample size, varies from a low of .427 for $\theta = 2$ to a high of .508 for $\theta = 0.2$. The often cited generating mechanisms for long-memory are temporal aggregation (Granger, 1980 and Anderson and Bollerslev, 1997) and shocks of different durations (Parke, 1999). Structural changes in level can also produce the long-

memory phenomenon. In particular, if a stationary short memory process is contaminated by random level shifts, the estimate of the memory parameter is biased away from zero and the autocorrelation function exhibits a slow rate of decay, akin to a long-memory process (see Diebold and Inoue, 2001). Mikosch and Stărică (2004) and Lamoureux and Lastrapes (1990) provided examples in the context of volatility persistence.

More closely related to our work are Perron and Qu (2007, 2008). They considered the following simple process involving random level shifts for some series z_t which can be thought as the natural logarithm of squared returns:

$$\begin{aligned} z_t &= c + u_t + v_t, \\ u_t &= u_{t-1} + \pi_t \eta_t, \end{aligned} \tag{2}$$

where c is a constant, $u_0 = 0$, v_t is a stationary short memory process, $\eta_t \sim i.i.d.\ N(0, \sigma_\eta^2)$ and π_t is a Bernoulli random variable that takes value 1 with probability p_n. They used $p_n = p/n$ to model rare shifts. They studied the statistical properties of this process in the time and spectral domains and obtained the following results. First, in the spectral domain, the level shift component affects the periodogram only up to $j = o(n^{1/2})$, where j indexes the frequency ordinates. Within this range, the rate of decrease of the periodogram is on average -2, implying $d = 1$. This is different from a long-memory process defined by (1) in which the effect is up to $j = o(n)$ and the rate of decrease is $-2d$. This result implies that if we use a local method to estimate the memory parameter of the level shift model (2), the estimate will depend on the number of frequencies used. It will tend to decrease as the number of frequencies used increases and, in particular, a sharp decrease will occur when m varies from roughly $n^{1/3}$ to $n^{1/2}$. The decrease after m reaches $n^{1/2}$ will be gradual as the effect of the short-memory component becomes more important. Second, in the time domain, the autocovariance function of z_t has a special structure. Its expected value at large lags is a function that depends only on the sample size. It is initially (almost) linearly decreasing, crossing the zero axis when $j/n \approx 0.293$, reaching a minimum at $j/n \approx 0.592$ and increasing back up to 0 when $j/n = 1$ (basically, a flattened version of the autocorrelation function of a random walk). This is again different from what is obtained with a stationary long-memory process of the form (1), where the autocorrelations decrease to zero at rate $2d-1$. The features of the level shift model were found to be empirically present when using transformations of S&P 500 daily returns covering the period 1928-2002 in Perron and Qu (2008). Here, we replicate their findings using the sample period 1980.1-2005.12 which is of interest in this paper. The results are shown in Figures 1 and 2 for the logarithm of

squared S&P 500 and NASDAQ returns, respectively. They show that the pattern of the log-periodogram estimates of d as a function of m and the sample autocorrelation functions are consistent with the level shift model (2). These results suggest that it is worthwhile to formally incorporate level shifts into volatility models and to develop methods for inference and forecasting. This paper addresses these issues.

3 The stochastic volatility model with level shifts

Let $\{x_t; t = 1, ..., n\}$ denote the demeaned return process. A standard stochastic volatility model (SV) is given by

$$\begin{aligned} x_t &= \exp(h_t/2)\varepsilon_t, \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\upsilon \upsilon_t, \end{aligned} \tag{3}$$

where h_t is the log volatility at time t assumed to follow an $AR(1)$ process with $|\phi| < 1$. The innovations ε_t and υ_t are sequences of independent standard normal random variables with $cov(\varepsilon_t, \upsilon_s) = 0$ for all t, s. Traditionally, SV models have been hard to estimate in comparison to ARCH type models due to the fact that h_t is a latent process and there is no tractable representation of the likelihood function. Classical treatments include quasi-maximum likelihood estimation based on the Kalman filter (Harvey, Ruiz and Shephard, 1994) and the method of moments (Duffie and Singleton, 1993). Bayesian treatments include the work of Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998), among others. The Bayesian approach is computationally demanding but it is fully efficient and allows for the analysis of models with complicated structures. In this paper, we adopt the Bayesian approach.

The stochastic volatility model with level shifts that we propose is given by

$$\begin{aligned} x_t &= \exp(h_t/2 + \mu_t/2)\varepsilon_t, \\ h_{t+1} &= \phi h_t + \sigma_\upsilon \upsilon_t, \\ \mu_{t+1} &= \mu_t + \delta_t \sigma_\eta \eta_t, \end{aligned} \tag{4}$$

with initial conditions

$$(h_0, \mu_0) = 0 \text{ and } (h_1, \mu_1)' \sim N(0, P), \tag{5}$$

where η_t, υ_t and ε_t are independent standard normal random variables and δ_t is a sequence of independent Bernoulli random variables taking value 1 with unknown probability p, i.e.,

$$\delta_t \sim B(1, p).$$

The variables $\eta_j, \delta_h, \varepsilon_k$ and v_l are mutually independent for all $1 \leq j, h, k, l \leq n$.

The presence of μ_t brings substantial flexibility and the model approximately nests the standard SV model when p is very small. If the standard model is correct, the posterior distribution of p will have large mass around zero and the model essentially reduces to (3). Otherwise, the posterior distribution of p is informative regarding the frequency of regime shifts. Such information, along with the posterior distribution of σ_η, can be used to assess the relative contributions of the level shift and stationary components to overall volatility. Note that the effect of innovations via h_t quickly die out while those occurring via μ_t remain in effect until the next shift. This is different from short or long-memory SV models where all shocks are assumed to affect the future path of the process in the same manner.

Remark 1 *For identification, we need to restrict* $0 < |\phi| < 1$. *This is straightforward to implement by restricting the support of the prior distribution to be strictly within the unit interval.*

Remark 2 *The model is different from that of Duffie, Pan and Singleton (2000), who incorporate jumps into the SV model to capture clustered extreme movements in returns. There, the jumps have a short lasting effect and their model accordingly generates a volatility process that is stationary while ours does not. As we shall see, this non-stationarity in the level shift component is crucial to generating features akin to a long-memory process.*

Remark 3 *Anderson and Bollerslev (1997) advocate a temporal aggregation explanation of the long-memory phenomenon along the lines of Granger (1980). The essential requirement for the presence of long memory is that a sufficient number of the individual information arrival processes have high persistence, i.e., mimicking a unit root process in finite samples. An important distinction is that we only use two components neither of which are restricted to be unit root processes. The crucial ingredient of our model is the explicit modeling of the probability of the occurrence of permanent shifts (i.e., we allow p to vary between 0 and 1, in contrast to a unit root process in which* $p = 1$*).*

4 The Bayesian inference procedure

The model (4) can be represented as

$$\begin{aligned} \log x_t^2 &= h_t + \mu_t + \log \varepsilon_t^2, \\ h_{t+1} &= \phi h_t + \sigma_\upsilon \upsilon_t, \\ \mu_{t+1} &= \mu_t + \delta_t \sigma_\eta \eta_t, \end{aligned} \tag{6}$$

with initial conditions $(h_0, \mu_0) = 0$ and $(h_1, \mu_1)' \sim N(0, P)$. The goal of the inference procedure is to obtain the posterior distributions of the parameters $(\sigma_\upsilon, \sigma_\eta, p, \phi)$ and the smoothed estimates of the latent processes h_t and μ_t given some priors. The Gibbs sampler and data augmentation are important ingredients that we use and thus we present a brief description of these tools in Appendix 1.

Under the assumption that ε_t is *i.i.d.* $N(0,1)$, $\log \varepsilon_t^2$ is the log of a Chi-square random variable. Hence (6) is a partial non-Gaussian state space model as analyzed by Shephard (1994), for which the optimal filtering is a nonlinear problem. To address this issue, we follow Shephard (1994), Carter and Kohn (1994) and Kim, Shephard and Chib (1998) and approximate the distribution of $\log \varepsilon_t^2$ by a mixture of normals. Then, conditional on a given realization of the mixture, the errors are normally distributed and the non-linearity due to $\log \varepsilon_t^2$ is no longer present. More specifically, define a new error process ε_t^* as

$$\varepsilon_t^* = \log \varepsilon_t^2 - E(\log \varepsilon_t^2).$$

We approximate its distribution by

$$\varepsilon_t^* \overset{d}{\sim} \sum_{i=1}^{K} q_i N(m_i, \sigma_i^2). \tag{7}$$

The choices of K, q_i, m_i and σ_i^2 follow Kim, Shephard and Chib (1998) and the details are presented in Appendix 2. Note that if ε_t follows a t distribution, the above procedure can still be applied with appropriate modifications for the weights q_i, the individual components (m_i, σ_i^2) and the number of components K.

The main difficulty for inference is that μ_t is a random level shift process with both the number and locations of shifts unknown. Also the current value of μ_t depends on the whole history of shifts. These features make inference more challenging than for a regime switching model in which the number of regimes is pre-specified and μ_t typically only depends on the current state or a finite number of past states. We address this issue using the technique of data augmentation and by appropriate conditioning. More specifically, the data is augmented by the innovation processes υ_t and η_t, the locations of shifts δ_t and the realization of the mixture in (7) at time t, denoted by ω_t. Conditional on a given realization of δ_t, the transition equation is then generated by a linear recursion with normal errors. Conditional on a given realization of the mixture (7), the measurement equation is also linear and Gaussian. Hence, the standard tools for Gaussian state space models can be applied. These allow us to sample efficiently from the posterior distribution of the parameters $(\phi, \sigma_\upsilon, \sigma_\eta)$, the innovation process

(υ_t, η_t) and the initial state (μ_1, h_1), conditional on δ_t and ω_t. Then, conditional on the initial state μ_1 and the path of η_t, the distribution of δ_t can be sampled recursively. Finally, ω_t and the other parameters can be sampled in the usual way.

Because of the logarithmic transformation, values of stock returns that are close to zero may have an undesirable effect on the inference procedure. To avoid this, we define

$$y_t = \log(x_t^2 + c) - E(\log \varepsilon_t^2),$$

where c is an "offset". That is, a small number is introduced to bound the term inside the logarithm away from zero, a technique introduced to the stochastic volatility literature by Fuller (1996). We use $c = 0.001$, though in practice it can be chosen based on the data. Then, model (6) can be expressed as

$$\begin{aligned} y_t &= h_t + \mu_t + \varepsilon_t^*, \\ h_{t+1} &= \phi h_t + \sigma_\upsilon \upsilon_t, \\ \mu_{t+1} &= \mu_t + \sigma_\eta \delta_t \eta_t. \end{aligned} \tag{8}$$

We now present the Bayesian procedure for inference. We start in Section 4.1 with the sampling algorithm to construct the posterior distributions. In Section 4.2, we discuss the priors to be used.

4.1 The sampling procedure for the posterior distributions

As a matter of notation, define $\alpha_t = (h_t, \mu_t)'$ for $t = 1, ..., n$; in particular, $\alpha_1 = (h_1, \mu_1)'$ stands for the initial state of the transition equation. Also define $R = \{(\upsilon_1, \eta_1)', ..., (\upsilon_n, \eta_n)'\}$, $\delta = (\delta_1, ..., \delta_n)$, $\omega = (\omega_1, ..., \omega_n)$, $\theta = (\phi, \sigma_\upsilon, \sigma_\eta, p)$ and let $\theta_{(-\phi)}$ denote the vector θ with ϕ excluded ($\theta_{(-\sigma_\upsilon)}, \theta_{(-p)}$, and $\theta_{(-\sigma_\eta)}$ are defined analogously). The sampling procedure used to obtain the posterior distributions consists of the following four blocks:

1. Sampling $\phi, \sigma_\upsilon, \sigma_\eta$, R and α_1 conditional on the location of the shifts and the mixture;
2. Sampling the process of shifts δ;
3. Sampling the probability of shifts p;
4. Sampling the mixture ω.

We discuss the details of each step below.

Step 1 (**Sampling conditional on the location of the shifts and the mixture**): We sample from the posterior distribution of $\phi, \sigma_\upsilon, \sigma_\eta$, R and α_1 conditional on the probability of shifts p, the realizations of the jumps δ and the realization of the mixture ω, i.e., we sample from the posterior conditional density

$$f(\theta_{(-p)}, R, \alpha_1 | p, \delta, \omega, y). \tag{9}$$

Now write this density as the product of a conditional and a marginal density, viz.,

$$f(\theta_{(-p)}, R, \alpha_1 | p, \delta, \omega, y) = f(R, \alpha_1 | \theta, \delta, \omega, y) f(\theta_{(-p)} | p, \delta, \omega, y),$$

suggesting a way to sample the distribution (9) by drawing R and α_1 from $f(R, \alpha_1 | \theta, \delta, \omega, y)$ and $\theta_{(-p)}$ from $f(\theta_{(-p)} | p, \delta, \omega, y)$. The first component $f(R, \alpha_1 | \theta, \delta, \omega, y)$ is a multivariate conditional Gaussian density since model (8) is a conditional Gaussian state space model. Accordingly, it can be sampled using a relatively straightforward extension of the simulation smoother developed by De Jong and Shephard (1995). The details are given in Appendix 4.

The second component $f(\theta_{(-p)} | p, \delta, \omega, y)$ is sampled using the Gibbs sampler, i.e., by drawing iteratively from $f(\phi | \theta_{(-\phi)}, \delta, \omega, y)$, $f(\sigma_\upsilon | \theta_{(-\sigma_\upsilon)}, \delta, \omega, y)$ and $f(\sigma_\eta | \theta_{(-\sigma_\eta)}, \delta, \omega, y)$. The procedure to sample from each of these conditional densities is similar and we provide details only for the first. Using Bayes' Theorem, $f(\phi | \theta_{(-\phi)}, \delta, \omega, y) \propto f(y | \theta, \delta, \omega) \pi(\phi)$ where $\pi(\phi)$ is the prior density. Note that the normalizing constant is unknown. We can apply the adaptive rejection sampler of Gilks and Wild (1992) and Gilks (1996). The basic idea is to construct an envelope function of the log of the target density without requiring the specification of the normalizing constant. This envelope function approaches the true density as sampling progresses, hence it provides an efficient algorithm [2].

Step 2 (**Sampling the process of shifts** δ): The conditional posterior density of δ is given by $f(\delta | R, \alpha_1, \theta, \omega, y)$. It can be sampled by drawing iteratively from

$$f(\delta_t | \delta_{(-t)}, R, \alpha_1, \theta, \omega, y) \qquad \text{for } t = n, n-1, ..., 1,$$

[2] The second component $f(\theta_{(-p)} | p, \delta, \omega, y)$ could also be sampled all at once using a Metropolis-Hastings algorithm, as in Kim, Shephard and Chib (1998). However, the envelope density for $f(\theta_{(-p)} | p, \delta, \omega, y)$ is difficult to construct and as a result, the algorithm often rejects a long sequence of proposed values. Hence, this approach is not used.

where $\delta_{(-t)}$ denotes the vector of δ with the t^{th} element excluded. Let $Y_t = (y_1, ..., y_t)$ and $Y_t^+ = (y_{t+1}, ..., y_n)$, then for each individual run,

$$\begin{aligned} & f(\delta_t = 1|\delta_{(-t)}, R, \alpha_1, \theta, \omega, y) \\ = & \frac{f(Y_t^+|\delta_t = 1, \delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_t) f(\delta_t = 1|\delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_t)}{\sum_{i=0}^{1} f(Y_t^+|\delta_t = i, \delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_t) f(\delta_t = i|\delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_t)}. \end{aligned}$$

Now, since

$$f(\delta_t = 1|\delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_t) = f(\delta_t = 1|p) = p,$$

we have

$$\begin{aligned} & f(\delta_t{=}1|\delta_{(-t)}, R, \alpha_1, \theta, \omega, y) \\ = & \frac{p \prod_{j=t+1}^{n} f(y_j|\delta_t = 1, \delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_{j-1})}{\left\{ p \prod_{j=t+1}^{n} f(y_j|\delta_t = 1, \delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_{j-1}) + \right. \\ & \left. (1-p) \prod_{j=t+1}^{n} f(y_j|\delta_t = 0, \delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_{j-1}) \right\}. \end{aligned}$$

Because δ_t is a binary variable, sampling is straightforward once the likelihood functions are evaluated sequentially. Computationally, it is more convenient to look at the posterior odds ratio:

$$\frac{f(\delta_t = 1|\delta_{(-t)}, R, \alpha_1, \theta, \omega, y)}{f(\delta_t = 0|\delta_{(-t)}, R, \alpha_1, \theta, \omega, y)} = \frac{p \prod_{j=t+1}^{n} f(y_j|\delta_t = 1, \delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_{j-1})}{(1-p) \prod_{j=t+1}^{n} f(y_j|\delta_t = 0, \delta_{(-t)}, R, \alpha_1, \theta, \omega, Y_{j-1})}$$

whose logarithm can be used to avoid numerical problems due to rounding errors.

The procedure for this step is similar to that of McCulloch and Tsay (1993) who study a finite order autoregressive process with random shifts in mean or variance. In practice a large number of draws may be necessary to ensure convergence because this is a one-step sampler, especially with a large sample size.

Step 3 (**Sampling the probability of shifts** p): For this case, the sampling is straightforward because the posterior conditional density satisfies

$$f(p|\delta, R, \alpha_1, \theta_{(-p)}, \omega, y) = f(p|\delta).$$

Using Bayes' Theorem, $f(p|\delta) \propto f(\delta|p)f(p)$. If the prior on p has a beta distribution, i.e., $p \sim beta(\gamma_1,\gamma_2)$, the conditional posterior distribution is given by

$$f(p|\delta) \sim beta(\gamma_1 + k, \gamma_2 + n - 1 - k),$$

where k is the number of shifts and $n-1$ is the effective sample size, see DeGroot (1970, p.160).

Step 4 (**Sampling the mixture** ω): This step is straightforward since

$$f(\omega_t = j|\delta, R, \alpha_1, \theta, y) = f(\omega_t = j|\varepsilon_t^*) \propto f(\varepsilon_t^*|\omega_t = j)f(\omega_t = j).$$

Note that $\omega_t = j$ means that ε_t^* is drawn from the j^{th} component of the mixture (7).

Remark 4 *The most important step is the second. Note that it would be a mistake to condition upon h_t and μ_t instead of R and α_1, since this would lead to degeneracy and δ_t could not be updated.*

Remark 5 *It is possible to model the probability of the level shifts as a function of some covariates in which case only steps 2 and 3 need to be modified. Although this is technically straightforward, in practice such covariates are difficult to find. We will not pursue this direction in the current paper, though it is an interesting avenue for exploration in subsequent work.*

4.2 The specifications of the priors

We shall use independent priors. Those for ϕ and σ_υ are the same as in Kim, Shephard and Chib (1998). For ϕ,

$$\pi(\phi) \propto \left\{\frac{(1+\phi)}{2}\right\}^{\phi^{(1)}-1} \left\{\frac{(1-\phi)}{2}\right\}^{\phi^{(2)}-1}, \quad \text{with } \phi^{(1)}, \phi^{(2)} > \frac{1}{2}.$$

This distribution has support on the interval $(-1, 1)$ with a mean of $2\phi^{(1)}/(\phi^{(1)} + \phi^{(2)}) - 1$. In the empirical applications reported below, we use $\phi^{(1)} = 20$ and $\phi^{(2)} = 1.5$, implying a prior mean of 0.86. For σ_υ,

$$\sigma_\upsilon^2 \sim \mathcal{IG}(\sigma_r/2, S_\sigma/2),$$

where $\mathcal{IG}(\sigma_r/2, S_\sigma/2)$ denotes the inverse gamma distribution with shape parameter $\sigma_r/2$ and scale parameter $S_\sigma/2$. We set $\sigma_r = 5$ and $S_\sigma = 0.01 \times \sigma_r$.

The prior distributions for p and σ_η are chosen to reflect the belief that the level shifts are infrequent and that their magnitude is large. For p, we specify

$$p \sim beta(\gamma_1, \gamma_2),$$

with $\gamma_1 = 1$ and $\gamma_2 = 40$. This implies a prior mean of $1/40$ so that a shift occurs on average every 40 days. For σ_η,

$$\sigma_\eta^2 \sim \mathcal{IG}(\sigma_r^*/2, S_\sigma^*/2),$$

with $\sigma_r^* = 20$ and $S_\sigma = 60$. This implies an approximate prior mean of 3.33 and variance of 1.39 [3]. Finally, we use the following prior distribution for the initial states $\alpha_1 = (h_1, \mu_1)$

$$\alpha_1 \sim N(0, P) \text{ with } P = Diag(1 \times 10^6, 1 \times 10^6).$$

5 Filtering, likelihood evaluation and forecasting

We now discuss issues related to filtering, likelihood evaluation and forecasting conditional on estimated parameters. With some abuse of notation, we use θ to denote the estimates, which can be the posterior means or medians. Let $X_t = (x_1, ..., x_t)'$ denote the vector of returns up to time t. The issue of interest is to obtain the filtered signals, $(\alpha_t|X_t, \theta)$, the p-step ahead predicted volatility, $E(\exp(h_{t+p} + \mu_{t+p})|X_t, \theta)$, and the likelihood function upon which diagnostics will be based.

5.1 Filtering

The objective of the filtering procedure is to recursively obtain a sample of draws from $(\alpha_t|X_t, \theta)$ for $t = 1, ..., n$. Then, the filtered signal $E(\exp(\alpha_t)|X_t, \theta)$ can be computed by taking an average over those draws. Formally, the link between the distributions of $(\alpha_{t+1}|X_{t+1}, \theta)$ and $(\alpha_t|X_t, \theta)$ is given by, using Bayes' Theorem,

$$\begin{aligned} f(\alpha_{t+1}|X_{t+1}, \theta) &= \frac{f(x_{t+1}|\alpha_{t+1}, X_t, \theta)}{f(x_{t+1}|X_t, \theta)} f(\alpha_{t+1}|X_t, \theta) \\ &= \frac{f(x_{t+1}|\alpha_{t+1}, X_t, \theta)}{f(x_{t+1}|X_t, \theta)} \int f(\alpha_{t+1}|\alpha_t, X_t, \theta) dP(\alpha_t|X_t, \theta). \end{aligned} \tag{10}$$

This is not directly useful since it involves an integral which cannot be evaluated analytically. A solution to this problem is to use the particle filter as in Gordon, Salmond and

[3] Specifying prior distributions for p and σ_η that allow frequent jumps of small magnitude would essentially lead to a random walk model which is undesirable in our context.

Smith (1993) and Kim, Shephard and Chib (1998). Particle filtering is a technique for implementing recursive filtering by Monte Carlo simulations. The main idea is to evaluate the required density by a set of random draws with associated weights and to compute estimates based on these draws and weights. The approximation error can be made arbitrarily small by using a large number of draws. More specifically, the required density $f(\alpha_{t+1}|X_{t+1},\theta)$ can be represented as random draws from $f(\alpha_{t+1}|X_t,\theta)$ with associated weights given by $f(x_{t+1}|\alpha_{t+1},X_t,\theta)/f(x_{t+1}|X_t,\theta)$. Using (10), for a given sample of M draws $\alpha_t^{(j)}$ $(j=1,...,M)$ from the distribution of $(\alpha_t|X_t,\theta)$, a sample from $f(\alpha_{t+1}|X_t,\theta)$ can be obtained by drawing from $f(\alpha_{t+1}|\alpha_t^{(j)},X_t,\theta)$. This distribution depends on whether a shift occurs at time t and is given by

$$(\alpha_{t+1}|\alpha_t^{(j)},X_t,\theta) \sim \delta_t W_{1t}^{(j)} + (1-\delta_t)W_{2t}^{(j)} \tag{11}$$

with

$$W_{1t}^{(j)} \sim N\left(\begin{bmatrix}\phi & 0\\ 0 & 1\end{bmatrix}\alpha_t^{(j)},\ \begin{bmatrix}\sigma_\upsilon^2 & 0\\ 0 & \sigma_\eta^2\end{bmatrix}\right)$$

and

$$W_{2t}^{(j)} \sim N\left(\begin{bmatrix}\phi & 0\\ 0 & 1\end{bmatrix}\alpha_t^{(j)},\ \begin{bmatrix}\sigma_\upsilon^2 & 0\\ 0 & 0\end{bmatrix}\right).$$

The associated weights are given by

$$w_{t+1}^{(j)} = \frac{f(x_{t+1}|\alpha_{t+1}^{(j)},X_t,\theta)}{\sum_{j=1}^M f(x_{t+1}|\alpha_{t+1}^{(j)},X_t,\theta)},$$

where

$$f(x_{t+1}|\alpha_{t+1}^{(j)},X_t,\theta) \sim N(0,\exp(h_{t+1}^{(j)}+\mu_{t+1}^{(j)})).$$

5.2 The likelihood function

The conditional density function $f(x_{t+1}|X_t,\theta)$ does not have a tractable analytical representation but it can be approximated using Monte Carlo simulations. Note that if one has a random sample from the joint density

$$f(x_{t+1},\alpha_{t+1},\alpha_t,|X_t,\theta), \tag{12}$$

one can then recover the marginal density $f(x_{t+1}|X_t,\theta)$. To sample from (12), we use the relationship

$$f(x_{t+1},\alpha_{t+1},\alpha_t,|X_t,\theta) = f(x_{t+1}|\alpha_{t+1},\theta)f(\alpha_{t+1}|\alpha_t,\theta)f(\alpha_t|X_t,\theta).$$

First, given a sample of draws $\alpha_t^{(j)}$ $(j = 1, ..., M)$ from $f(\alpha_t|X_t, \theta)$, we draw $\alpha_{t+1}^{(j)}$ from (11). Then, $f(x_{t+1}|\alpha_{t+1}^{(j)}, \theta)$ can easily be computed. Accordingly, $f(x_{t+1}|X_t, \theta)$ can be estimated by

$$\frac{1}{M}\sum_{j=1}^{M} f(x_{t+1}|\alpha_{t+1}^{(j)}, X_t, \theta)$$

and the likelihood function can be approximated by

$$\prod_{t=0}^{n-1}\left(\frac{1}{M}\sum_{j=1}^{M} f(x_{t+1}|\alpha_{t+1}^{(j)}, X_t, \theta)\right).$$

The approximation can be made precise by choosing M large enough. From this likelihood function, various diagnostics can be constructed.

5.3 Forecasting

We now discuss the construction of the one and multiple-steps ahead predictions for the density $f(\alpha_{t+p}|X_t, \theta)$, volatility $E(\exp(h_{t+p} + \mu_{t+p})|X_t, \theta)$ and log volatility $E(h_{t+p} + \mu_{t+p}|X_t, \theta)$.

The predictive density $f(\alpha_{t+p}|X_t, \theta)$ can be obtained by the same method as in Jacquier, Polson and Rossi (1994). Let $\alpha_f = (\alpha_{t+1}, ..., \alpha_{t+p})'$ denote a vector of future states and let $X_f = (x_{t+1}, ..., x_{t+p})$ denote a vector of future returns. The key insight is that if X_f were known, it would be possible to draw from $f(\alpha_f|X_f, X_t, \theta)$. This suggests augmenting the data by X_f. More specifically, the procedure is the following using an initial value for α_f:

1. Draw from $f(X_f|X_t, \alpha_f, \theta)$, which is straightforward since

$$X_f|X_t, \alpha_f, \theta \sim N(0, Diag(\exp(\alpha_f)));$$

2. Draw from $f(\alpha_f|X_f, X_t, \theta)$, which can be done using the algorithm developed in Section 4.1, in particular steps 1, 2 and 4;

3. Repeat to generate a random sample.

Let $(\alpha_f^{(1)}, ..., \alpha_f^{(M)})$ denote the resulting vector of M draws. The predictive density can then be estimated using a kernel smoothing method.

We now discuss the problem of predicting volatility. The minimum mean squared error (MSE) estimate of the of the p-step ahead forecast of volatility can be approximated by

$$E(\exp(h_{t+p} + \mu_{t+p})|X_t, \theta) \approx \frac{1}{M}\sum_{j=1}^{M} \exp(h_{t+p}^{(j)} + \mu_{t+p}^{(j)})$$

and the estimate of average volatility over the one to p-step ahead forecasts is

$$\bar{\sigma}^2_{t,p} \equiv \sum_{i=1}^{p} E(\exp(h_{t+i} + \mu_{t+i})|X_t, \theta) \approx \frac{1}{M} \sum_{i=1}^{p} \sum_{j=1}^{M} \exp(h^{(j)}_{t+i} + \mu^{(j)}_{t+i}).$$

Again, these approximations can be made precise by choosing M large enough.

If level shifts are expected to be rare, as will be the case in our applications, it is sensible to ignore them when forecasting, especially since their timing and magnitude exhibit considerable uncertainty. The algorithm then substantially simplifies and simulations for the predictive density are not needed for the volatility forecasts. This is so because ignoring shifts,

$$\begin{aligned} \mu_{t+p} &= \mu_t, \\ h_{t+p} &= \phi^p h_t + \sum_{i=1}^{p} \phi^{i-1} \sigma_\upsilon \upsilon_{t+p-i}, \end{aligned}$$

and the predicted volatility p-steps ahead is

$$\begin{aligned} E(\exp(h_{t+p} + \mu_{t+p})|X_t) &= E(\exp(h_{t+p} + \mu_t)|X_t) \\ &= E(\exp(\phi^p h_t + \mu_t + \sum_{i=1}^{p} \phi^{i-1} \sigma_\upsilon \upsilon_{t+p-i})|X_t) \\ &= E(\exp(\phi^p h_t + \mu_t)|X_t) \prod_{i=1}^{p} E(\exp(\phi^{i-1} \sigma_\upsilon \upsilon_{t+p-i})) \\ &= E(\exp(\phi^p h_t + \mu_t)|X_t) \prod_{i=0}^{p-1} \exp(\frac{\phi^{2i} \sigma^2_\upsilon}{2}), \end{aligned} \tag{13}$$

where the last equality follows by normality of υ_t. Note that $E(\exp(\phi^p h_t + \mu_t)|X_t)$ can be evaluated using particle filtering, as discussed in Section 5.1.

If we only wish to predict the logarithm of volatility, $E(h_{t+p} + \mu_{t+p}|X_t)$, then information about its predictive density is unnecessary regardless of whether out-of-sample shifts are allowed since

$$\begin{aligned} & E(h_{t+p} + \mu_{t+p}|X_t, \theta) \\ = \; & E(\phi^p h_t + \sum_{i=1}^{p} \phi^{i-1} \sigma_v v_{t+p-i}|X_t) + E(\mu_t + \sum_{i=1}^{p} \delta_{t+i-1} \sigma_v \eta_{t+i-1}|X_t) \\ = \; & \phi^p h_t + \mu_t. \end{aligned}$$

This result follows from the assumption that the distribution of the level shifts are symmetric about zero.

6 Simulation evidence on the performance of the method

In this section we assess whether the estimation method proposed above can produce informative results. The key issues of interest are whether, for practically relevant cases, the procedure is able to detect level shifts when they occur and whether it tends to report spurious shifts (i.e., report shifts when the process does not have any). To this end, we simulated processes with and without level shifts and examine if the results of our method yield correct characterizations of the processes. To make the evaluation practically relevant, the parameters used for the simulations were those estimated from the S&P 500 returns for the sample period 1980.1-2005.12, to be discussed in more detail below. For processes with level shifts, the parameter values were the posterior means obtained from estimating model (4) using the algorithms and priors presented earlier. For processes without level shifts, the values were the posterior means obtained from estimating (3), with the same data, using the priors and algorithm of Kim, Shephard and Chib (1998). The simulations used samples of sizes $n = 2000$, 4000 and all relevant statistics were obtained using 5000 draws, discarding the first 3000. The code was written in the software package Ox with the aid of the SVPack 2.0 developed by Neil Shephard. The seeds used for the random numbers were the default values set by Ox.

Consider first the case with level shifts. The series were generated using (4) with $\phi = 0.953, \sigma_v = 0.148, \sigma_\eta = 1.679, \delta_t \sim B(1, 0.00187)$ and $(h_1, \mu_1) = (0.450, -0.210)$. Figure 3 presents the result from a typical realization with $n = 2000$. The series contains two shifts, with the second negligible in magnitude. The model captures both the magnitude and the location of the first shift (see Figure 3.b). The posterior mean of the shift probability, displayed in Figure 3.c, shows that the shift is significant and the location is precisely estimated. The estimated model also successfully extracts the overall signal, shown in Figure 3.d. The posterior densities (Figures 3.e-3.h) are correctly centered and present clear evidence that the data are informative about the parameters of interest. The correlograms show that the algorithm performs well in generating invariant distributions: draws for p and σ_η are nearly independent while draws for the AR(1) parameters, ϕ and σ_v, exhibit correlation which basically vanishes after lag 50. We also tried altering the initial values and examined different realizations; the patterns were identical. Figure 4 pertains to the case with $n = 4000$ for which more shifts occur but similar general conclusions apply.

Consider now the case with no shift. The series were generated using (3) with $\phi = 0.986$, $\sigma_v = 0.123$ and $\mu = -0.297$, the posterior means from estimating the model (3). Figures

5 and 6 present the results for $n = 2000$ and $n = 4000$, respectively. They show that our estimation procedure does not report spurious shifts because the posterior mean of the shift component is essentially zero and the probability of shifts is negligible. Overall, we conclude that our estimation method is reliable.

7 Applications to S&P 500 and NASDAQ returns

In this section, we use our model to study the volatility of the S&P 500 and NASDAQ daily returns over the period 1980.1-2005.12. The following are discussed: a) the posterior distributions of the parameters using the full sample; b) whether our model evaluated at the posterior means can replicate keys features of the data; c) the forecasting performance of our model relative to other popular forecasting models.

7.1 The posterior distributions based on the full sample

We used the priors discussed in Section 4 and generated the posterior distributions based on 5000 draws, discarding the first 3000. The initial parameter values used to start the sampling chains were $\phi = 0.98, \sigma_v^2 = 0.01, \sigma_\eta^2 = 5, p = 0.005, \omega_t = 4.94$ $(t = 1, ..., n)$ and $\delta_t = 1$ if t is a multiple of 50 and $\delta_t = 0$ otherwise. We experimented with different initial values and the results did not change.

Consider first the results for the S&P 500 daily returns. Figure 7 presents the posterior means of the level shift component and the log volatility process. Figure 8 summarizes the distributions of the parameters and the correlograms for the draws. Several interesting features emerge. First, the shifts are rare. The posterior distribution of p has a mean of 0.00187 with a 95% confidence interval of (0.00095 0.00301). In terms of duration, this implies that on average, a shift occurs every 535 days (with a 95% confidence interval of (332, 1053) days). The results are substantially different from the prior which implies an average of one shift per 40 days, indicating that the data are informative about this parameter. Second, although rare, the level shift component is quantitatively the most important. To see this, consider the following decomposition with $s_t = \mu_t + h_t$ and with $\bar{s}$, $\bar{u}$ and $\bar{h}$ denoting the sample means of the corresponding processes,

$$s_t = \bar{s} + (\mu_t - \bar{\mu}) + (h_t - \bar{h}).$$

The quantities

$$\frac{\sum_{i=1}^n (\mu_t - \bar{\mu})^2}{\sum_{i=1}^n s_t^2} \text{ and } \frac{\sum_{i=1}^n (h_t - \bar{h})^2}{\sum_{i=1}^n s_t^2} \tag{14}$$

measure the relative contributions of μ_t and h_t to the overall variation in volatility. Their values are 0.52 and 0.21, respectively. This strong result indicates that the dominant contribution of the time-variation in volatility consists of movements with a horizon of more than one year. In light of this, a standard SV model can only capture a narrow piece of the picture. Finally, the posterior distribution of the AR(1) coefficient has a mean of 0.953 with a 95% confidence interval of (0.929,0.970). This implies that the short-memory component is indeed much less persistent when level shifts are accounted for, as typical estimates of ϕ using standard SV models are very close to 1. Indeed, in our model the half life of a typical shock is only about 14 days with a 95% confidence interval of (9.4, 22) days. This contrasts sharply with the estimate obtained using a standard SV model for which the posterior mean of ϕ implies a half life of 49 days with a 95% confidence interval of (31, 86) days.

A close look at Figure 7 suggests that the shifts often coincide with important events. The first shift occurred in August, 1982, often referred to as the start of the five year bull market, and was followed by a brief period of relatively high volatility that ended in March, 1983. Then the market returned to a volatility level that was roughly the same as before August, 1982 until the market was hit by "Black Monday" in 1987, the end of the five year bull market. The smoothed estimates of μ_t around "Black Monday" are:

Date (Weekday)	μ_t
Before 10/9/1987 (Friday)	-0.26
10/12/1987 (Monday) to 10/16/1987 (Friday)	2.21
10/19/1987 (Monday) to 10/26/1987 (Monday)	3.53
After 10/27/1987 (Tuesday)	1.01

The results are quite informative. First, the market was already showing signs of unusually high volatility one week before the crash occurred. Second, the mean level of volatility further increased on "Black Monday" and this high volatility lasted one week, until October 26, 1987. The volatility then dropped suddenly on October 27, 1987, the new regime lasting until Monday January 18, 1988, when the market returned to a volatility level comparable to that prior to the crash. The market entered a long period of low volatility starting in April, 1992. Volatility started to increase again in June, 1996, predating the Asian financial crisis. (Note that Southeast Asia's export growth slowed dramatically in the Spring of 1996 and the crisis was triggered in July 1997 in Thailand.) However, this transition was gradual and increases in volatility continued until July, 1998. Then, a sudden decrease in volatility occurred between June and July, 2003, this new regime persisting to the end of the sample.

For the NASDAQ returns, the results are summarized in Figures 9 and 10. The implications are similar to those for the S&P 500 returns. The ratios defined in (14) are 0.78 and 0.12, respectively, again indicating that the level shift component accounts for most of the variability in volatility. The ratios indicate that this feature is even more pronounced for NASDAQ volatility. The posterior distribution of the probability of shifts has a mean of 0.00248 with a 95% confidence interval of (0.00131, 0.00410). In terms of regime duration, this implies, on average, one shift every 403 days (with a 95% confidence interval of (244, 758) days); in terms of the number of shifts, the mean is 15 with a 95% confidence interval of (12, 20). The estimates of the level shift component are similar to those for the S&P 500 index. This suggests that common underlying causes affect volatility of the two markets. With respect to the AR(1) coefficient of the short-memory component, the posterior distribution has a mean of 0.918 with a 95% interval of (0.883, 0.943). This estimate again implies much less persistence than that typically reported using standard SV or other models. The half life of a typical shock is only about 8 days with a tight 95% confidence interval of (5.6, 11.8) days. This again contrasts sharply with the estimates obtained using the standard SV model for which the posterior mean implies a half life of 28 days with a 95% confidence interval of (22, 40) days.

Our model also fares well in terms of likelihood values. The log-likelihood value of our model (resp., the SV model) are -8559.2 (resp., -8589.2) for the S&P 500 and -8974.3 (resp., -9036.9) for the NASDAQ index. The Bayesian Information Criterion (BIC) also indicates that our model provides a better in-sample fit than a SV model without level shifts. For the S&P 500 (resp., NASDAQ) index, the value of the BIC is 2.613 (resp., 2.740) for the level shift model and 2.621 (resp., 2.757) for the standard SV model.

Since long-memory is often entertained as a plausible feature of volatility series, we also evaluated a fractionally integrated generalized autoregressive conditional heteroskedasticity model, FIGARCH(1,d,1), given by

$$\begin{aligned} x_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \gamma + \beta_1 \sigma_{t-1}^2 + [1 - \beta_1 L - (1 - \phi_1 L)(1 - L)^d] x_t, \\ \varepsilon_t &\sim i.i.d.\ N(0,1), \end{aligned}$$

where x_t denotes demeaned returns. We estimated the model using the G@RCH 4.2 package (see, Laurent and Peters, 2006). The parameter estimates (standard errors in parentheses) obtained for the S&P 500 are $\hat{\gamma} = 1.4360$ (0.286), $\hat{d} = 0.427$ (0.022), $\hat{\phi}_1 = 0.299$ (0.023) and $\hat{\beta}_1 = 0.619$ (0.053). The log-likelihood value is -8714.9 with 2.661 as the corresponding value

of the BIC. For the NASDAQ series the results are: $\hat{\gamma} = 0.968$ (0.184), $\hat{d} = 0.392$ (0.023), $\hat{\phi}_1 = 0.231$ (0.044) and $\hat{\beta}_1 = 0.460$ (0.053). The log-likelihood value is -9080.5 with 2.772 as the corresponding value of the BIC. Hence, our level shift model also provides a better in-sample fit than the FIGARCH model for both series.

7.2 Does the model explain key features of the data?

In this section, we consider two issues. First, we provide some standard diagnostic statistics to assess the quality of the fitted process of our model. Second, and more importantly, we analyze simulated artificial samples using the estimated parameter values (the posterior means of our model). These allow us to examine whether our model explains the key features of the data shown in Figures 1 and 2 pertaining to the full sample autocorrelation function and the path the log-periodogram estimates of the long-memory parameter take as a function of the number of frequency ordinates used.

Figure 11 presents a normal QQ plot of the estimated residuals for the model fitted to the S&P 500 return series and the autocorrelations for various transforms of these residuals. The QQ plot shows that the residuals are slightly skewed to the left but otherwise closely follow a standard normal distribution. We experimented with different subsamples and the results were identical. The autocorrelations suggest that no significant correlation remains in the log and power transforms of the standardized returns (x_t/σ_t).

It would be sensible to argue that the above diagnostics are limited. To better understand the implications of our model and better assess its differences with others, we examined the time and spectral properties of the simulated samples. To this end, we generated 500 samples of size $n = 6564$, using $\phi = 0.953, \sigma_v = 0.148, \sigma_\eta = 1.679, \delta_t \sim B(1, 0.00187)$ and $(h_1, \mu_1) = (0.450, -0.210)$. For each simulated sample, we computed the autocorrelations for all lags between 1 and 6563 and the log-periodogram estimates of the long-memory parameter using a wide range for the number of frequency ordinates. The averages over the 500 samples are reported in Figure 12. They generate patterns closely resembling those of the actual data depicted in Figure 1. To the best of our knowledge, there is no convincing evidence that a covariance stationary short or long-memory model is able to reproduce such a match. The only type of process we found able to match such features is a strongly mean-reverting non-stationary long-memory process with a long-memory parameter well above 0.5 (e.g., around 0.8), see Perron and Qu (2008). Such non-stationary processes are unlikely candidates for a useful model of volatility. Hence, we believe this is an interesting and important result that clearly indicates the relevance of our model for describing the volatility of the return

series analyzed. Similar results were obtained for the NASDAQ series, the details of which are omitted.

7.3 Forecasting volatility for S&P 500 and NASDAQ daily returns

We now consider one and multiple-step ahead forecasts for volatility using a setup similar to that of Stărică and Granger (2005). For our level shift model, we used the first 2000 observations for initial estimation and re-estimated the model with the addition of 20 observations. We then made forecasts for horizons up to 20 days ahead. We continued this and recorded the relevant values. This forecast specification is labeled recursive. For the standard SV model, we examined two specifications: a) rolling window forecasts with the window size fixed at 2000 and b) recursive forecasts as described above. We consider rolling window forecasts for the standard SV model to allow it to accommodate nonstationarity to some extent. This is not relevant for our level shift model since the issue of nonstationarity is automatically taken into account via the level shift process.

The basis for comparison is the Mean Squared Error (MSE) criterion (see Hansen, 2006, for a justification). Let $\hat{\sigma}^2_{t+j}$ denote the volatility forecast j days ahead. The average volatility forecast at horizon p is then given by

$$\hat{\sigma}^2_{t,p} = \sum_{j=1}^{p} \hat{\sigma}^2_{t+j}.$$

An unbiased estimate for the volatility at horizon p is given by the cumulative squared demeaned returns, i.e.,

$$x^2_{t,p} = \sum_{j=1}^{p} x^2_{t+j},$$

which suggests estimating the MSE by

$$MSE(p) = \frac{1}{J} \sum_{t=1}^{J} (\hat{\sigma}^2_{t,p} - x^2_{t,p})^2,$$

where J is the total number of forecasts produced. For $p > 1$, $\hat{\sigma}^2_{t,p}$ involves averaging in order to cancel some of the idiosyncratic noise in the daily data, with the objective of obtaining a better measure of the quality of the forecasts. Note that the series $x^2_{t,p}$ contains large outliers which may have an overwhelming effect on the measure $MSE(p)$. To avoid this problem, we discarded observations whose absolute values were above 6%. As a result, we eliminated 8 observations from the S&P 500 series. None were discarded from the NASDAQ series since,

in this case, the results did not change. After discarding, we were left with 6556 observations for the S&P 500 series and 6564 for the NASDAQ series.[4]

Figures 13 and 14 present the estimated MSE ratios of the forecasts obtained from the level shift model relative to those obtained from the SV and FIGARCH models. Consider first the comparisons with the standard SV model. For S&P 500 returns, the forecasts from both models perform similarly, with the MSE ratios ranging between 0.98 and 1.09. For NASDAQ returns, the level shift model noticeably outperforms the SV model with the MSE ratios ranging between 0.79 and 0.97, when considering the recursive version of the forecasts from the SV model. With the rolling version, the range is 0.90 to 1.08. Note that when using the recursive SV forecasts our model does better at all horizons while with rolling SV forecasts the improvements occur at short horizons. This reflects to some extent the importance of non-stationarity. When comparing our model to the FIGARCH model, the forecast accuracy improvement is not as pronounced. But here also, our model performs better at short horizons for NASDAQ returns.

Following Anderson et al. (2003), and in the tradition of Mincer and Zarnowitz (1969), we also evaluated relative forecasting accuracy by regressing proxies for volatility on a constant and the forecasts of the various models. We first considered regressions of the form

$$x_{t,p}^2 = b_0 + b_1 \hat{\sigma}_{t,p}^2 + u_t, \tag{15}$$

where $\hat{\sigma}_{t,p}^2$ denotes the p-step ahead forecast for each of the following five models: a) SV with level shifts (labeled LS), b) SV without level shifts in which forecasts are obtained using a recursive method (labeled SV-Rec), c) SV without level shifts in which forecasts are obtained using a rolling window (labeled SV-Rol), d) FIGARCH with forecasts obtained using a recursive method (labeled FIG-Rec) and e) FIGARCH with forecasts obtained using a rolling window (labeled FIG-Rol). A good forecast will have an estimate of b_0 insignificantly different from 0 and an estimate of b_1 insignificantly different from 1. For such forecasts, the "best" yields the largest adjusted R^2 value. We estimated regression (15) for forecast horizons between 1 and 5 days. The results are presented in Table 1. Consider first the case of the NASDAQ index. For the level shift model, we have that in all cases the value 0 is contained in the confidence interval for b_0 and the value 1 is contained in the confidence interval for b_1. This is also the case for both FIGARCH forecasting methods. For the SV

[4]If no observations are discarded from the S&P 500 series, the forecast MSE ratios of our level shift model relative to the standard SV model are between 1.10 and 1.54. A closer examination showed that this result is indeed dominated by a few extremal observations and the relative MSE ratios stablize once they are excluded.

model without level shifts, the value 1 is not in the confidence interval for b_1 at horizon 1, though it is for the other horizons. More importantly, the adjusted R^2 of the regression is highest for the level shift model at all horizons. Hence by this measure, the level shift model forecasts better than the other methods. For the S&P 500 index, the results are more ambiguous. The value 1 is in the confidence interval for b_1 only at horizons 1, 4 and 5 for all forecasting methods. The adjusted R^2 is the highest for the level shift model at horizons 1, 2 and 3, but the difference is only 0.01 for horizons 4 and 5. So for this index, there is no clear evidence that one of the methods is superior or inferior to the others.

To further assess the relative performance of the forecasting methods, we also considered regressions of the form

$$x_{t,p}^2 = b_0 + b_1 \hat{\sigma}_{t,p}^2 + b_2 \hat{\sigma}_{t,p,i}^2 + u_t, \tag{16}$$

where $\hat{\sigma}_{t,p}^2$ denotes the p-step ahead forecast from the level shift model and $\hat{\sigma}_{t,p,i}^2$ denotes the forecast from one of the other four methods. In this case, the level shift model can be said to generate better forecasts than a competitor if the values (0,1,0) are in the confidence intervals for (b_0, b_1, b_2), respectively. Also, one should expect the difference in the adjusted R^2 to be small compared to the case in which the forecast from the alternative method is not included. The results are presented in Table 2. For the NASDAQ series the results, with few exceptions, suggest that the level shift model is the best forecaster: the values $(0, 1, 0)$ tend to lie in the confidence intervals for (b_0, b_1, b_2) and minimal gain is achieved, in terms of adjusted R^2, from including forecasts from other models. For the S&P 500 index, the results are again more ambiguous, though in general tend to favor the level shift model.

The above results are important since they imply that not only does the level shift model provide a better description of the data in sample, but often also increases the forecasting performance. Hence, the level shift component appears to be a genuine element of the data, not simply a modeling convenience allowing for a better in-sample fit. This implication contrasts with the common perception that structural change models are not useful for forecasting [5].

We conjecture that forecasts for longer horizons could be further improved by explicit modeling of the sign of shifts and by allowing for out-of-sample breaks. Such extensions are clearly non-trivial. Our results nevertheless indicate that investigations along these lines are worth pursuing and we hope to report results to that effect in the near future.

[5] Pesaran, Pettenuzzo and Timmermann (2006) provided another interesting example in which they considerd forecasting U.S. Treasury bill rates. They found that accounting for structural changes leads to better out-of-sample forecasts compared to a variety of alternative methods.

8 Conclusion

We considered methods of inference and forecasting for a stochastic volatility model with random level shifts. We showed that a very simple stochastic volatility model incorporating both a random level shift and a short-memory component indeed provides a better in-sample fit of the data and produces forecasts that are no worse, and sometimes better, than standard stationary short or long-memory models. These results are encouraging because our model can be extended in several directions to provide a deeper understanding of the time series structure of volatility and improve forecasting performance. This could be done by making the sign, magnitude and probability of shifts a function of covariates. This would allow for the possibility of forecasting future level shifts. However, such extensions are not trivial, though we feel that our results point to the need for future work along these lines.

References

[1] Anderson, T.G., and T. Bollerslev (1997): "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *Journal of Finance*, 52, 975-1005.

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

[3] Arteche, J. (2004): "Gaussian Semiparametric Estimation in Long Memory in Stochastic Volatility and Signal Plus Noise Models," *Journal of Econometrics*, 119, 131-154.

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

[5] Beran, J. (1994): *Statistics for Long Memory Processes*. New York: Chapman & Hall.

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

[7] 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.

[8] Carter, C.K., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541-553.

[9] Chib, S. (1995): "Marginal Likelihood from Gibbs Output," *Journal of the American Statistical Association*, 90, 1313-1321

[10] Chib, S. (1998): "Estimation and Comparison of Multiple Change Point Models," *Journal of Econometrics*, 86, 221-241.

[11] Chib, S., and E. Greenberg (1995): "Understanding the Metroplois-Hastings Algorithm," *The American Statistician*, 49, 327-335.

[12] De Jong, P. (1991): "The Diffuse Kalman filter," *Annals of Statistics*, 19, 1073-1083.

[13] De Jong, P., and N. Shephard (1995): "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339-350.

[14] DeGroot, M. (1970): *Optimal Statistical Decisions*, New York: McGraw-Hill.

[15] Deo, R.S., and C.M. Hurvich (2001): "On the Log Periodogram Regression Estimator of the Memory Parameter in the Long Memory Stochastic Volatility Models," *Econometric Theory*, 17, 686-710.

[16] Diebold, F.X., and A. Inoue (2001): "Long Memory and Regime Switching," *Journal of Econometrics,* 105, 131–159.

[17] Ding, Z., R.F. Engle, and C.W.J. Granger (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83-106.

[18] Duffie, D., J. Pan, and K.J. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

[19] Duffie, D., and K.J. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929-952.

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

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

[22] Gilks, W.R., and P. Wild (1992): "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337-348.

[23] Gilks, W.R. (1996): "Full Conditional Distributions," in *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

[24] Gordon, N.J., D.J. Salmond, and A. F. M. Smith (1993): "A Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation," *IEE-Proceedings F*, 140, 107-133.

[25] Granger, C.W.J. (1980): "Long Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227-238.

[26] Granger, C.W.J., and Z. Ding (1995): "Some Properties of Absolute Return: an Alternative Measure of Risk," *Annales d'Économie et de Statistique*, 40, 67-91.

[27] 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.

[28] Hansen, P.R., and A. Lunde (2006): "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131, 97-121.

[29] 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.

[30] Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Volatility Models," *Review of Economic Studies*, 61, 247–264.

[31] Jacquier, E., N. Polson, and P. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12 , 371-418.

[32] Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65, 361-393.

[33] Koop, G. and S.M. Potter (2007): "Estimation and Forecasting in Models with Multiple Breaks," *Review of Economic Studies*, 74, 763-789.

[34] Koopman, S.J. (1993): "Disturbance Smoother for State Space Models," *Biometrika*, 80, 117-126.

[35] 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.

[36] Laurent, S. and J.-P. Peters (2006): *Estimating and Forecasting ARCH Models*. Timberlake Consultants Press.

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

[38] McCulloch, R.E., and R.S. Tsay (1993): "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88, 968-978.

[39] Mikosch, T., and C. Stărică (2004): "Nonstationarities in Financial Time Series, the Long-range Dependence, and the IGARCH Effects," *Review of Economics and Statistics,* 86, 378-390.

[40] 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.

[41] Parke, W.R. (1999): "What is Fractional Integration?" *Review of Economics and Statistics* 81, 632–638.

[42] 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.

[43] 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.

[44] Pesaran, M.H., D. Pettenuzzo, and A. Timmermann (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.

[45] Shephard, N. (1994): "Partial Non-Gaussian State Space," *Biometrika*, 81, 115-131.

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

[47] Stărică, C., and C.W.J. Granger (2005): "Nonstationarities in Stock Returns," *Review of Economics and Statistics* 87, 503-522.

Appendix 1: The Gibbs sampler and data augmentation

Suppose that the objective is to obtain a sample from the posterior distribution of $\theta = (\theta_1, ..., \theta_p)'$, given by $f(\theta_1, ..., \theta_p|X)$ where $X = \{x_1, ..., x_n\}$. In practice, it is often difficult or impossible to sample directly from the joint distribution. The Gibbs sampler allows one to achieve the same objective by iteratively sampling from the univariate distributions, which is often straightforward. The basic steps involved are as follows, where without loss of generality, we let $p = 2$:

1. Consider a set of initial values, say θ_1^0 and θ_2^0;
2. Draw from the univariate distribution of θ_1 conditional on θ_2^0, i.e., $f(\theta_1|X, \theta_2^0)$. Denote the draw as θ_1^1;
3. Draw from the univariate distribution of θ_2 conditional on θ_1^1, i.e., $f(\theta_2|X, \theta_1^1)$. Denote the draw as as θ_2^1;
4. Repeat steps 2 and 3 M times with the values of the conditioning variables updated sequentially.

For the general case with p variables the draw in each step is obtained by conditioning on all but one parameter. Under mild conditions, the random samples obtained from the M draws converges weakly to the joint distribution of interest; see Geman and Geman (1984).

The method of data augmentation is useful when a latent process is involved and it is difficult to sample from the univariate distributions. For illustration, consider the following simple process:

$$\begin{aligned} x_t &= \exp(h_t/2)\varepsilon_t, \\ h_{t+1} &= \theta_1 h_t + \theta_2 \upsilon_t. \end{aligned}$$

where ε_t and υ_t are *i.i.d.* $N(0, 1)$. In this case, if the process $h = (h_1, ..., h_n)$ were known, sampling would be straightforward since

$$\begin{aligned} f(\theta_1|X, h, \theta_2) &= f(\theta_1|h, \theta_2), \\ f(\theta_2|X, h, \theta_1) &= f(\theta_2|h, \theta_1), \end{aligned}$$

and the structure essentially reduces to an AR(1) model. This suggests augmenting the data by h, i.e., treating the elements of h as additional parameters to sample from. This reasoning leads to the following Gibbs Sampler with data augmentation:

1. Consider a set of initial values, say θ_1^0, θ_2^0 and h;
2. Draw h^1 from $f(h|X, \theta_1^0, \theta_2^0)$;
3. Draw θ_1^1 from $f(\theta_1|X, h^1, \theta_2^0)$;

4. Draw θ_2^1 from $f(\theta_2|X, h^1, \theta_1^1)$;

5. Repeat steps 2 to 4 M times with the values of the conditioning variables updated sequentially.

This method can easily be extended to cover more general cases.

Appendix 2: The mixture of Kim, Shephard and Chib (1998)

As stated in Section 4, we approximate the distribution of ε_t^* by the following mixture of normals:

$$\varepsilon_t^* \overset{d}{\sim} \sum_{i=1}^{K} q_i N(m_i, \sigma_i^2).$$

We set $K = 7$ and the parameters q_i, m_i and σ_i^2 we use are those suggested by Kim, Shephard and Chib (1998) as described in the following table.

i	q_i	m_i	σ_i^2
1	0.00730	-10.1299	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	-1.08819	1.26261

Appendix 3: SV model in state space form and the Kalman filter

To simplify notation and relate the model to the state space modeling literature, we write the model as follows:

$$\begin{aligned} y_t &= Z\alpha_t + G_t u_t, \\ \alpha_{t+1} &= T\alpha_t + H_t u_t, \end{aligned} \tag{A.1}$$

where the observations are

$$y_t = \log(x_t^2 + c) - E(\log \varepsilon_t^2),$$

the state vector is $\alpha_t = (h_t, \mu_t)'$, the innovations are $u_t \sim i.i.d.\ N(0, I_3)$, the coefficients are $Z = (1, 1)$, $G_t = (\sigma(\omega_t), 0, 0)$, with ω_t the state of the mixture,

$$T = \begin{pmatrix} \phi & 0 \\ 0 & 1 \end{pmatrix}, \text{ and } H_t = \begin{pmatrix} 0 & \sigma_v & 0 \\ 0 & 0 & \sigma_\eta \delta_t \end{pmatrix}.$$

The initial distribution of the state vector is given by $\alpha_0 = 0$ and $\alpha_1 \sim N(0, P)$, where a large P indicates a diffuse prior. In our applications, we use $P = 1 \times 10^6$.

Model (A.1) is a linear Gaussian state space model conditional on the probability of shifts p, the realizations of the shifts δ and the mixture ω. The standard Kalman filtering algorithm to construct the likelihood function can thus be applied without modification. The detailed steps are as follows. As a matter of notation, let

$$a_t = E(\alpha_t | y_1, ..., y_{t-1})$$

and

$$P_t = E(a_t - \alpha_t)(a_t - \alpha_t).$$

Then for $t = 1, ..., n$, the Kalman filter uses the following recursions (see De Jong, 1991 and De Jong and Shephard, 1995):

$$\begin{aligned} e_t &= y_t - Z a_t, \\ D_t &= Z P_t Z' + G_t G_t', \\ K_t &= (T P_t Z') D_t^{-1}, \\ a_{t+1} &= T a_t + K_t e_t \end{aligned}$$

and

$$P_{t+1} = T P_t (T - K_t Z)' + H_t H_t',$$

with the initialization $a_1 = 0$ and $P_1 = P$. If there is no shift at t,

$$H_t = \begin{pmatrix} 0 & \sigma_v & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and if a shift is present,

$$H_t = \begin{pmatrix} 0 & \sigma_v & 0 \\ 0 & 0 & \sigma_\eta \end{pmatrix}.$$

Since the only matrix inversion involves D_t, which depends on the shifts only through P_t, no degeneracy is implied by the shifts.

Appendix 4: The simulation smoother

In this appendix, we discuss the algorithm to sample from

$$f(R, \alpha_1|\phi, \sigma_\upsilon, \sigma_\eta, \delta, \omega, y) = f(R|\phi, \sigma_\upsilon, \sigma_\eta, \delta, \omega, y) f(\alpha_1|\phi, \sigma_\upsilon, \sigma_\eta, \delta, \omega, y, R). \tag{A.2}$$

We use an extension of the simulation smoother developed by De Jong and Shephard (1995). Let e_t, D_t and K_t be the output obtained from applying the Kalman filter discussed in Appendix 3 and define F_t for $t = 1, ..., n$, by

$$F_t = \begin{cases} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \text{if } \delta_t = 1, \\ \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, & \text{if } \delta_t = 0. \end{cases} \tag{A.3}$$

Then, starting from $r_n = 0$ and $U_n = 0$, we apply the following backward recursion for $t = n, n-1, ..., 1$:

$$\begin{aligned} C_t &= F_t(I - H_t' U_t H_t) F_t', \qquad z_t \sim i.i.d.\ N(0, C_t), \\ V_t &= F_t H_t' U_t L_t, \\ r_{t-1} &= Z' D_t^{-1} e_t + L_t' r_t - V_t' C_t^{-1} z_t \end{aligned}$$

and

$$U_{t-1} = Z' D_t^{-1} Z + L_t' U_t L_t + V_t' C_t^{-1} V_t,$$

where $L_t = T - K_t Z$. We also draw $z_t^* \sim N(0, 1)$, independent of z_t if $\delta_t = 0$, and store

$$R_t = \begin{cases} F_t H_t' r_t + z_t, & \text{if } \delta_t = 1, \\ (F_t H_t' r_t + z_t, z_t^*)', & \text{if } \delta_t = 0. \end{cases} \tag{A.4}$$

Finally, we set α_1 to

$$\alpha_1 = P r_0. \tag{A.5}$$

Then, the values of $R = (R_1, ..., R_n)$ and α_1 obtained yield draws from the desired posterior distribution.

We shall prove the validity of this algorithm. The proof proceeds along the lines of De Jong and Shephard (1995, p. 348-349) and involves verifying that their conditions are satisfied in our context. All the subsequent arguments are understood to be conditional upon $(\phi, \sigma_\upsilon, \sigma_\eta, \delta, \omega)$ and we shall make explicit notice only when conditioning upon other variables. Let $\varphi_t = (\upsilon_t, \eta_t)'$ for $t = 0, 1, ..., n$, then $R = (\varphi_1, ..., \varphi_n)$. The target distributions to be generated are those of

$$\varphi_t|y, \varphi_{t+1}, ..., \varphi_n; \text{ for } t = 0, ..., n, \tag{A.6}$$

and

$$\alpha_1|y, \varphi_0, ..., \varphi_n \stackrel{d}{=} E(\varphi_0|y, \varphi_1, ..., \varphi_n). \tag{A.7}$$

We need to show that the first and second moments of (A.6) and (A.7) correspond to those of (A.4) and (A.5). Let $\xi_t = F_t u_t$ $(t = 0, 1, ..., n)$ with F_t defined by (A.3). If $\delta_t = 1$, then $\xi_t = \varphi_t$, otherwise ξ_t is a subset of φ_t.

Consider first the case with $t = n$. A shift occurring at time n affects the system starting at $n+1$, hence it is not identified based on a sample of size n. Indeed, both specifications of F_n lead to the same result. To see this, using $cov(e_n) = D_n^{-1}$ and $cov(u_n, e_n) = cov(u_n, G_n u_n) = G_n'$, we have,

$$\begin{aligned} E(\xi_n|y) &= E(\xi_n|e) = E(\xi_n|e_n) = cov(\xi_n, e_n)D_n^{-1}e_n = F_n cov(u_n, e_n)D_n^{-1}e_n = 0 \\ cov(\xi_n|y) &= E(\xi_n - E(\xi_n|y))(\xi_n - E(\xi_n|y))' = E(\xi_n\xi_n') = F_n F_n'. \end{aligned}$$

Hence, $\xi_n|y$ has the same distribution as $F_n H_n' r_n + z_n$ which implies that $\varphi_n|y$ has the same distribution as R_n. Now consider the case with $t < n$. Let

$$\begin{aligned} \epsilon_t &= \xi_t - E(\xi_t|y, \xi_{t+1}, ..., \xi_n), \text{ for } t = 0, ..., n-1, \\ \epsilon_n &= \xi_n - E(\xi_n|y). \end{aligned}$$

Then, using the relationship between φ_t and ξ_t, we have,

$$E(\varphi_t|y, \varphi_{t+1}, ..., \varphi_n) = \begin{cases} E(\xi_t|y, \varphi_{t+1}, ..., \varphi_n), & \text{if } \delta_t = 1, \\ (E(\xi_t|y, \varphi_{t+1}, ..., \varphi_n), 0)', & \text{if } \delta_t = 0. \end{cases}$$

In both cases, it is sufficient to study $E(\xi_t|y, \varphi_{t+1}, ..., \varphi_n)$. We have,

$$\begin{aligned} E(\xi_t|y, \varphi_{t+1}, ..., \varphi_n) &= E(\xi_t|y_1, e_1, ..., e_n, \varphi_{t+1}, ..., \varphi_n) \\ &= E(\xi_t|e_t, ..., e_n, \varphi_{t+1}, ..., \varphi_n) \\ &= E(\xi_t|e_t, ..., e_n, \xi_{t+1}, ..., \xi_n) \\ &= \sum_{s=t}^{n} E(\xi_t|e_s) + \sum_{s=t+1}^{n} E(\xi_t|\epsilon_s) \\ &= E(\xi_t|e_t) + \sum_{s=t+1}^{n} \{E(\xi_t|e_s) + E(\xi_t|\epsilon_s)\} \\ &= \sum_{s=t+1}^{n} \left\{cov(\xi_t, e_s)D_s^{-1}e_s + cov(\xi_t, \epsilon_s)var(\epsilon_s)^{-1}\epsilon_s\right\}, \end{aligned} \tag{A.8}$$

where the first equality holds due to the Gaussian assumption, the second is due to the uncorrelatedness of u_t, the third is due to that fact that if there is no shift at time s the value of η_s does not affect the likelihood function, the fourth is due to the fact that the

variables in the information sets are mutually orthogonal, the sixth is due to $E(\xi_t|e_t) = F_t cov(u_t, e_t) var(e_t)^{-1} e_t = (F_t G_t') var(e_t)^{-1} e_t = 0$ and the Gaussian assumption. Concerning the variance of $\varphi_t|y, \varphi_{t+1}, ..., \varphi_n$, we have,

$$var(\varphi_t|y, \varphi_{t+1}, ..., \varphi_n) = \begin{cases} var(\xi_t|y, \varphi_{t+1}, ..., \varphi_n), & \text{if } \delta_t = 1, \\ \text{Diag}(var(\xi_t|y, \varphi_{t+1}, ..., \varphi_n), 1)', & \text{if } \delta_t = 0. \end{cases}$$

Again, it is sufficient to study $var(\xi_t|y, \varphi_{t+1}, ..., \varphi_n)$. We have,

$$\begin{aligned} & var(\xi_t|y, \varphi_{t+1}, ..., \varphi_n) \\ = \; & cov(\xi_t - E(\xi_t|y, \varphi_{t+1}, ..., \varphi_n), \xi_t) \\ = \; & F_t F_t - cov(E(\xi_t|y, \varphi_{t+1}, ..., \varphi_n), \xi_t) \\ = \; & F_t F_t - \sum_{s=t+1}^{n} \left\{ cov(\xi_t, e_s) D_s^{-1} cov(e_s, \xi_t) + cov(\xi_t, \epsilon_s) var(\epsilon_s)^{-1} cov(\epsilon_s, \xi_t) \right\}. \quad \text{(A.9)} \end{aligned}$$

To study (A.8) and (A.9), the key is to verify the following results for $s > t$:

$$cov(e_s, \xi_t) = Z L_{s,t+1} H_t F_t', \quad \text{(A.10)}$$

$$cov(\epsilon_s, \xi_t) = -V_s L_{s,t+1} H_t F_t', \quad \text{(A.11)}$$

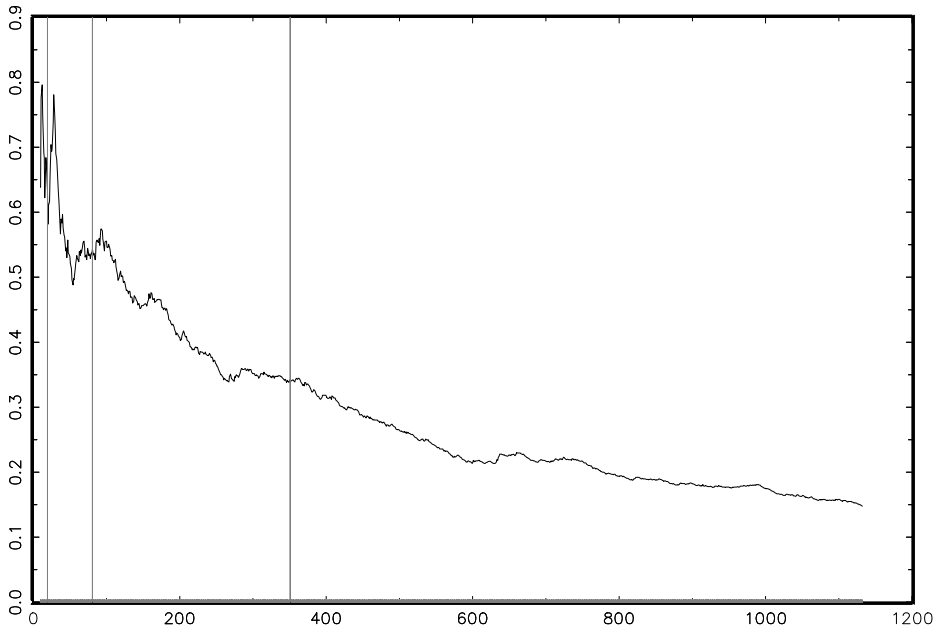
where the matrix $L_{s,j}$ is given by the product

$$L_{s,j} = L_{s-1} L_{s-2} ... L_j \quad \text{for } j = 1, ..., t-1,$$

and $L_{s,s}$ is the identity matrix ($s = 1, ..., n$). The result (A.10) was shown by Koopman (1993, p. 124) and (A.11) can be proved following De Jong and Shephard (1995, pp. 349). Substituting (A.10) and (A.11) into (A.8) and (A.9), it then follows that the first two moments of (A.6) and (A.7) correspond to those of (A.4) and (A.5). The proof is complete.

Figure 1: Spectral and time domain properties of log squared S&P500 daily returns: 1980:1-2005:12

(a) The log-periodogram estimate of d as a function of m



(b) The autocorrelation function

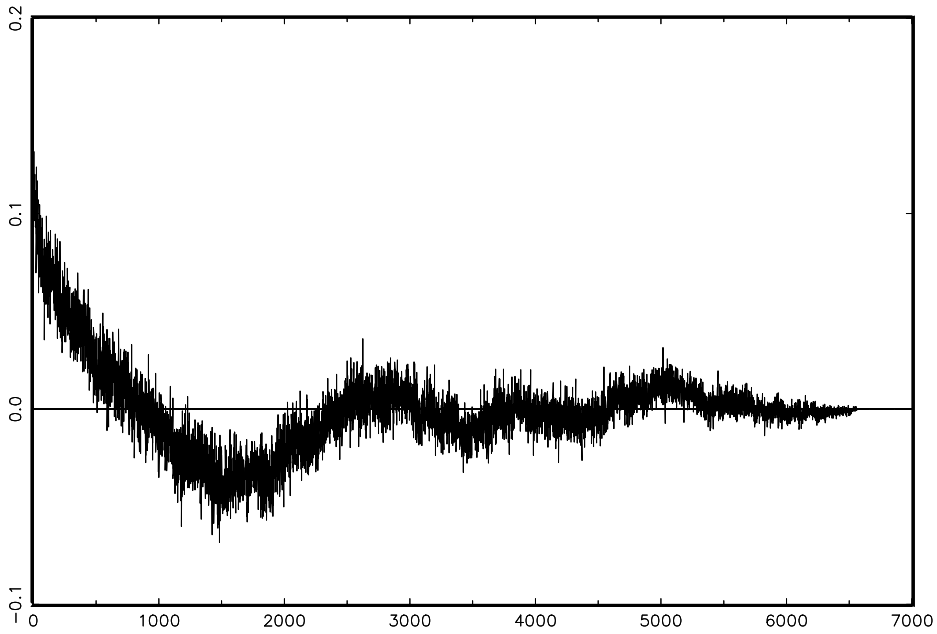
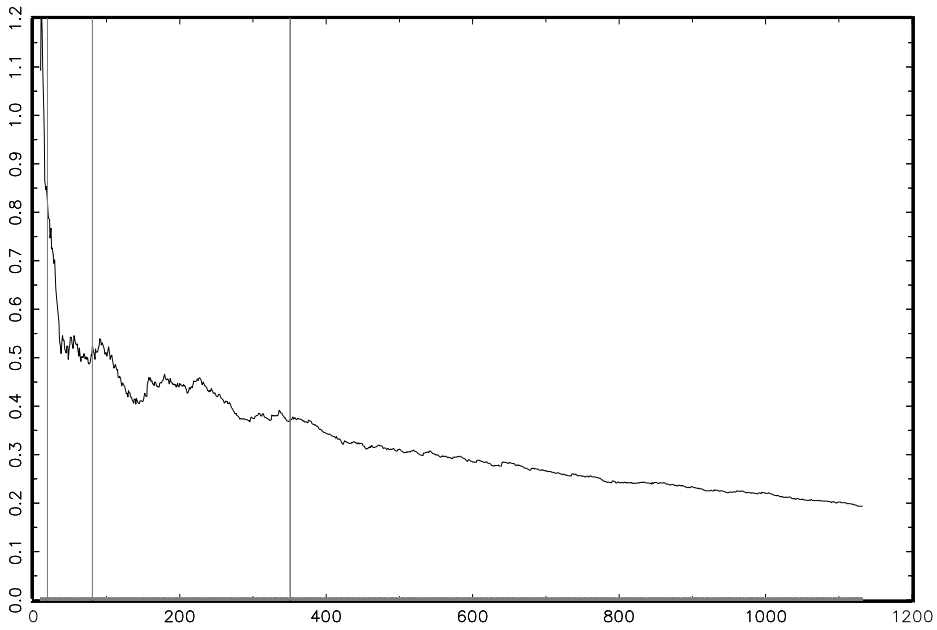


Figure 2: Spectral and time domain properties of log squared NASDAQ daily returns: 1980:1-2005:12

(a) The log-periodogram estimate of d as a function of m



(b) The autocorrelation function

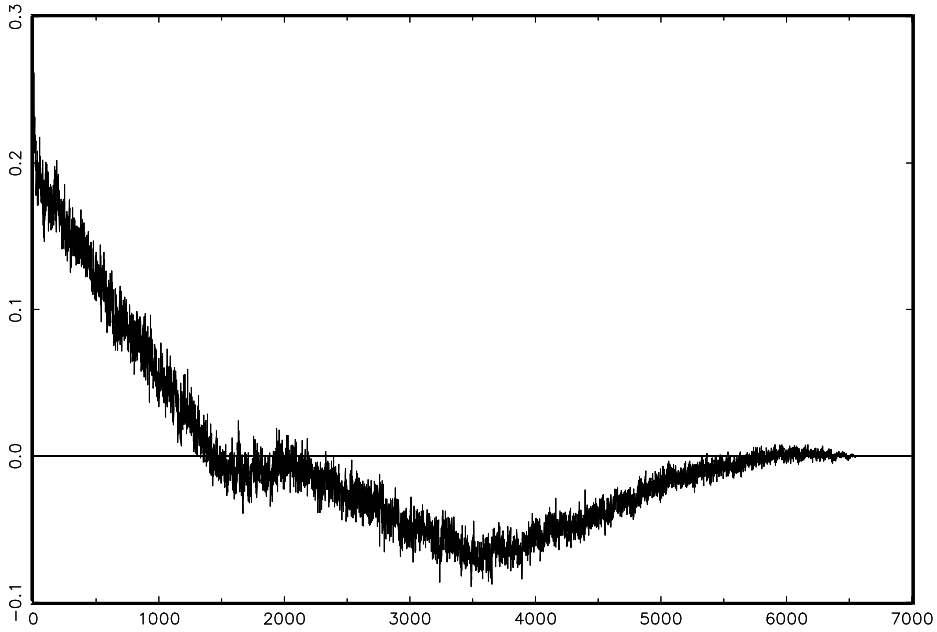


Figure 3: Simulation results with level shifts; n=2000

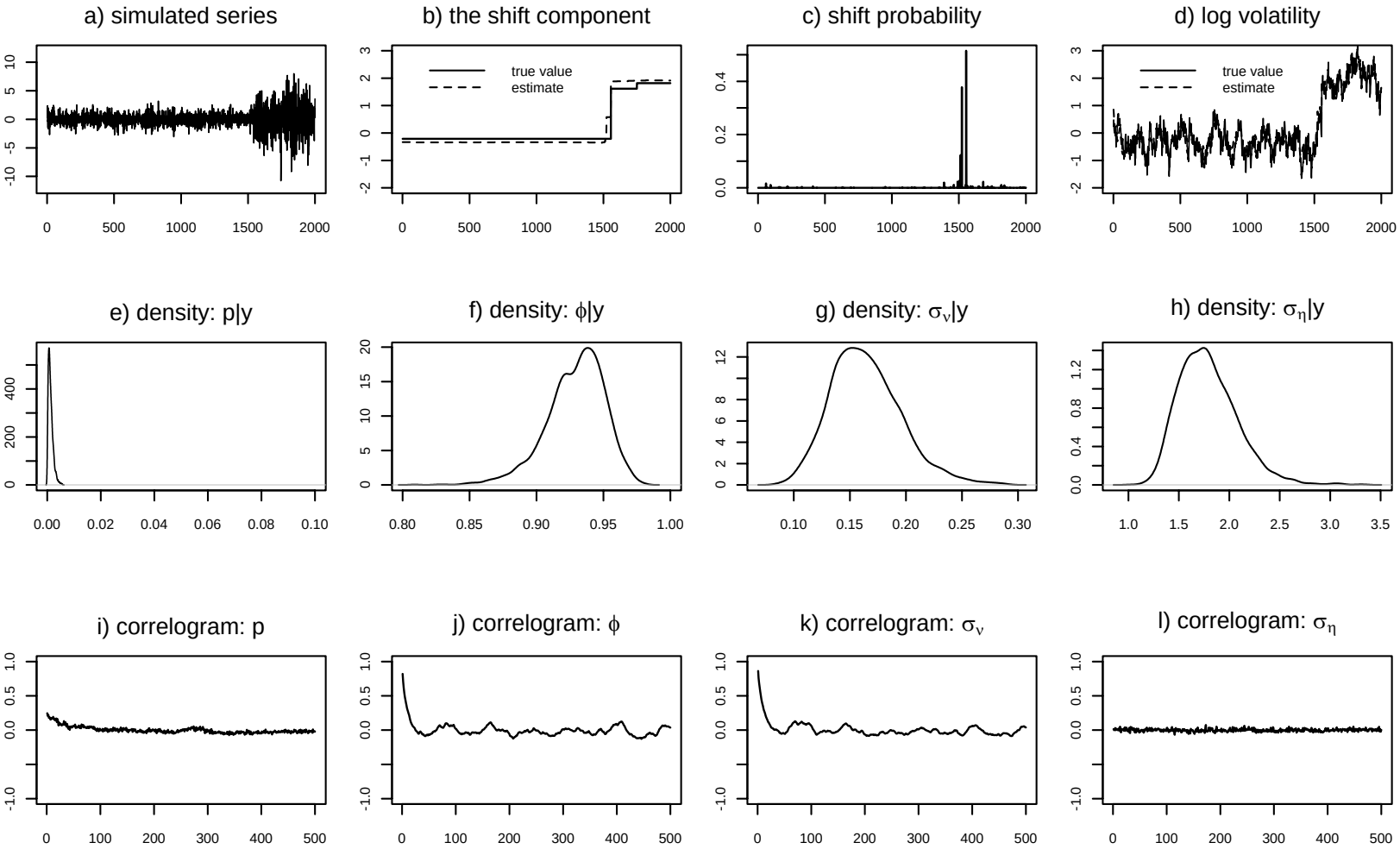


Figure 4: Simulation results with level shifts; n=4000

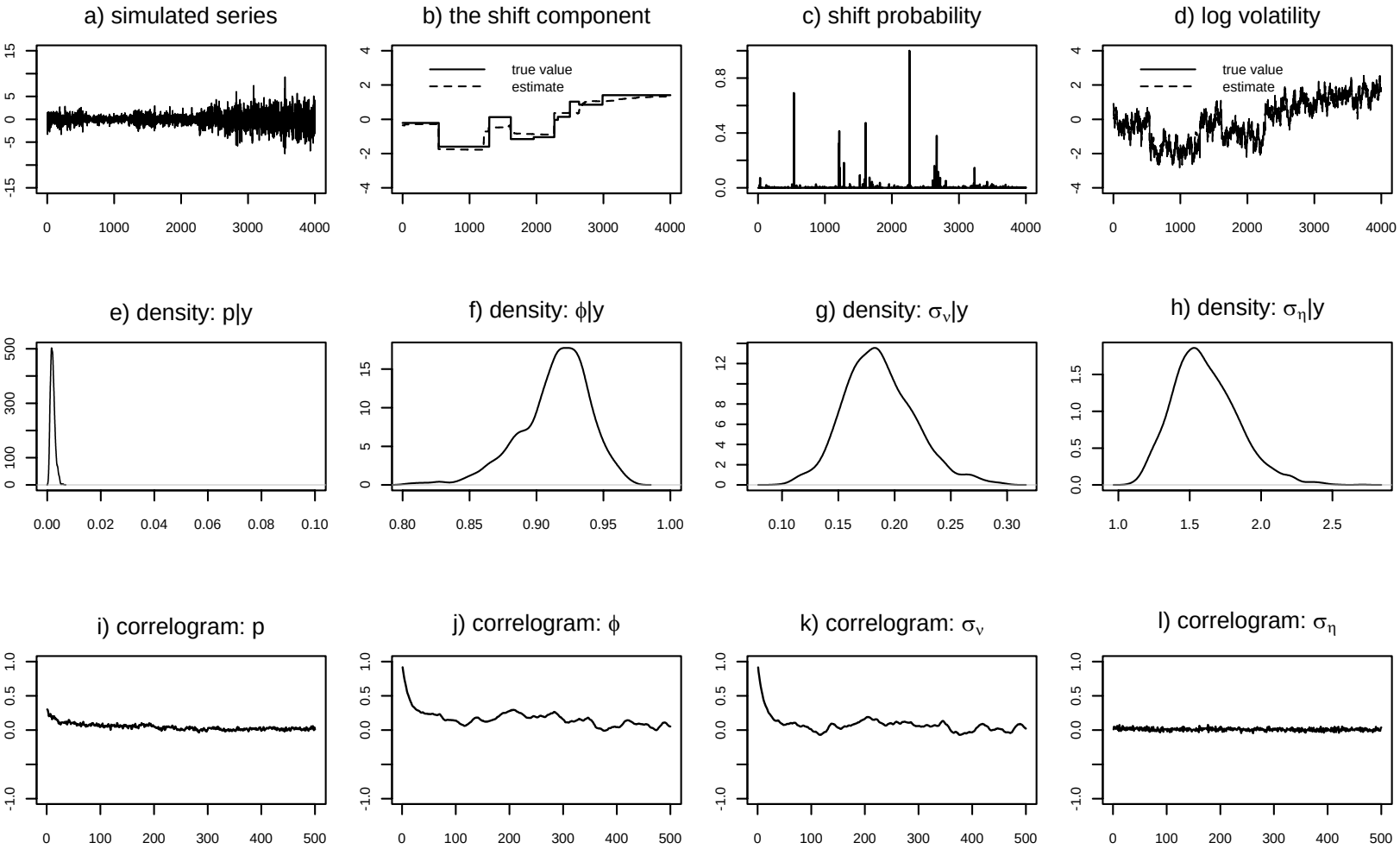


Figure 5: Simulation results without level shifts; n=2000

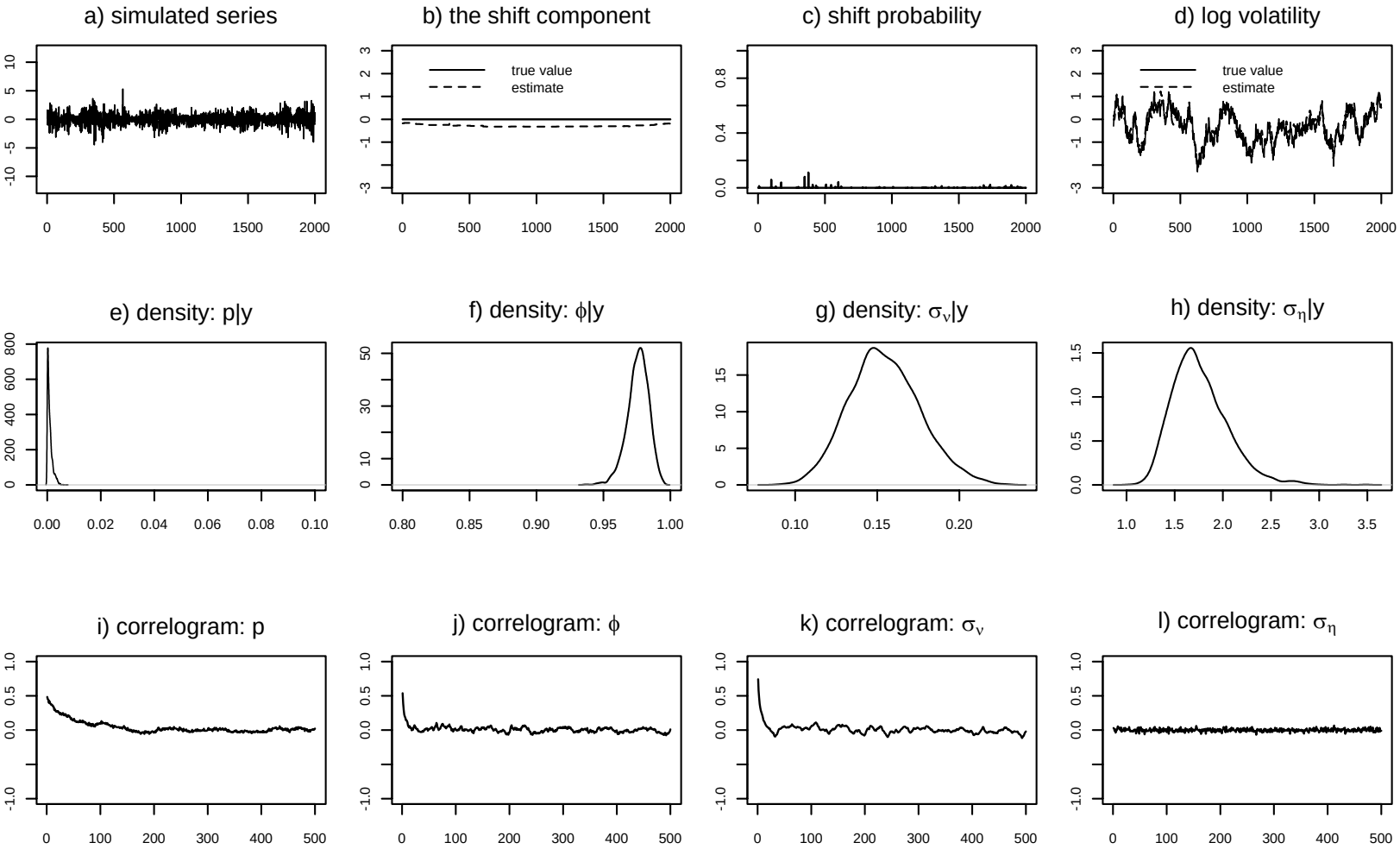


Figure 6: Simulation results without level shifts; n=4000

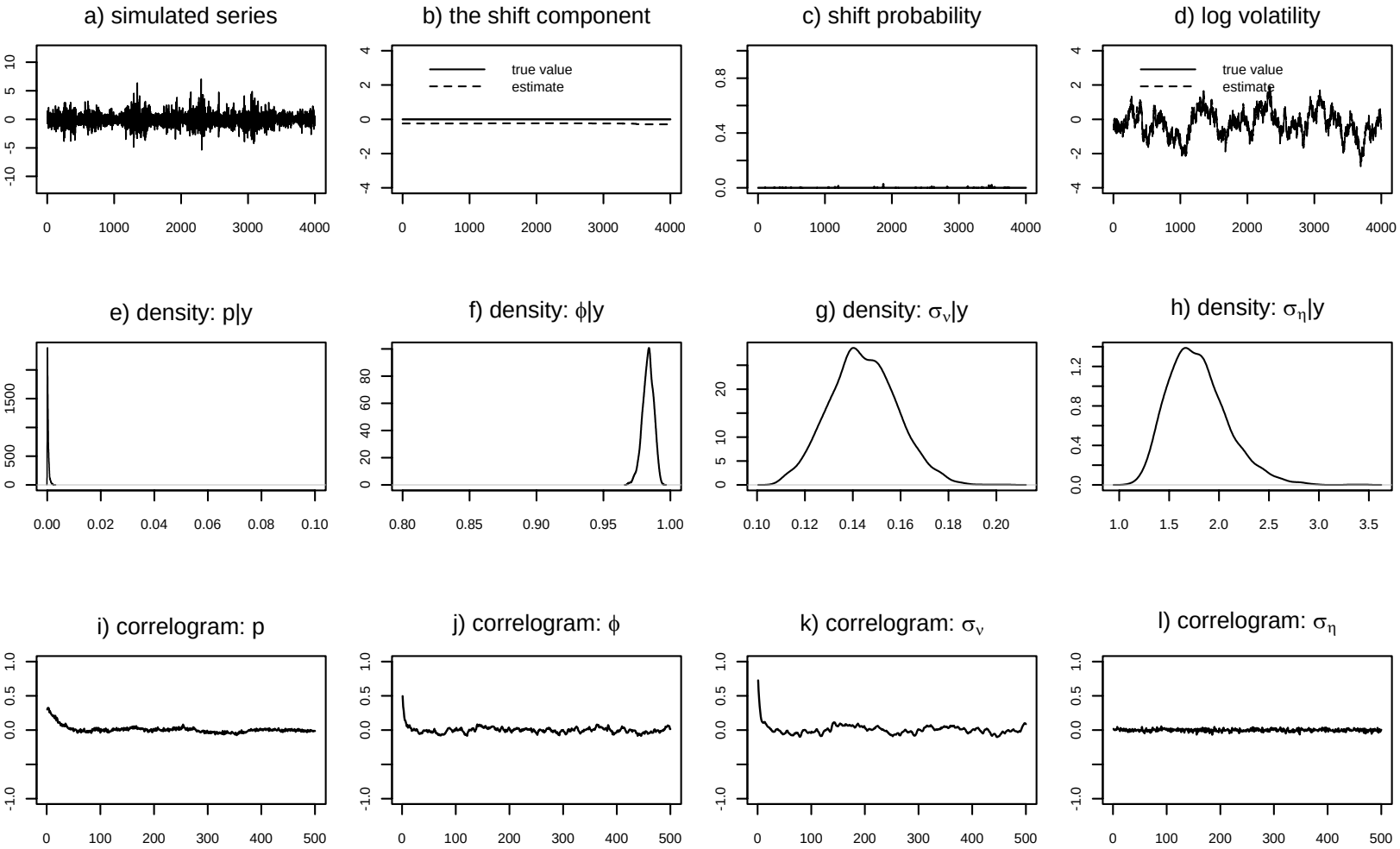
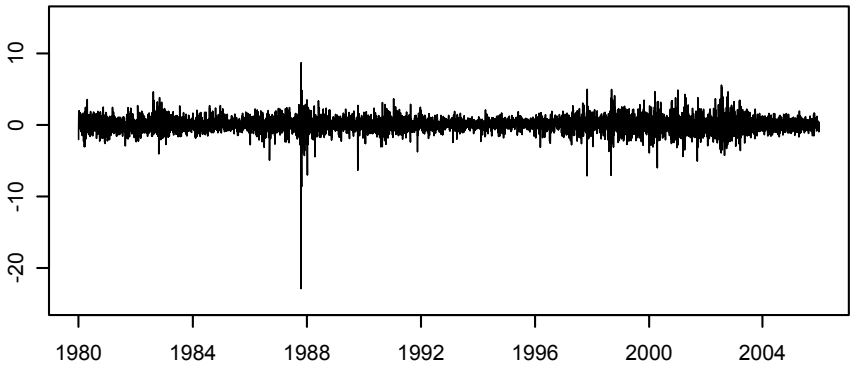
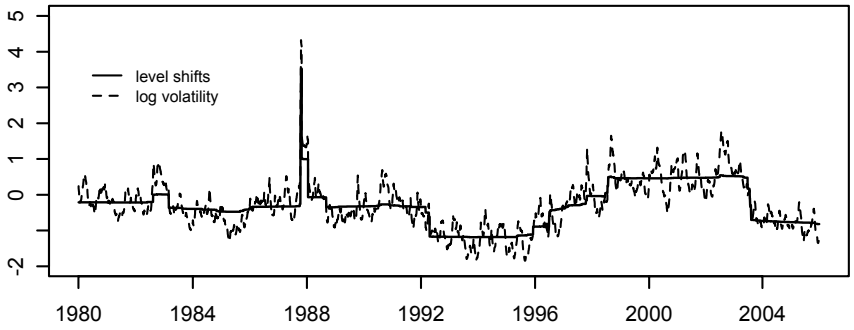


Figure 7: Results for S&P500 volatility

a) the return series



b) smoothed estimates of the level shift component and the log volatility



c) smoothed estimates of the probability of level shifts

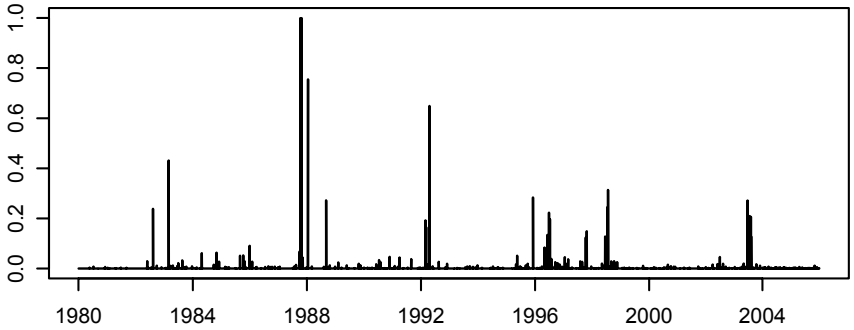


Figure 8: Results for S&P500 volatility (cont'd)

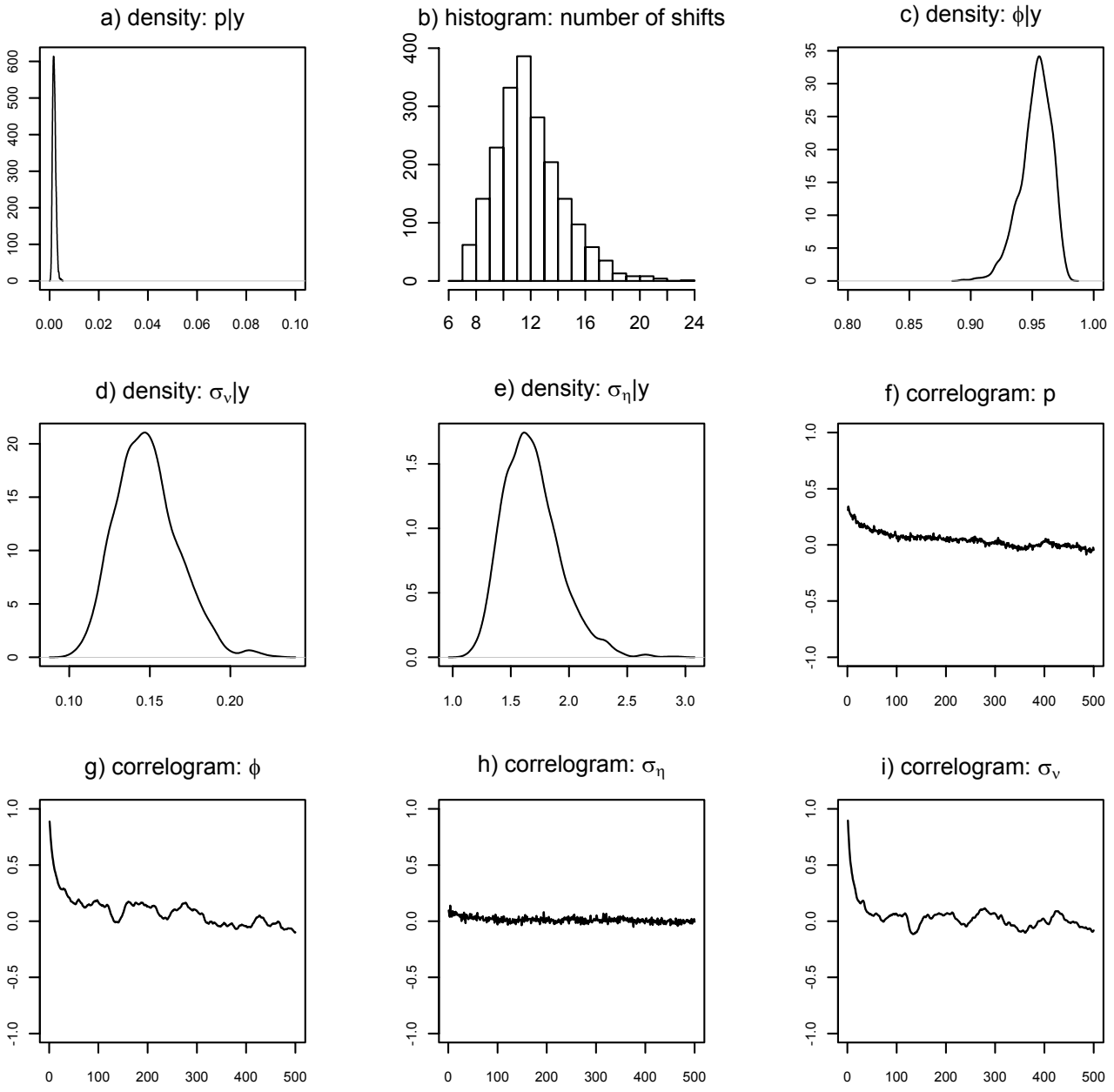
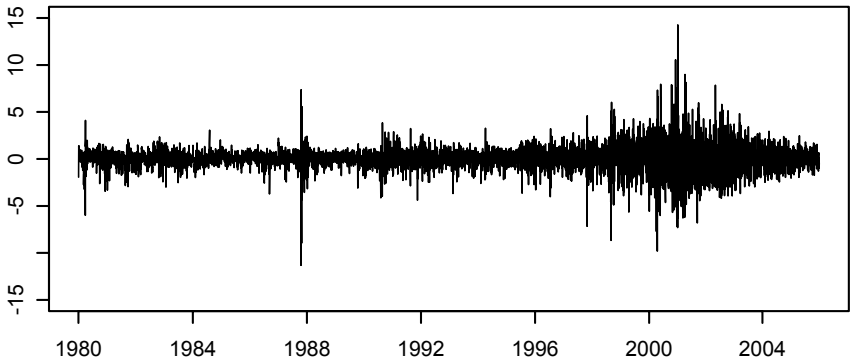
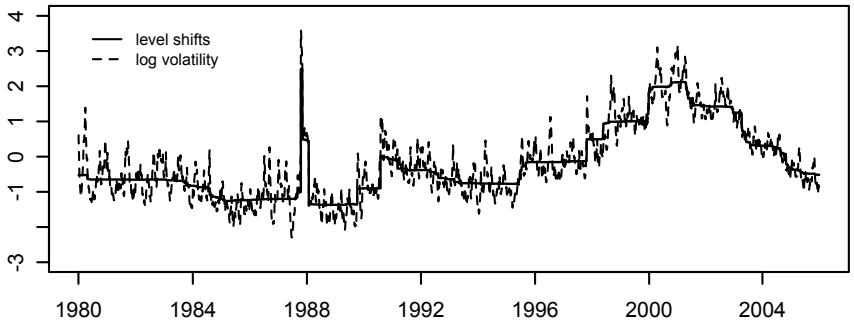


Figure 9: Results for NASDAQ volatility

a) the return series



b) smoothed estimates of the level shift component and the log volatility



c) smoothed estimates of the probability of level shifts

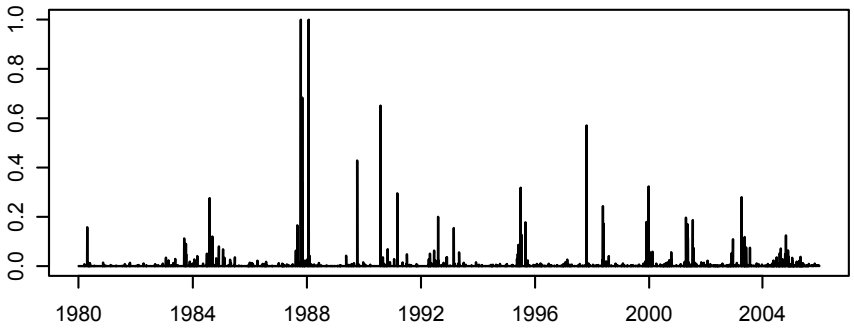


Figure 10: Results for NASDAQ volatility (cont'd)

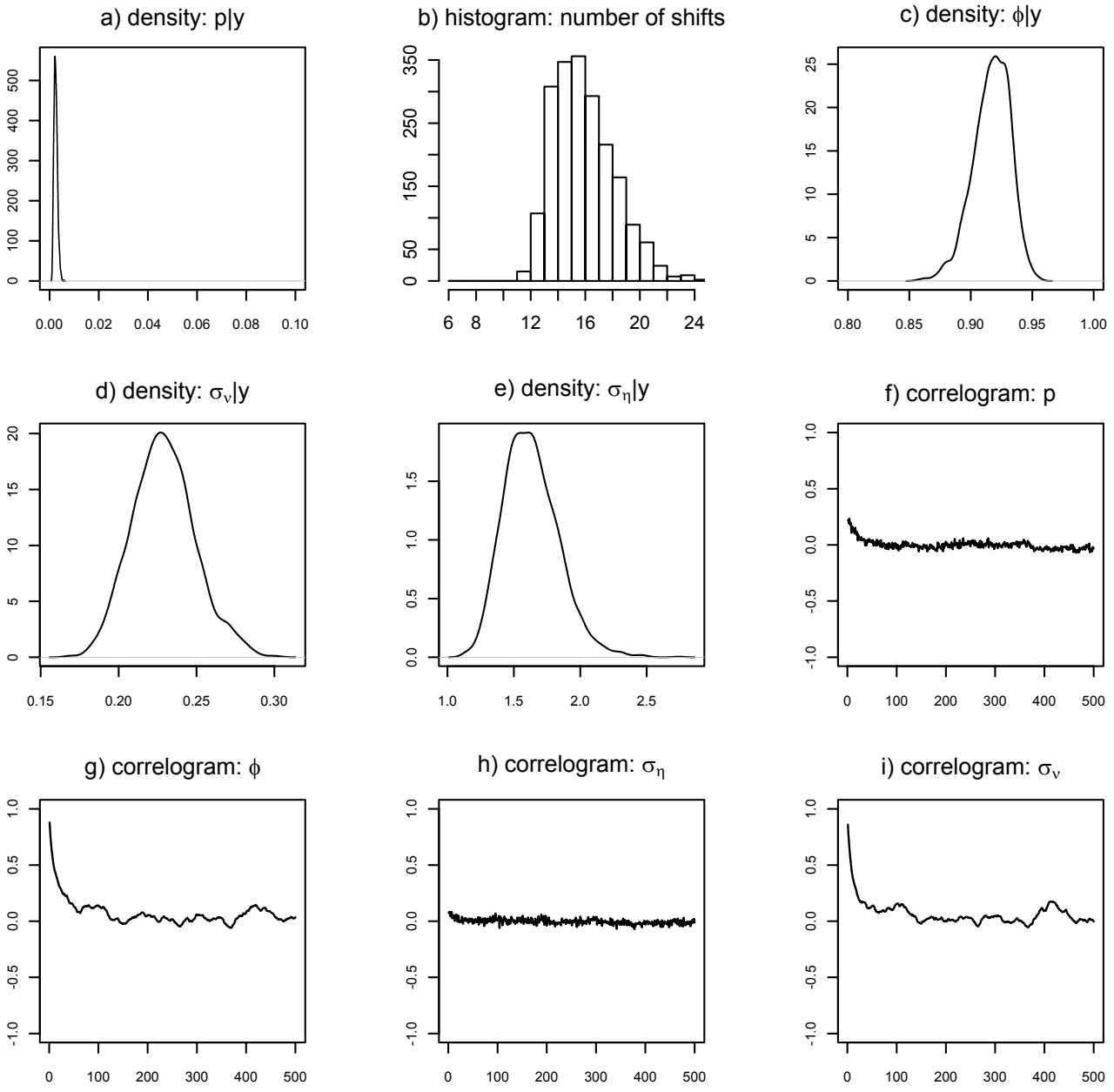


Figure 11: Diagnostic results for S&P500 volatility

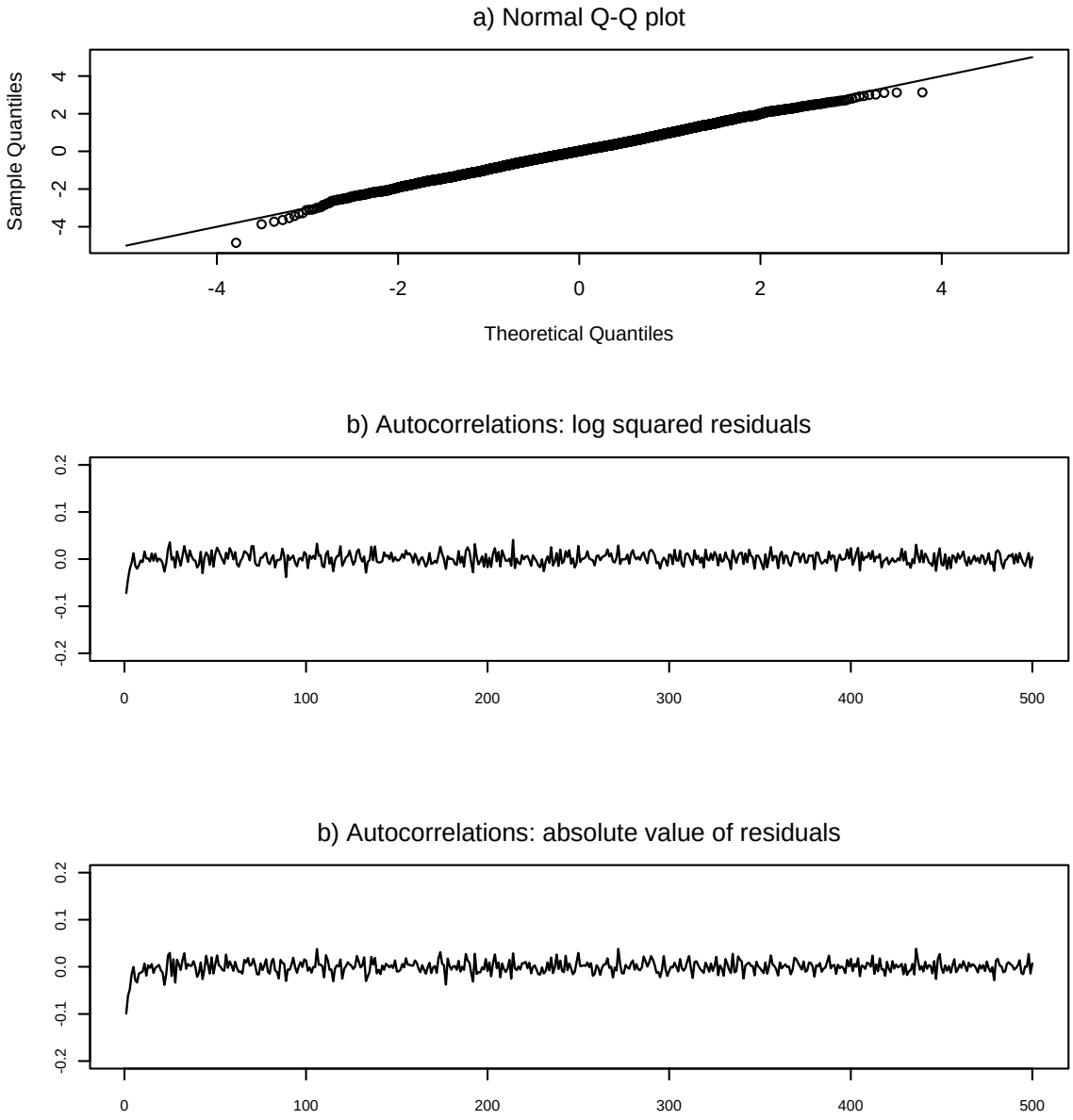
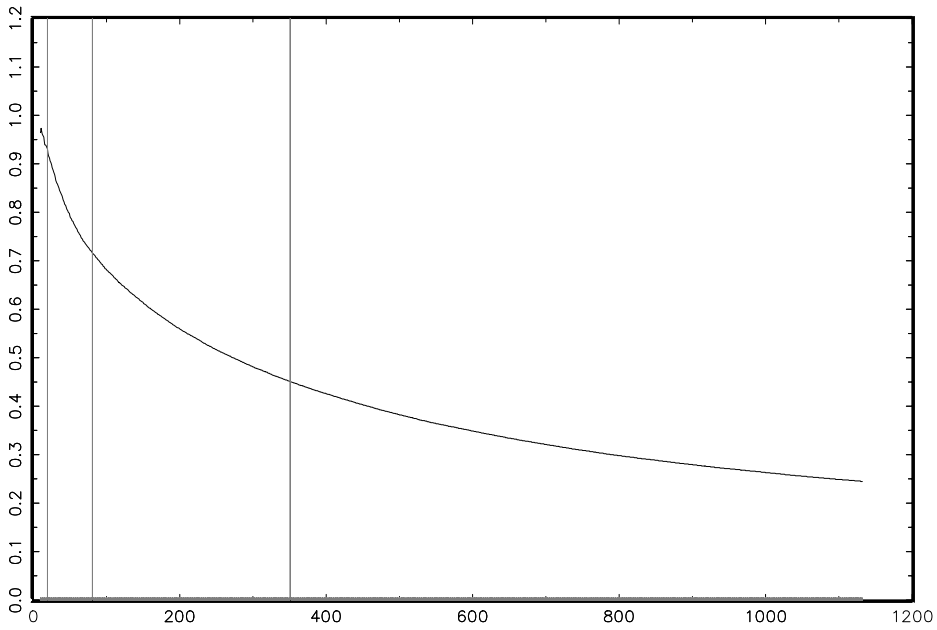


Figure 12: Spectral and time domain properties of the simulated returns (log-squared)

(a) The mean of the log-periodogram estimate of d as a function of m



(b) The mean of the autocorrelations

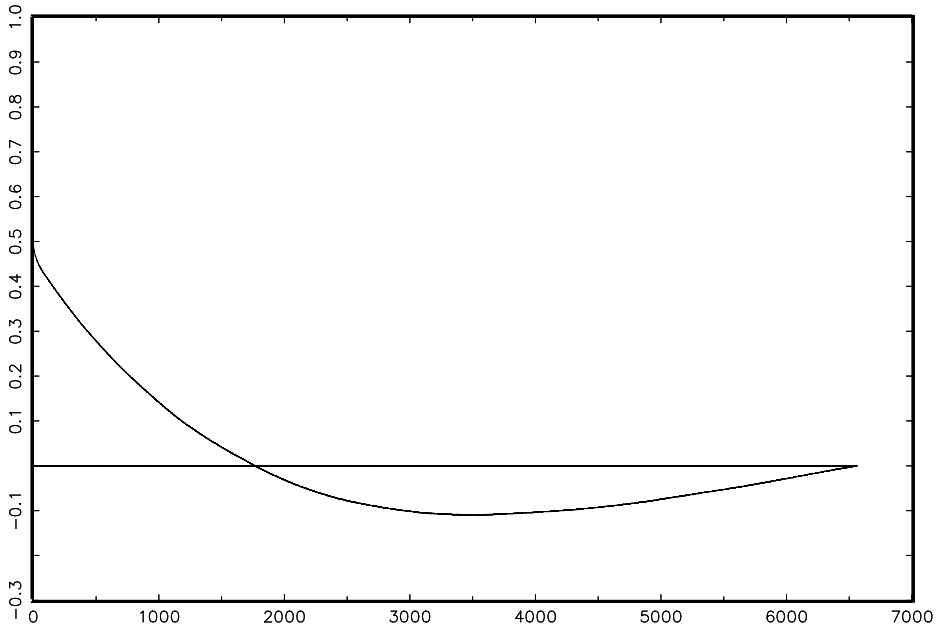
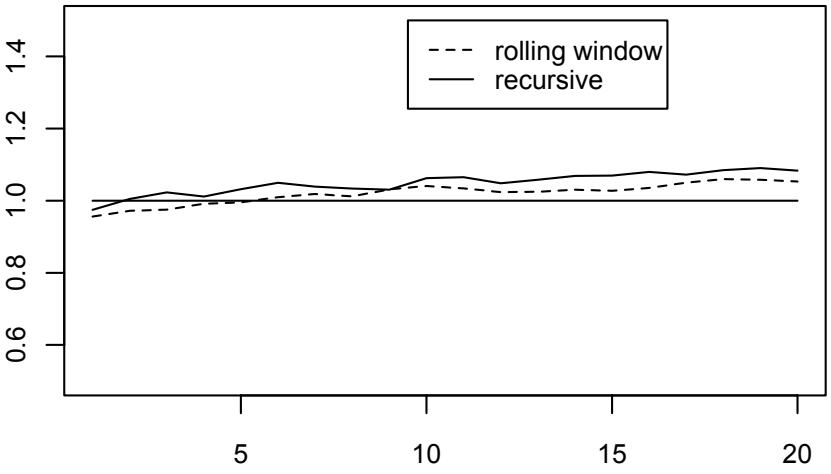
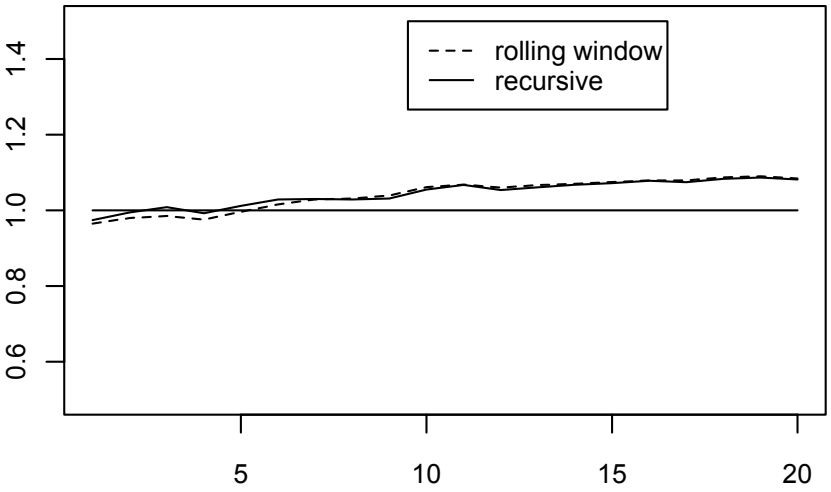


Figure 13: Forecasting comparisons: S&P 500

a) MSEs for the level shift model relative to SV



b) MSEs for the level shift model relative to FIGARCH(1,d,1)



(p-step ahead forecast horizon)

Figure 14: Forecasting comparisons: NASDAQ

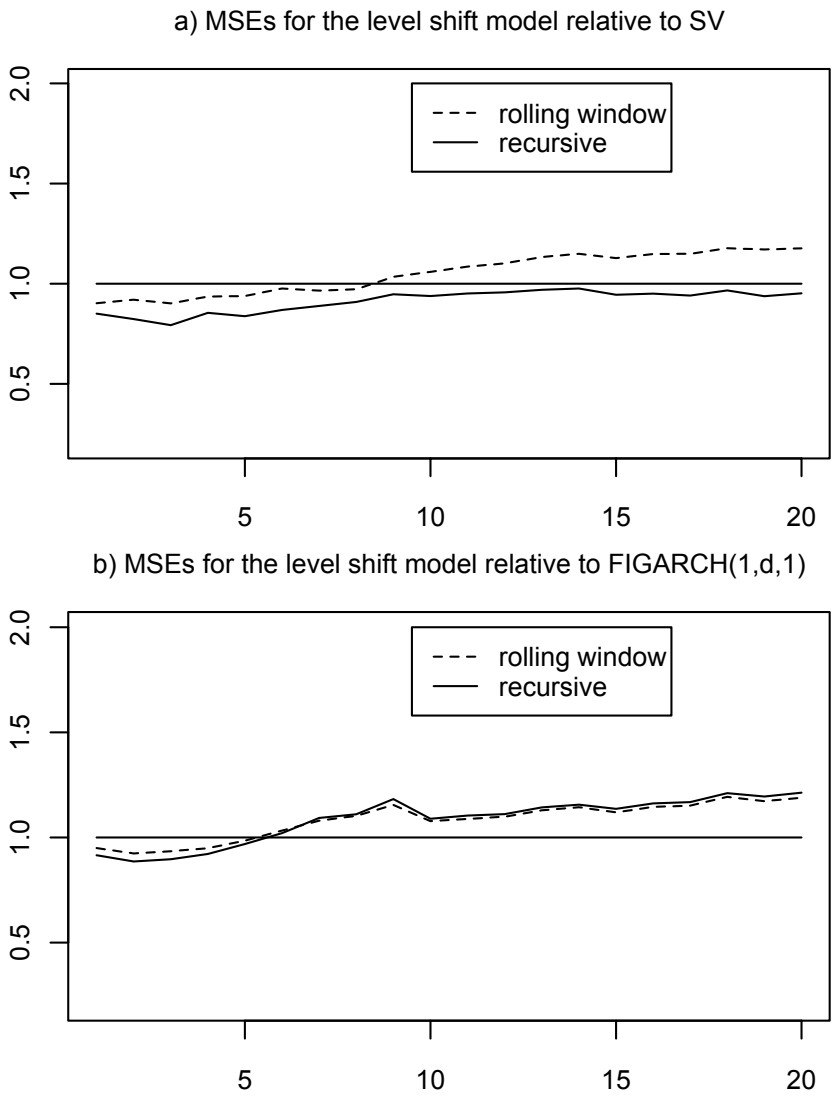


Table 1: Estimates from the forecasting regression (15).

		NASDAQ			S&P 500		
Horizon		b_o	b_1	$\bar{R}^2$	b_o	b_1	$\bar{R}^2$
$h=1$	LS	0.25 (−0.49,0.98)	0.83 (0.64,1.02)	**0.25**	0.05 (−0.36,0.47)	0.91 (0.60,1.22)	**0.13**
	SV-Rec	0.92 (0.17,1.67)	0.66 (0.45,0.87)	0.14	0.06 (−0.38,0.51)	0.95 (0.59,1.32)	0.10
	SV-Rol	0.70 (−0.04,1.45)	0.73 (0.53,0.93)	0.18	0.19 (−0.24,0.61)	0.80 (0.47,1.33)	0.09
	FIG-Rec	0.64 (−0.12,1.39)	0.79 (0.57,1.00)	0.18	0.07 (−0.37,0.51)	0.91 (0.56,1.26)	0.10
	FIG-Rol	0.52 (−0.22,1.27)	0.84 (0.63,1.06)	0.21	0.13 (−0.30,0.57)	0.86 (0.52,1.20)	0.10
$h=2$	LS	−0.54 (−1.83,0.75)	1.03 (0.87,1.20)	**0.40**	0.09 (−0.16,1.22)	0.53 (0.37,0.89)	**0.09**
	SV-Rec	0.82 (−0.54,2.18)	0.92 (0.73,1.11)	0.28	0.51 (−0.23,1.24)	0.68 (0.37,0.98)	0.07
	SV-Rol	0.24 (−1.06,1.54)	1.00 (0.82,1.18)	0.35	0.68 (−0.25,1.38)	0.57 (0.30,0.84)	0.07
	FIG-Rec	0.14 (−1.20,1.50)	1.05 (0.87,1.25)	0.32	0.54 (−0.19,1.27)	0.64 (0.35,0.93)	0.07
	FIG-Rol	−0.05 (−1.37,1.27)	1.11 (0.92,1.31)	0.36	0.63 (−0.08,1.34)	0.59 (0.31,0.87)	0.07
$h=3$	LS	−0.35 (−1.82,1.13)	0.93 (0.81,1.06)	**0.49**	0.63 (−0.16,1.42)	0.70 (0.50,0.90)	**0.17**
	SV-Rec	1.30 (−0.29,2.90)	0.88 (0.73,1.04)	0.35	0.46 (−0.36,1.30)	0.80 (0.57,1.03)	**0.17**
	SV-Rol	0.69 (−0.81,2.20)	0.92 (0.78,1.06)	0.43	0.75 (−0.06,1.55)	0.68 (0.47,0.89)	0.15
	FIG-Rec	0.32 (−1.24,1.87)	1.01 (0.85,1.16)	0.42	0.52 (−0.31,1.35)	0.75 (0.53,0.97)	0.16
	FIG-Rol	0.15 (−1.36,1.67)	1.05 (0.90,1.20)	0.45	0.68 (−0.14,1.49)	0.70 (0.49,0.92)	0.15
$h=4$	LS	−0.36 (−2.20,1.47)	0.95 (0.84,1.07)	**0.53**	0.16 (−0.74,1.05)	0.93 (0.76,1.10)	0.34
	SV-Rec	1.19 (−0.73,3.09)	1.01 (0.86,1.15)	0.45	−0.28 (−1.22,0.67)	1.10 (0.90,1.29)	**0.35**
	SV-Rol	0.79 (−1.05,2.62)	0.99 (0.86,1.11)	0.50	0.14 (−0.77,1.05)	0.96 (0.78,1.14)	0.33
	FIG-Rec	0.15 (−1.74,2.05)	1.08 (0.94,1.22)	0.49	−0.09 (−1.03,0.86)	1.01 (0.82,1.20)	0.33
	FIG-Rol	0.03 (−1.83,1.89)	1.12 (0.97,1.26)	0.51	0.11 (−0.83,1.04)	0.97 (0.78,1.15)	0.32
$h=5$	LS	−0.25 (−2.17,1.68)	0.92 (0.82,1.02)	**0.60**	0.30 (−0.67,1.26)	0.90 (0.75,1.04)	0.39
	SV-Rec	1.45 (−0.60,3.49)	1.01 (0.89,1.14)	0.52	−0.26 (−1.27,0.75)	1.06 (0.89,1.23)	**0.40**
	SV-Rol	1.10 (−0.83,3.03)	0.97 (0.86,1.08)	0.57	0.26 (−0.72,1.24)	0.93 (0.77,1.08)	0.38
	FIG-Rec	0.04 (−0.92,1.99)	1.09 (0.97,1.21)	0.59	−0.05 (−1.06,0.95)	0.98 (0.82,1.15)	0.39
	FIG-Rol	−0.01 (−1.94,1.92)	1.11 (0.99,1.23)	0.59	0.14 (−0.86,1.14)	0.95 (0.79,1.11)	0.38

Table 2: Estimates from the forecasting regression (16).

		NASDAQ				S&P 500			
Horizon	LS versus	b_o	b_1	b_2	$\bar{R}^2$	b_o	b_1	b_2	$\bar{R}^2$
$h=1$	SV-Rec	0.22 (−0.52,0.95)	1.27 (0.87,1.67)	−0.51 (−0.93,−0.08)	0.26	0.25 (−0.20,0.70)	2.52 (1.02,4.03)	−1.90 (−3.64,−0.16)	0.14
	SV-Rol	0.22 (−0.52,0.95)	1.27 (0.76,1.78)	−0.49 (−1.01,0.04)	0.26	0.15 (−0.26,0.56)	3.43 (2.01,4.86)	−2.68 (−4.15,−1.20)	0.17
	FIG-Rec	0.15 (−0.57,0.87)	2.20 (1.49,2.92)	−1.57 (−2.36,−0.78)	0.29	0.31 (−0.14,0.75)	3.43 (1.63,5.25)	−2.83 (−4.83,−0.83)	0.15
	FIG-Rol	0.15 (−0.58,0.88)	2.04 (1.17,2.91)	−1.38 (−2.35,−0.40)	0.27	0.28 (−0.13,0.70)	4.51 (2.70,6.32)	−3.83 (−5.88,−1.98)	0.18
$h=2$	SV-Rec	−0.57 (−1.85,0.72)	1.34 (0.97,1.70)	−0.36 (−0.76,0.03)	0.41	0.71 (−0.05,1.46)	1.33 (0.08,2.57)	−0.93 (−2.28,0.62)	0.09
	SV-Rol	−0.54 (−1.84,0.75)	1.05 (0.59,1.51)	−0.02 (−0.50,0.46)	0.40	0.64 (−0.05,1.33)	1.96 (0.77,3.15)	−1.42 (−2.65,−0.18)	0.10
	FIG-Rec	−0.64 (−1.89,0.61)	2.46 (1.77,3.16)	−1.67 (−2.46,−0.88)	0.44	0.84 (0.10,1.60)	2.25 (0.62,3.87)	−1.81 (−3.60,−0.02)	0.10
	FIG-Rol	−0.66 (−1.94,0.62)	2.10 (1.26,2.93)	−1.25 (−2.20,−0.29)	0.42	0.86 (0.16,1.56)	3.48 (1.80,5.17)	−3.10 (−4.90,−1.29)	0.13
$h=3$	SV-Rec	−0.36 (−1.83,1.11)	1.15 (0.86,1.43)	−0.26 (−0.58,0.06)	0.49	0.61 (−0.26,1.47)	0.63 (−0.32,1.57)	0.08 (−1.01,1.18)	0.17
	SV-Rol	−0.35 (−1.83,1.13)	0.94 (0.58,1.30)	−0.01 (−0.39,0.37)	0.48	0.71 (−0.09,1.50)	1.28 (0.39,2.18)	−0.62 (−1.56,0.31)	0.18
	FIG-Rec	−0.38 (−1.82,1.07)	1.82 (1.26,2.39)	−1.06 (−1.71,−0.40)	0.51	0.75 (−0.12,1.62)	1.11 (−0.21,2.44)	−0.46 (−1.92,1.00)	0.17
	FIG-Rol	−0.41 (−1.88,1.06)	1.58 (0.92,2.23)	−0.77 (−1.54,0.01)	0.49	0.89 (0.07,1.70)	2.30 (0.93,3.66)	−1.73 (−3.18,−0.27)	0.19
$h=4$	SV-Rec	−0.36 (−2.20,1.47)	0.86 (0.58,1.13)	0.12 (−0.18,0.44)	0.53	−0.20 (−1.18,0.78)	0.24 (−0.55,1.04)	0.82 (−0.10,1.74)	0.35
	SV-Rol	−0.29 (−2.13,1.55)	0.74 (0.40,1.08)	0.24 (−0.12,0.61)	0.53	0.11 (−0.80,1.02)	0.67 (−0.09,1.43)	0.28 (−0.51,1.07)	0.34
	FIG-Rec	−0.35 (−2.19,1.48)	1.23 (0.69,1.79)	−0.34 (−0.99,0.30)	0.53	0.16 (−0.83,1.15)	0.94 (−0.23,2.11)	−0.01 (−1.30,1.27)	0.34
	FIG-Rol	−0.36 (−2.21,1.48)	0.99 (0.38,1.62)	−0.05 (−0.80,0.70)	0.53	0.33 (−0.60,1.26)	1.73 (0.54,2.92)	−0.87 (−2.14,0.41)	0.34
$h=5$	SV-Rec	−0.26 (−2.19,1.67)	0.80 (0.57,1.04)	0.15 (−0.12,0.43)	0.60	−0.21 (−1.26,0.84)	0.14 (−0.53,0.81)	0.91 (0.12,1.69)	0.40
	SV-Rol	−0.14 (−2.06,1.78)	0.68 (0.40,0.97)	0.28 (−0.03,0.59)	0.61	0.23 (−0.75,1.21)	0.61 (−0.04,1.25)	0.31 (−0.36,0.98)	0.39
	FIG-Rec	−0.27 (−2.20,1.66)	0.73 (0.26,1.20)	0.24 (−0.32,0.80)	0.60	0.08 (−0.99,1.14)	0.40 (−0.62,1.42)	0.55 (−0.57,1.67)	0.39
	FIG-Rol	−0.25 (−2.18,1.67)	0.58 (0.05,1.10)	0.43 (−0.21,1.06)	0.60	0.29 (−0.72,1.30)	0.89 (−0.14,1.92)	0.01 (−1.10,1.11)	0.39