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

Zhongjun Qu† Boston University

Pierre Perron‡ Boston University

November 1, 2007; This version: September 15, 2012.

**Abstract**

This paper proposes a framework for the modeling, inference and forecasting of volatility in the presence of level shifts of unknown timing, magnitude and frequency. First, we consider a stochastic volatility model comprising both a level shift and a short-memory component, with the former modeled as a compound binomial process and the latter as an AR(1). Next, we adopt a Bayesian approach for inference and develop algorithms to obtain posterior distributions of the parameters and the two latent components. Then, we apply the model to daily S&P 500 and NASDAQ returns over the period 1980.1-2010.12. The results show that although the occurrence of a level shift is rare, about once every two years, this component clearly contributes most to the variation in the volatility. The half-life of a typical shock from the AR(1) component is short, on average between 9 and 15 days. Interestingly, isolating the level shift component from the overall volatility reveals a stronger relationship between volatility and business cycle movements. Although the paper focuses on daily index returns, the methods developed can potentially be used to study the low frequency variation in realized volatility or the volatility of other financial or macroeconomic variables.

**Keywords**: Low frequency volatility; Long-memory; Structural change; State-space models.
**JEL Classification**: C11, C12, C53, G12.

*Pierre Perron acknowledges financial support from the National Science Foundation under Grant SES-0649350. We are grateful to Adam McCloskey for detailed comments and the co-editor Jianqing Fan and two anonymous referees for constructive comments and suggestions.

†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. For extensive reviews and collected works, see Engle (1995) and Shephard (2005). In their standard forms, the ensuing volatility processes are stationary and weakly dependent with autocorrelations that decrease exponentially. This contrasts sharply 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. Baillie, Bollerslev and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996) considered fractionally integrated GARCH and EGARCH models. Breidt, Crato and De Lima (1998) and Harvey (1998) proposed long-memory SV models, in which log-volatility is modeled as a fractionally integrated latent process.

Level shifts has been advanced as a possible explanation to the strong persistence in volatility. Using daily data on S&P 500 returns, Granger and Hyung (2004) documented 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 level shifts in unconditional variance. Perron and Qu (2010) analyzed the spectral domain properties of a stationary short-memory process affected by random level shifts. When applied to daily S&P 500 log absolute returns over the period 1928-2002, the level shift model explains the path of the log periodogram estimates as a function of the number of frequency ordinates used. The only type of long-memory process able to match such a feature is a perturbed long-memory process with a long-memory parameter well above 0.5 (e.g., around 0.8). Such infinite variance processes are unlikely candidates for a useful model of volatility. It is worthwhile to formally incorporate level shifts into volatility models and to develop methods for inference and forecasting. This paper addresses these issues.

We propose a simple stochastic volatility model that allows for level shifts of unknown timing, magnitude and frequency and confront it with S&P 500 and NASDAQ return volatility. Let $x_t$ denote the mean corrected return process. The proposed model is given by

$$x_t = \exp(h_t/2 + \mu_t/2)\varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim i.i.d.\ N(0,1)$. The first process $h_t$ is a stationary AR(1) component. It is common to standard SV models and is intended to capture short run dynamics. The second process $\mu_t$ is a compound binomial process that models random level shifts. More specifically, $\mu_{t+1} = \mu_t + \delta_t \sigma_\eta \eta_t$,

where $\eta_t \sim i.i.d.\ N(0,1)$ and $\delta_t$ is a sequence of independent Bernoulli random variables taking value 1 with unknown probability $p$. This 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. In the model, shocks to $h_t$ die out quickly while shocks to $\mu_t$ have a long lasting effect, remaining in the system until the next level shift occurs. This feature is useful for disentangling the effect of rare but influential events from that of frequent but transitory shocks on the dynamics of volatility.

The model (1) is related to three families of models proposed in the literature. The first is the Markov switching ARCH model of Hamilton and Susmel (1994). A key difference is that, in (1), the number of regimes is not specified a priori, but rather determined endogenously by the sequence $\{\delta_t, \eta_t\}$. The second is the Spline-GARCH model of Engle and Rangel (2008), where the volatility is also composed of two components but with the low frequency component modeled as an exponential quadratic spline. In our case, this component follows a stochastic process, opening the possibility of procedures for filtering, smoothing and out-of-sample forecasting. The third family includes models considered in Duffie, Pan and Singleton (2000), Eraker, Johannes and Polson (2003) and Eraker (2004), where jumps are incorporated into the SV model to capture clustered extreme movements in returns. In their models, the jumps have a short lasting effect and accordingly they generate a volatility process that is stationary while ours does not. As we shall see, this difference is crucial to generating features akin to a long-memory process.

We propose methods for inference and forecasting. For inference, we adopt a Bayesian approach with data augmentation following the work of Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998). The method addresses two difficulties: (1) as common to all SV models, the processes determining volatility are latent, and (2) $\mu_t$ is a compound binomial process making both the number and locations of shifts unknown. We provide algorithms to obtain the posterior distributions of all parameters and estimates of the processes $h_t$ and $\mu_t$. We also present algorithms for filtering and forecasting. They use the technique of particle filtering and are relatively straightforward to implement.

The model is then applied to the volatility of daily S&P 500 and NASDAQ returns over the period 1980.1-2010.12, which covers two market crashes in 1987 and 2008. The choice of priors pertaining to the level shift component reflects the belief that shifts are rare and their magnitudes are large. 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. Importantly, although

rare, the level shift component clearly contributes more to total variation in the volatility process than the AR(1) component. Second, the model explains the path of the log periodogram estimates as a function of the number of frequency ordinates used, as documented by Perron and Qu (2010). Third, the level-shift component is closely linked to business cycle fluctuations while the AR(1) component is not. After isolating the former component from the overall volatility, we observe a stronger volatility-business cycle relationship.

The paper contributes to the literature on modeling the strong persistence in the volatility of stock returns and other financial/macroeconomic variables. It fills a gap by providing a unified framework for modeling, inference and forecasting of volatility in the presence of regime change of unknown 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. In this paper, level shifts affects the volatility with the latter itself being a latent stochastic process. This generates technical difficulties and our analysis provides technical solutions to address 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 the stochastic volatility model with random level shifts. Section 4 presents the Bayesian inference procedure. Section 5 discusses methods for filtering and forecasting. Section 6 presents the results from empirical applications related to the volatility of daily S&P 500 and NASDAQ returns. Section 7 offers brief concluding remarks and some technical derivations are contained in appendices.

## 2 Preliminaries: long memory features in return volatility

Let $z_t$ be a stationary time series with autocorrelation function $\gamma_z(\tau)$ at delay $\tau$. It is said to have long-memory (Beran, 1994) if $\gamma_z(\tau) = g(\tau)\tau^{2d-1}$ as $\tau \to \infty$, where $d > 0$ and $g(\tau)$ is a slowly varying function as $\tau \to \infty$. For $d \in (0, 1/2)$, this implies that the $\gamma_z(\tau)$ decreases to zero at a hyperbolic rate, in contrast to the fast geometric rate that applies to a short-memory process.

Several papers have reported that transformations of stock returns, say $x_t$, of the form $|x_t|^\theta$ 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). The often cited generating mechanisms for long-memory are contemporaneous aggregation (Granger, 1980 and Andersen and Bollerslev, 1997) and shocks of different durations (Parke, 1999). Level

shifts can also produce features akin to a long-memory process. 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 (Diebold and Inoue, 2001). Mikosch and Stărică (2004) and Lamoureux and Lastrapes (1990) provided illustrations of such a feature in the context of volatility persistence. Smith (2005) derived the bias of the log periodogram estimator of the memory parameter when the underlying process is white noise with a slowly varying mean.

More closely related to our work is Perron and Qu (2010), who studied the spectral and time domain features of the following process

$$z_t = c + u_t + v_t \quad \text{with} \quad u_t = u_{t-1} + \pi_t \eta_t, \tag{2}$$

where $z_t$ can be thought as the logarithm of squared asset returns, $c$ 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 $\pi_t$ is a Bernoulli random variable that takes value 1 with probability $p_n$. They used $p_n = p/n$ to model rare shifts and obtained the following results. 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 stationary long-memory process 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 (say $m$) 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 $v_t$ becomes more important. They found this feature is empirically present when using transformations of S&P 500 daily returns covering the period 1928-2002. Here, we replicate their findings using the sample period 1980.1-2010.12 which is of interest in this paper. Figures 1 depicts the path of the log periodogram estimates, for the logarithm of squared S&P 500 and NASDAQ returns, where the three vertical lines correspond to $m = T^{1/3}, T^{1/2}$ and $T^{2/3}$, respectively. They suggest that the pattern is consistent with the level shift model (2). To the best of our knowledge, the only type of long-memory process able to match such features is a perturbed long-memory process with a long-memory parameter well above 0.5 (e.g., around 0.8), see Perron and Qu (2010). Such infinite variance processes are unlikely candidates for a useful model of volatility. Hence, it is worthwhile to formally incorporate level shifts into volatility models and to develop methods for inference and forecasting.

## 3 A stochastic volatility model with random level shifts

Let $\{x_t\}_{t=1}^n$ be a demeaned return process. The standard stochastic volatility (SV) model is

$$x_t = \exp(h_t/2)\varepsilon_t \quad \text{with} \quad h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\upsilon \upsilon_t, \tag{3}$$

where $h_t$ is the log volatility at time $t$ assumed to follow an $AR(1)$ process with $|\phi| < 1$ and $\varepsilon_t$ and $\upsilon_t$ are independent standard normal random variables with $cov(\varepsilon_t, \upsilon_s) = 0$ for all $t, s$. 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 $\eta_t, \upsilon_t$ and $\varepsilon_t$ are independent standard normal random variables and $\delta_t$ 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 $\upsilon_l$ are mutually independent for all $1 \leq j, h, k, l \leq n$. Note that the model implies that the log squared returns, $\log(x_t^2)$, follows the level shift process (2).

The presence of $\mu_t$ brings substantial flexibility. If the standard model (3) is adequate, the posterior distribution of $p$ will have large mass around zero and the model (4) essentially reduces to (3). Otherwise, the posterior distribution of $p$ is informative regarding the frequency of level shifts. Such information, along with the posterior distribution of $\sigma_\eta$, can be used to assess the relative contributions of the level shift and stationary components to overall volatility. The effects of innovations via $h_t$ quickly die out while those occurring via $\mu_t$ remain in effect until the next shift. This is different from short or long-memory SV models where all shocks affect the future path of the process in the same manner.

The model can be viewed as a member of the class of time varying parameter models. Early contributions include, among others, Rosenberg (1973), Cooley and Prescott (1976) and Tjøstheim (1986). However, being different from the majority of models in the literature, here $\mu_t$ changes only infrequently. This is essential for modelling regimes changes, although it also generates some technical difficulty that will be addressed in the next Section.

## 4 The Bayesian inference procedure

The model (4) can be represented as

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

The key ingredients for inference are the Gibbs sampler and data augmentation.

Under the assumption that $\varepsilon_t$ is *i.i.d.* $N(0,1)$, (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 $\varepsilon_t^*$ as

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

and approximate its distribution using

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

Choices of $K, q_i, m_i,\ \sigma_i^2$ follow Kim, Shephard and Chib (1998) and are described in Appendix 1. We write

$$\omega_t = j$$

if $\varepsilon_t^*$ is a realization from the $j^{th}$ component of the mixture (7).

The main challenge for inference is that both the number and locations of the level shifts are unknown. We address this issue using the technique of data augmentation and by appropriate conditioning. More specifically, the data is augmented by the locations of the shifts $\delta_t$. Conditional on $\delta_t$, the transition equation is generated by a linear recursion with normal errors. Meanwhile, as discussed above, conditional on the mixture, the measurement equation is also linear and Gaussian. Hence, the standard tools for Gaussian state space models can be applied.

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 it can be chosen based on the data. Then, model (6) can be expressed as

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

with initial conditions $(h_0, \mu_0) = 0$ and $(h_1, \mu_1)' \sim N(0, P)$. 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 procedures for the posterior distributions

We first express relevant quantities in vector notations. Let $\alpha_1 = (h_1, \mu_1)'$, $R = \{(\upsilon_1, \eta_1)', ..., (\upsilon_n, \eta_n)'\}$, $\delta = (\delta_1, ..., \delta_n)$, $\omega = (\omega_1, ..., \omega_n)$, $\theta = (\phi, \sigma_\upsilon, \sigma_\eta, p)$ and $y = (y_1, ..., y_n)$. Note that $\delta$ delivers the locations of shifts and $\delta, \alpha_1$ and $R$ jointly deliver the two components in the volatility process. The goal is to sample from the following joint posterior distribution $f(\theta, \alpha_1, R, \delta, \omega | y)$. We use a Gibbs sampler that draws from the following four blocks: (1) $f(\theta_{(-p)}, \alpha_1, R | p, \delta, \omega, y)$, where $\theta_{(-p)}$ denotes $\theta$ with $p$ excluded; (2) $f(\delta | \theta, \alpha_1, R, \omega, y)$; (3) $f(p | \theta_{(-p)}, \alpha_1, R, \delta, \omega, y)$; (4) $f(\omega | \theta, \alpha_1, R, \delta, y)$. We discuss the details of each step below.

**Step 1** (**Sampling $\theta_{(-p)}$ and the volatility process**): Write

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

This suggests sampling $R$ and $\alpha_1$ from $f(R, \alpha_1 | \theta, \delta, \omega, y)$ and $\theta_{(-p)}$ from $f(\theta_{(-p)} | p, \delta, \omega, y)$. Because $f(R, \alpha_1 | \theta, \delta, \omega, y)$ is a multivariate conditional Gaussian density, 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 3.

The density $f(\theta_{(-p)} | p, \delta, \omega, y)$ 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 objects is similar and we provide details only for the first. Using Bayes' Theorem,

$$f(\phi | \theta_{(-\phi)}, \delta, \omega, y) \propto f(y | \theta, \delta, \omega) \pi(\phi) \tag{9}$$

where $\pi(\phi)$ is the prior density and the likelihood $f(y | \theta, \delta, \omega)$ can be computed using the Kalman filter. The details for computing $f(y | \theta, \delta, \omega)$ are given in Appendix 2. Implementing (9) requires choosing a proposal density. In the empirical applications, we use the adaptive rejection Metropolis

sampler of Gilks, Best and Tan (1995). 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 providing an effective sampling algorithm. Note that because this is a Metropolis type procedure, iterations are needed for the chain to converge to the target distribution.

In the above, $\alpha_1$ and $R$ are integrated out when drawing $\theta_{(-p)}$. Instead, we could sample from $f(\theta_{(-p)}|p, \alpha_1, R, \delta, \omega, y)$. The resulting algorithm is less efficient because there is strong dependence between $\theta_{(-p)}$ and $(\alpha_1, R)$. In other words, the computational cost induced by the strong dependence between $\theta_{(-p)}$ and $R$ outweighs the gain from being able to sample from $f(\theta_{(-p)}|\alpha_1, R, p, \delta, \omega, y)$ analytically. Such a feature was also observed by Kim, Shephard and Chib (1998), who also recommended integrating out the volatility process when sampling the parameters.

**Step 2** (**Sampling the process of shifts** $\delta$): We draw iteratively from $f(\delta_t|\theta, \delta_{(-t)}, \alpha_1, R, \omega, y)$ (for $t = n, n-1, ..., 1$), where $\delta_{(-t)}$ denotes the vector of $\delta$ with the $t^{th}$ element excluded. Specifically, define $Y_t = (y_1, ..., y_t)$ and $Y_t^+ = (y_{t+1}, ..., y_n)$. For each $t$, we have

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

Because $f(\delta_t = 1|\theta, \alpha_1, R, \delta_{(-t)}, \omega, Y_t) = f(\delta_t = 1|p) = p$, the preceding display can be rewritten as

$$\begin{aligned} & f(\delta_t{=}1|\theta, \alpha_1, R, \delta_{(-t)}, \omega, y) \\ = \; & \frac{p \prod_{j=t+1}^{n} f(y_j|\theta, \alpha_1, R, \delta_t = 1, \delta_{(-t)}, \omega, Y_{j-1})}{\left\{ p \prod_{j=t+1}^{n} f(y_j|\theta, \alpha_1, R, \delta_t = 1, \delta_{(-t)}, \omega, Y_{j-1}) + \right.} \\ & \left. (1-p) \prod_{j=t+1}^{n} f(y_j|\theta, \alpha_1, R, \delta_t = 0, \delta_{(-t)}, \omega, Y_{j-1}) \right\}. \end{aligned} \tag{10}$$

Computationally, it is more convenient to look at the posterior odds ratio instead of (10):

$$\frac{f(\delta_t{=}1|\theta, \alpha_1, R, \delta_{(-t)}, \omega, y)}{f(\delta_t{=}0|\theta, \alpha_1, R, \delta_{(-t)}, \omega, y)} = \frac{p \prod_{j=t+1}^{n} f(y_j|\theta, \alpha_1, R, \delta_t = 1, \delta_{(-t)}, \omega, Y_{j-1})}{(1-p) \prod_{j=t+1}^{n} f(y_j|\theta, \alpha_1, R, \delta_t = 0, \delta_{(-t)}, \omega, Y_{j-1})}, \tag{11}$$

whose logarithm can be used to avoid numerical problems because the ratio occasionally takes on extremal values. 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.

As pointed out by a referee, instead of sampling from $f(\delta|\theta,\alpha_1,R,\omega,y)$, one could integrate $\alpha_1$ and $R$ out and sample from $f(\delta|\theta,\omega,y)$. The would require computing, instead of (11),

$$\frac{f(\delta_t=1|\theta,\delta_{(-t)},\omega,y)}{f(\delta_t=0|\theta,\delta_{(-t)},\omega,y)}=\frac{pf(y|\theta,\delta_t=1,\delta_{(-t)},\omega)}{(1-p)f(y|\theta,\delta_t=0,\delta_{(-t)},\omega)},$$

where $f(y|\theta,\delta,\omega)$ can be computed using the Kalman filter as described in Appendix 2. The resulting algorithm is valid without modifying the other steps. We implemented both procedures in the empirical section. The results suggest that the latter algorithm requires less iterations for the chain to converge to the target distribution because it avoids the dependence between $(\alpha_1,R)$ and $\delta$. However, the computational cost for obtaining a single draw of $\delta$ is also substantially higher because it requires repeating Kalman filtering $T=7823$ times. For the two empirical applications considered, the latter effect outweighs the former. Meanwhile, our applications also confirm that they lead to the same results regarding the posterior distributions of the parameters and the volatility process.

**Step 3** (**Sampling the probability of shifts** $p$): The sampling is straightforward because $f(p|\theta_{(-p)},\alpha_1,R,\delta,\omega,y)=f(p|\delta)$. 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** $\omega$): The posterior probability of $\varepsilon_t^*$ being from the $j^{th}$ component of the mixture (7) is given by

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

To obtain a draw, we first compute $\varepsilon_t^*=y_t-h_t-\mu_t$ with $h_t$ and $\mu_t$ obtained from $\alpha_1$ and $R$ using (8). Next, we compute $f(\varepsilon_t^*|\omega_t=j)$, for $j=1,...,K$, using $\varepsilon_t^*|(\omega_t=j)\sim N(m_j,\sigma_j^2)$. Then, we simply draw from the resulting multinomial distribution.

The approximation error induced by the offset mixture approximation can be corrected using a reweighting step as in Kim, Shephard and Chib (1998), or a Metropolis-Hastings step as described below. The basic idea is to treat the approximating model as one that generates a proposal distribution. Given any $(\theta,\alpha_1,R,\delta)$, the likelihood functions for the approximating and the true models are

$$\begin{aligned}k(y|\theta,\alpha_1,R,\delta) &= \textstyle\prod_{t=1}^n\sum_{i=1}^K q_i f_N\left(y_t|m_i+h_t+\mu_t,\sigma_i^2\right),\\ g(y|\theta,\alpha_1,R,\delta) &= \textstyle\prod_{t=1}^n f_N\left(x_t|0,\exp\left(h_t+\mu_t\right)\right),\end{aligned}$$

where $f_N(y_t|a,b)$ stands for the Normal density with mean $a$ and variance $b$, $(q_i, m_i,\ v_i^2)$ and $K$ follow from the mixture, $h_t + \mu_t$ can be computed from $\alpha_1$ and $R$ and $x_t$ is the return at time $t$. To implement the Metropolis-Hastings algorithm, we start with some initial values, say $(\theta^{(0)}, \alpha_1^{(0)}, R^{(0)}, \delta^{(0)})$, take a random draw from those given by Steps 1-4 say $(\theta^*, \alpha_1^*, R^*,\ \delta^*)$, and accept it with probability

$$r = \min\left(\frac{g(y|\theta^*, \alpha_1^*, R^*, \delta^*)/k(y|\theta^*, \alpha_1^*, R^*, \delta^*)}{g(y|\theta^{(0)}, \alpha_1^{(0)}, R^{(0)}, \delta^{(0)})/k(y|\theta^{(0)}, \alpha_1^{(0)}, R^{(0)}, \delta^{(0)})}, 1\right).$$

This is then continued to deliver the desired number of draws. We implemented both procedures (i.e., with and without reweighting) in the empirical applications and obtained very similar posterior distributions.

Although the current paper focuses on daily returns, the sampling procedure proposed above can also be used to analyze realized volatility allowing for level shifts. In that case, the left hand side variable in Equation (6) becomes the logarithm of realized volatility, the processes for $h_t$ and $\mu_t$ remain the same, while $\varepsilon_t^*$ can for example be modeled as $i.i.d.N(0, \sigma_\varepsilon^2)$. The parameter $\sigma_\varepsilon^2$ can be treated as part of $\theta$. Then, Steps 1 to 3 can be applied to deliver the posterior distributions. Note that the offset mixture approximation and Step 4 are no longer relevant.

### 4.2 The specifications of the priors

We use independent priors. Those for $\phi$ and $\sigma_\upsilon$ follow Kim, Shephard and Chib (1998). For $\phi$,

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

This distribution has support on $(-1, 1)$ with a mean of $2\phi^{(1)}/(\phi^{(1)} + \phi^{(2)}) - 1$. In the empirical applications, we set $\phi^{(1)} = 20$ and $\phi^{(2)} = 1.5$, implying a prior mean of 0.86. For $\sigma_\upsilon$, we use $\sigma_\upsilon^2 \sim \mathcal{IG}(\sigma_r/2, S_\sigma/2)$, i.e., 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 $\sigma_\eta$ 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 $1/41$ so that a shift occurs on average every 41 days. For $\sigma_\eta$, we specify $\sigma_\eta^2 \sim \mathcal{IG}(\sigma_r^*/2, S_\sigma^*/2)$ with $\sigma_r^* = 20$ and $S_\sigma^* = 60$. This implies an approximate prior mean of 3.33 and variance of 1.39. Finally, we use the following prior distribution for the initial states: $(h_1, \mu_1) \sim N(0, P)$ with $P = Diag(1 \times 10^6, 1 \times 10^6)$.

# 5 Filtering and forecasting

We discuss filtering and forecasting conditional on estimated parameters. We use $\theta$ to denote the parameter estimates, which can be the posterior means or medians. Let $X_t = (x_1, ..., x_t)'$ denote the vector of returns up to time $t$ and define $\alpha_t = (h_t, \mu_t)$. Since we work with the formulation (6) instead of (8), the mixture approximation is no longer needed.

## 5.1 Filtering

The objective is to recursively obtain a sample of draws from $(\alpha_t|X_t, \theta)$ for $t = 1, ..., n$. The filtered signal $E(\exp(h_t+\mu_t)|X_t, \theta)$ can then 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,

$$f(\alpha_{t+1}|X_{t+1}, \theta) = \frac{f(x_{t+1}|\alpha_{t+1}, X_t, \theta)}{f(x_{t+1}|X_t, \theta)} \int f(\alpha_{t+1}|\alpha_t, X_t, \theta) dP(\alpha_t|X_t, \theta). \tag{12}$$

This is not directly useful because 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 Smith (1993) and Kim, Shephard and Chib (1998). More specifically, 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 $f(\alpha_{t+1}|X_{t+1}, \theta)$ can be obtained by drawing from $f(\alpha_{t+1}|\alpha_t^{(j)}, X_t, \theta)$ and reweighting them using $f(x_{t+1}|\alpha_{t+1}^{(j)}, X_t, \theta)$. The distribution $f(\alpha_{t+1}|\alpha_t^{(j)}, X_t, \theta)/f(x_{t+1}|X_t, \theta)$ depends on whether a shift occurs at time $t$ and is given by

$$\alpha_{t+1}|(\alpha_t^{(j)}, X_t, \theta) \sim \delta_t W_{1t}^{(j)} + (1 - \delta_t) W_{2t}^{(j)} \tag{13}$$

with

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

The associated weights are given by $w_{t+1}^{(j)} = f(x_{t+1}|\alpha_{t+1}^{(j)}, X_t, \theta)/\sum_{j=1}^M f(x_{t+1}|\alpha_{t+1}^{(j)}, X_t, \theta)$, 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 Forecasting

The subsequent analysis is conditional on a given $\theta$ (say the posterior mean or mode), the same objects can be obtained by averaging over the posterior distribution of the parameters.

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 $X_f$, which leads to the following procedure: (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 $f(\alpha_f|X_f, X_t, \theta)$ using the algorithm in Section 4.1 (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.

Using these draws, the minimum mean squared error (MSE) estimate of the of the p-step ahead forecast of volatility can be approximated by $E(\exp(h_{t+p} + \mu_{t+p})|X_t, \theta) \approx (1/M) \sum_{j=1}^{M} \exp(h_{t+p}^{(j)} + \mu_{t+p}^{(j)})$ and the estimate of average volatility over the one to p-step ahead forecasts is

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

These approximations can be made precise by choosing $M$ large enough.

The preceding algorithm allows for out-of-sample level shifts. If level shifts are expected to be rare and the forecasting horizon is relatively short, as will be the case in our applications, it is sensible to assume that they do not occur out-of-sample, 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 $\mu_{t+p} = \mu_t$ and $h_{t+p} = \phi^p h_t + \sum_{i=1}^{p} \phi^{i-1} \sigma_\upsilon \upsilon_{t+p-i}$. The predicted p-step ahead volatility is then

$$E(\exp(h_{t+p} + \mu_{t+p})|X_t) = E(\exp(\phi^p h_t + \mu_t)|X_t) \prod_{i=0}^{p-1} \exp(\frac{\phi^{2i} \sigma_\upsilon^2}{2}), \tag{14}$$

where the last equality follows by normality of $\upsilon_t$. The term $E(\exp(\phi^p h_t + \mu_t)|X_t)$ can be evaluated using particle filtering, as discussed in Section 5.1. Note that the quantities appearing in the forecast, $\phi, \sigma_\upsilon, h_t$ and $\mu_t$, are estimated using all observations up to time $t$, not merely observations from the most recent regime. Therefore, in-sample shifts continue to affect the forecast.

# 6 Applications to S&P 500 and NASDAQ returns

We apply the model to the volatility of the S&P 500 and NASDAQ daily returns over the period 1980.1-2005.12. We consider: a) the posterior distributions of the parameters using the full sample; b) whether the model evaluated at the posterior means can replicate keys features of the data; c) the comovement between the estimated volatility components and some indicators of the business cycle; and d) the forecasting performance of the model relative to other popular forecasting models. All the reported results are without the reweighting step described in Section 4.1.

### 6.1 The posterior distributions based on the full sample

We used the priors discussed in Section 4 and generated the posterior distributions based on 10,000 draws, discarding the first 5,000. 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 tried different initial values and the results did not change.

Consider first the results for the S&P 500 daily returns. Figure 2 presents the posterior means of the level shift component and the log volatility process. Figure 3 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 0.00218 with a 95% confidence interval of (0.00107 0.00365). In terms of duration, this implies that on average, a shift occurs every 459 days (with a 95% confidence interval of (274, 938) days). The results are substantially different from the prior which implies an average of one shift per 41 days. 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, $s_t = \bar{s} + (\mu_t - \bar{\mu}) + (h_t - \bar{h})$. Then

$$\frac{\sum_{i=1}^n (\mu_t - \bar{\mu})^2}{\sum_{i=1}^n s_t^2} \text{ and } \frac{\sum_{i=1}^n (h_t - \bar{h})^2}{\sum_{i=1}^n s_t^2} \tag{15}$$

measure the relative contributions of $\mu_t$ and $h_t$ to the overall variation in volatility. Their values are 0.59 and 0.20, respectively. This 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 0.956 with a 95% confidence interval of (0.934,0.974). This implies that the short-memory component is indeed much less persistent when level shifts are accounted for, as typical estimates of $\phi$ using standard SV models are very close to 1. Indeed, in our model the half life of a typical shock is only about 15 days with a 95% confidence interval of (10, 26) days. This contrasts with the estimate obtained using a standard SV model for which the posterior mean of $\phi$ implies a half life of 58 days with a 95% confidence interval of (51, 90) days.

Figure 2 suggests that the shifts often coincide with important events. The first major shift occurred on 10/12/1987 (Monday), leading to the "Black Monday" on 10/19/1987. The posterior mean $\mu_t$ around "Black Monday" are:

| Date | 10/9 | 10/12-10/14 | **10/15-10/23** | 10/26-01/11/88 | 01/12-01/15 | 01/18 |
|---|---|---|---|---|---|---|
| $\mu_t$ | -0.31 | 0.69 | **2.26** | 1.58 | 0.88 | -0.06 |

.

The results are quite informative. First, the market was already showing unusually high volatility one week before the crash occurred. Second, the high volatility surrounding the "Black Monday" lasted about a week, until October 23, 1987 (Friday). The volatility then dropped suddenly on October 26, 1987 (Monday). The new regime lasted until January 11, 1988 (Monday), when the market returned to a volatility level comparable to that prior to the crash on January 18, 1988 (Monday). The market then entered a long period of low volatility starting in April, 1992. Volatility started to increase again in June, 1996, predating the Asian financial crisis. This transition was more gradual and increases in volatility continued until August, 1998. Then, a substantial decrease occurred between June and October, 2003, with this new regime persisting to February, 2007.

The volatility started to pick up again in early 2007, which continued till the market crash in 2008. The posterior mean of $\mu_t$ around "Black Monday", September 15, 2008 are:

| Date | 09/05 | 09/08-09/11 | **09/12-11/13** | 11/14-06/02/09 | 06/03-07/15 | 07/16 |
|---|---|---|---|---|---|---|
| $\mu_t$ | 0.67 | 1.91 | **2.16** | 2.15 to 1.38 | 0.39 | 0.21 |

.

The results are again informative. First, the market was showing unusually high volatility before the crash occurred. Second, the high volatility surrounding the "Black Monday" lasted about two months until November 13, 2008 (Thursday). Then, it started to decrease gradually over an approximate six-month period to a level of 1.38. Afterwords, the volatility dropped suddenly on June 3, 2009 (Wednesday) and remained at relatively mild levels throughout 2010.

Our model ascribes the very large variances in October 1987 and September 2008 as brief changes in volatility level instead of extreme draws from the underlying distribution of returns. We do not view this as a defect of our procedure. On the contrary, events like the crashes of 1987 and 2008 are highly unlikely to be draws from the stationary short-memory stochastic component of volatility. Our method therefore has the advantage of purging the volatility process from both low frequency shifts and rare highly influential events. This allows the short-memory component to be more representative of the stochastic process underlying regular fluctuations in volatility. Also, this feature does not imply a misspecification of our model.

For the NASDAQ returns, the results are summarized in Figures 4 and 5. The implications are broadly similar to those for the S&P 500 returns. The ratios defined in (15) are 0.80 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 0.00301 with a 95% confidence interval of (0.00160, 0.00481). In terms of regime duration, this implies, on average, one shift every 332 days (with a 95% confidence interval of (208, 624) days). The estimates of the level shift

component are similar to those for the S&P 500 index. This suggests that common underlying causes affect the volatility of the two markets. With respect to the AR(1) coefficient of the short-memory component, the posterior distribution has a mean of 0.925 with a 95% interval of (0.899, 0.948). 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 9 days with a tight 95% confidence interval of (6.5, 13.0) days. This again contrasts sharply with standard SV model for which the posterior mean implies a half life of 46 days with a 95% confidence interval of (65, 97) days.

Figures 2 and 4 show that the durations of regimes exhibit bimodality. Very high volatility regimes tend to be shorter lived relative to lower volatility regimes. Our model delivers this result without requiring a mixture prior. This can be viewed as further evidence supporting the simple level shift model. Meanwhile, there are two possible ways to explicitly allow for bimodality. First, a more flexible (may be a mixture) prior for $p$ can be used. Second, $p$ can be modeled as a function of $\mu_t$, thus the probability of entering a new regime can be higher when the current volatility level is high. The latter generalization can be particularly interesting. It is relatively straightforward to allow for the two generalizations in the sampling procedure.

We carried out some experimentations to examine the sensitivity of the posterior distributions to the priors about $p$ and $\sigma_\eta^2$. Specifically, the parameters $\gamma_1, \gamma_2, \sigma_r^*$ and $S_\sigma^*$ are changed one at a time while keeping the others fixed at their original values. The results for S&P 500 (posterior means and 95% confidence sets) are summarized in Table 1. Findings for NASDAQ are similar and thus omitted. The changes in the priors have little effect on $\phi$ and $\sigma_\upsilon$, while at the same time they affect the posteriors of $p$ and $\sigma_\eta$. When using priors favoring smaller (i.e., larger $\sigma_r^*$ or smaller $S_\sigma^*$) or more frequent (i.e., larger $\gamma_1$ or smaller $\gamma_2$) shifts, the posterior distributions of $p$ and $\sigma_\eta$ also move in the same direction. The most substantial effect is that of $\sigma_r^*$ and $S_\sigma^*$ on $\sigma_\eta$. This is unsurprising because the shifts are rare and $\sigma_\eta$ gets updated only when a shift occurs. Therefore, the strong effect of the prior is not particular to the current model, but rather reflecting a common feature of all models with rare shifts. In spite of the differences in parameter estimates, upon plotting the estimated level shift components in the eight cases, we find that they are very similar to the one reported in Figure 2. (The details are omitted to save space.) Thus the key findings reported above stay qualitatively the same under the eight priors considered here.

### 6.2 Does the model explain key features of the data?

We consider three issues. First, we assess the fit of our model using some standard diagnostic statistics. Second, we estimate standard SV models using the subsamples identified by our model

to assess the time variations in $\phi$ and $\mu$ (c.f. (3)). This allows us to look into two issues: (1) whether assuming a constant $\phi$ is consistent with the data and (2) whether the changes in $\mu$ are statistically significant. Third, and more importantly, we analyze simulated artificial samples using the estimated parameter values (the posterior means). These allow us to examine whether our model explains the key feature of the data shown in Figures 1, the path the log-periodogram estimates of the long-memory parameter take as a function of the number of frequency ordinates used.

Figure 6 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 $(x_t/\sigma_t)$.

Table 2 reports the posterior means and 95% credible sets of $\phi$ and $\mu$ estimated using subsamples. The periods corresponding to the two market crashes are excluded because they have too few observations. The same priors as in the full sample case are used throughout. For S&P 500 returns, the estimates of $\phi$ are between 0.866 and 0.979. For NASDAQ returns, they are between 0.804 and 0.965. The 95% credible sets overlap in both cases. They also overlap with the 95% credible sets obtained using the full sample, in which case they are (0.934,0.974) and (0.899, 0.948) respectively. Therefore, the results support the modeling assumption that $\phi$ remains constant throughout the sample. In contrast, the credible sets for $\mu$ vary significantly over time. There is no single value of $\mu$ that is contained in all the credible sets. This is true for both S&P 500 and NASDAQ returns.

To further 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 $n = 7823$, using $\phi = 0.956, \sigma_v = 0.152, \sigma_\eta = 1.623, \delta_t \sim B(1, 0.00218)$ and $(h_1, \mu_1) = (0.539, -0.232)$. For each simulated sample, we computed 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 7. They generate patterns closely resembling those of the actual data depicted in Figure 1. We believe this 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.

### 6.3 Comovement with business cycle indicators

It has long been of interest to study the connection between stock market volatility and business cycle conditions (Schwert, 1989). Here, we examine comovements between the estimated volatility components and nine indicator variables for the U.S. economy. The first eight indicators are business cycle leading indicators, including interest rate spread, consumer sentiment, money supply growth, vendor performance, new orders for capital goods, average weekly work hours, new building permits and initial claims for unemployment insurance. The ninth variable is the Coincident Index for the US economy. Detailed definitions and the data source are in Appendix 4. We report results from bivariate linear regressions of the indicator variables on the estimated volatility components. Special attention is paid to how the $R^2$ and statistical significance change when different volatility components are included. Note that when considering the Coincident Index, the regressor is lagged by two quarters relative to the dependent variable because volatility is expected to lead the business cycle; the other regressions are run contemporaneously. All regressions use monthly observations starting at 1992:2 rather than 1980:1 constrained by the availability of observations on the new orders series. We consider two sample periods: 1992:2-2005:12 and 1992:2-2010:12. Their difference is informative about the recent financial crisis on the comovements between variables.

The first Panel in Table 3 reports results for the S&P 500 series for the period 1992:2-2005:12. When the indicator variable is regressed on the level shift component ($\mu_t$), the estimated slope coefficient is statistically significant (at 10% or higher level) in six among the nine regressions. The $R^2$ for the six regressions are between 6% and 31%. When the indicator is regressed on $h_t$, only one regression produces a statistically significant coefficient. This pattern is further strengthened by regressing the indicator on $\mu_t + h_t$. The resulting $R^2$ generally decreases relative to the regression on $\mu_t$ (except for the money supply and work hour series, whose $R^2$ increase by 0.01) and the coefficient tends to become less significant. We also considered reversed regressions by regressing $\mu_t$ and $h_t$ on the eight leading indicators in the table. The adjusted $R^2$ equals 64% for $\mu_t$ and 9% for $h_t$. The findings for the NASDAQ volatility series are qualitatively similar. Clearly, these results do not have any causal interpretation. However, they do strongly suggest that the level-shift component is closely linked to business cycle fluctuations while $h_t$ is not. Consequently, after isolating the level shift component from the overall volatility, we observe a stronger volatility-business cycle relationship.

Table 4 presents the results for the sample period that includes the recent financial crisis (1992:2-2010:12). There, the comovements between the volatility and macroeconomic indicators are weakened. The $R^2$ generally decreases. Three regression that include interest spread, consumer

sentiment and new orders no longer show statistical significance. Nevertheless, the overall pattern remains the same, i.e., the relationship with $\mu_t$ continue to be stronger than with $\mu_t + h_t$.

### 6.4 Forecasting volatility for S&P 500 and NASDAQ daily returns

We consider one and multi-step ahead forecasts for volatility using a setup similar to that of Stărică and Granger (2005). We used the first 2000 observations for initial estimation and re-estimated the model with the addition of 20 observations. In each step, the same priors as discussed in Section 4.2 were used. We then made forecasts for horizons up to 20 days ahead. Filtered states are used to ensure that only in-sample information is applied when forecasting is conducted. 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 irrelevant for our level shift model since the issue of nonstationarity is automatically taken into account via the level shift process.

Our metric for comparison follows Andersen, Bollerslev, Diebold and Labys (2003) and is in the tradition of Mincer and Zarnowitz (1969). We evaluate the relative forecasting accuracy by considering 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 $x_{t,p}^2$ is an unbiased estimate for the volatility at horizon $p$. Here we use cumulative squared demeaned returns, i.e., $x_{t,p}^2 = \sum_{j=1}^{p} x_{t+j}^2$, although it can be more desirable to use returns at a higher frequency. Note that the series $x_{t,p}^2$ contains large outliers which may have an overwhelming effect on the measure. To avoid this problem, we discarded observations whose absolute values were above 6%. None were discarded from the NASDAQ series since, in this case, the results did not change. After discarding, we were left with 7803 observations for the S&P 500 series and 7823 for the NASDAQ series. In the above, $\hat{\sigma}_{t,p}^2$ denotes the $p$-step ahead forecast from the level shift model (label LS) and $\hat{\sigma}_{t,p,i}^2$ denotes the forecast from one of the following four models: a) SV without level shifts in which forecasts are obtained using a recursive method (labeled SV-Rec), b) SV without level shifts in which forecasts are obtained using a rolling window (labeled SV-Rol), c) FIGARCH with forecasts obtained using a recursive method (labeled FIG-Rec) and d) FIGARCH with forecasts obtained using a rolling window (labeled FIG-Rol). The level shift model can be said to generate better (or no worse) forecasts than a competitor if the values $(0, 1, 0)$ are in the confidence intervals for $(b_0, b_1, b_2)$, respectively. We estimate regression (16) for forecast horizons between 1 and 5 days. Table 5 presents the parameter estimates, their 95% confidence intervals

and the adjusted $R^2$ from the corresponding regression. Out of the 40 regressions, there are 37 regressions with $(0, 1, 0)$ lying inside the confidence intervals for $(b_0, b_1, b_2)$. In the S&P 500 case, the intervals for $b_1$ is in general quite wide and in many case the intervals for $b_2$ also include 1. This suggests the level shift model performs similarly to the other models, but do not dominate them. In the NASDAQ case, the evidence clearly favors the level shift model.

In summary, the forecasting performance of the proposed model is at least comparable and can be better than the standard SV and FIGARCH models. This shows that the level shift component is not a modeling convenience allowing for a better in-sample fit. The result also contrasts with the common perception that structural change models are not useful for forecasting. Pesaran, Pettenuzzo and Timmermann (2006) provided another interesting example in which they considered 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.

# 7 Conclusion

This paper has provided a framework for modeling, inference and forecasting of volatility in the presence of level shifts of random timing, magnitude and frequency. We showed that a very simple stochastic volatility model incorporating both a random level shift and a short-memory component provides a good 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 the model can be extended in several directions to provide a deeper understanding of the structure of volatility and improve forecasting performance. This could be done by incorporating covariates into the model. Such extensions are not trivial, though we feel that our results point to the need for work along these lines.

## References


Andersen, 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.

Andersen, T., T. Bollerslev, F.X. Diebold and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 579-626.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 73, 3–30.

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

Bollerslev, T. and H.O. Mikkelsen (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73, 151-184.

Breidt, J.F., N. Crato and P.J.F. de Lima (1998). On the detection and estimation of long-memory in stochastic volatility. *Journal of Econometrics* 83, 325–348.

Carter, C.K., and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81, 541-553.

Chib, S. (1998). Estimation and comparison of multiple change point models. *Journal of Econometrics* 86, 221-241.

Cooley, T.F. and E.C. Prescott (1976). Estimation in the presence of stochastic parameter variation. *Econometrica* 44, 167-184.

De Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics* 19, 1073-1083.

De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika* 82, 339-350.

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

Diebold, F.X. and A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics* 105, 131–159.

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.

Duffie, D., J. Pan and K.J. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.

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

Engle, R.F. and J.G. Rangel (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21, 1187-1222.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance* 59, 1367-1404,

Eraker, B., M. Johannes and N. Polson (2003). The impact of jumps in volatility and returns. *The Journal of Finance* 58, 1269-1300.

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

Gilks, W.R. and P. Wild (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41, 337-348.

Gilks, W.R. (1996). Full conditional distributions. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Gilks, W.R., N.G. Best and K.K.C. Tan (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* 44, 455-473.

Gordon, N.J., D.J. Salmond and A.F.M. Smith (1993). A novel approach to non-linear and non-Gaussian bayesian state estimation. *IEE-Proceedings F* 140, 107-133.

Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14, 227-238.

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.

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.

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

Hamilton, J.D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307-333.

Hyung, N., S.H. Poon and C.W.J. Granger (2008). A source of long memory in volatility. In: Rapach, D., Wohar, M. (Eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Emerald Group Publishing, Howard House, UK, p. 329–380

Jacquier, E., N. Polson and P. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371-418.

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.

Koop, G. and S.M. Potter (2007). Estimation and forecasting in models with multiple breaks. *Review of Economic Studies* 74, 763-789.

Koopman, S.J. (1993). Disturbance smoother for state space models. *Biometrika* 80, 117-126.

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.

Laurent, S and J.-P. Peters (2006). *Estimating and forecasting ARCH models.* Timberlake Consultants Press.

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.

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.

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.

Mincer, J. and V. Zarnowitz (1969). The evaluation of economic forecasts. *Economic Forecasts and Expectations*, ed. by J. Mincer. New York: National Bureau of Economic Research.

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

Perron, P. and Z. Qu (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economics Statistics 28, 275-290.*

Pesaran, M.H., D. Pettenuzzo and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *Review of Economic Studies* 73, 1057–1084.

Rosenberg, B. (1973). The analysis of a cross-section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement* 2, 399-428.

Schwert, G.W. (1989). Why does stock market volatility change over time? *Journal of Finance* 44, 1207–1239.

Shephard, N. (1994). Partial non-Gaussian state space. *Biometrika* 81, 115-131.

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

Smith, A. (2005). Level shifts and the illusion of long memory in economic time series. *Journal of Business and Economic Statistics* 23, 321-335.

Stărică, C., and C.W.J. Granger (2005). Nonstationarities in stock returns. *Review of Economics and Statistics* 87, 503-522.

Tjøstheim, D.(1986). Some doubly stochastic lime series models. *Journal of Time Series Analysis* 7, 51-72.

## Appendix 1: The mixture of Kim, Shephard and Chib (1998)

As stated in Section 4, $\varepsilon_t^*$ is approximated by a 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 $q_i$, $m_i$ and $\sigma_i^2$ are the same as in Kim, Shephard and Chib (1998):

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

.

## Appendix 2: SV model in state space form and the Kalman filter

To relate the model to the state space modeling literature, we write it 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 $y_t = \log(x_t^2 + c) - E(\log \varepsilon_t^2)$, $\alpha_t = (h_t, \mu_t)'$, $u_t \sim i.i.d.\ N(0, I_3)$, $Z = (1, 1)$, $G_t = (\sigma(\omega_t), 0, 0)$ with $\omega_t = j$ denoting the j$^{th}$ component in the mixture,

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

The initial state is $\alpha_0 = 0$ and $\alpha_1 \sim N(0, P)$, where $P = 1 \times 10^6$, indicating a diffuse prior.

Given the locations of shifts $\delta$ and the realization of the mixture $\omega$, (A.1) reduces to a standard linear Gaussian state space model, consequently the standard Kalman filtering algorithm to construct the likelihood function can be applied without modification. The steps involved are as follows. 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): $e_t = y_t - Za_t$, $D_t = ZP_tZ' + G_tG_t'$, $K_t = (TP_tZ')D_t^{-1}$, $a_{t+1} = Ta_t + K_te_t$ and $P_{t+1} = TP_t(T - K_tZ)' + H_tH_t'$, with the initialization $a_1 = 0$ and $P_1 = P$. Also,

$$H_t = \begin{pmatrix} 0 & \sigma_\upsilon & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ if } \delta_t = 0 \text{ and } H_t = \begin{pmatrix} 0 & \sigma_\upsilon & 0 \\ 0 & 0 & \sigma_\eta \end{pmatrix} \text{ if } \delta_t = 1.$$

## Appendix 3: The simulation smoother

We adapt the simulation smoother of De Jong and Shephard (1995) to sample from

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

Let $e_t, D_t$ and $K_t$ be the output obtained from applying the Kalman filter discussed in Appendix 2 and define $F_t$ $(t = 1, ..., n)$ as

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

Starting from $r_n = 0$ and $U_n = 0$, apply the following backward recursion for $t = n, n-1, ..., 1$: $C_t = F_t(I - H_t' U_t H_t)F_t'$, $z_t \sim i.i.d.\ N(0, C_t)$, $V_t = F_t H_t' U_t L_t$, $r_{t-1} = Z' D_t^{-1} e_t + L_t' r_t - V_t' C_t^{-1} z_t$ 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 compute

$$\widetilde{R}_t = (F_t H_t' r_t + z_t, z_t^*)' \text{ if } \delta_t = 0 \text{ and } \widetilde{R}_t = F_t H_t' r_t + z_t \text{ if } \delta_t = 1.$$

Save $R_t = Diag(\sigma_\upsilon, \sigma_\eta)\widetilde{R}_t$. Finally, set $\alpha_1 = Pr_0$. Then, the values of $R = (R_1, ..., R_n)$ and $\alpha_1$ yield draws from (A.2). The validity of this algorithm can be verified along the lines of De Jong and Shephard (1995, p. 348-349). We omit the details.

## Appendix 4: Data used in the empirical application

The datasets were retrieved from the web page of the Federal Reserve Bank of St. Louis. They contain monthly observations from 1992-02 to 2010-12.

1. Interest Rate Spread: 10-Year Treasury Bond minus Federal Funds Rate. Series ID: GS10 and FEDFUNDS. Source: Board of Governors of the Federal Reserve System. Not Seasonally Adjusted.

2. University of Michigan: Consumer Sentiment Index. Series ID: UMCSENT. Source: Survey Research Center: University of Michigan. Not Seasonally Adjusted.

3. Money Supply Growth (M2). Series ID: M2SL. Source: Board of Governors of the Federal Reserve System. Compounded Annual Rate of Change. Seasonally Adjusted.

4. Supplier Deliveries Index: Manufacturing. Series ID: NAPMSDI. Source: Institute for Supply Management. Seasonally Adjusted. Natural logarithm used in the regression.

5. Manufacturers' New Orders: Nondefense Capital Goods Excluding Aircraft. Series ID: NEWORDER. Source: Department of Commerce: Census Bureau. Seasonally Adjusted. Natural logarithm used.

6. Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing. Series ID: AWHMAN. Source: U.S. Department of Labor: Bureau of Labor Statistics. Seasonally Adjusted.

7. New Private Housing Units Authorized by Building Permits. Series ID: PERMIT. Source: Department of Commerce: Census Bureau. Seasonally Adjusted. Natural logarithm used.

8. Four-Week Moving Average of Initial Claims for Unemployment Insurance. Series ID: IC4WSA. Source: U.S. Department of Labor: Employment and Training Administration. Seasonally Adjusted. Natural logarithm used in the regression.

9. Coincident Economic Activity Index for the United States. Series ID: USPHCI. Source: Federal Reserve Bank of Philadelphia. Percent Change. Seasonally adjusted.

Table 1. Posterior means and 95% confidence sets under different priors for $p$ and $\sigma_\eta^2$

| | (a) Vary $\gamma_1$ | | | | (b) Vary $\gamma_2$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $\gamma_1 = 0.25$ | | $\gamma_1 = 4$ | | $\gamma_2 = 10$ | | $\gamma_2 = 160$ | |
| $p$ | 0.00198 | [0.00096, 0.00345] | 0.00300 | [0.00158, 0.00522] | 0.00211 | [0.00089,0.00379] | 0.00199 | [0.00101,0.00332] |
| $\phi$ | 0.960 | [0.936,0.976] | 0.952 | [0.924, 0.974] | 0.952 | [0.922, 0.973] | 0.958 | [0.937,0.974] |
| $\sigma_\upsilon$ | 0.146 | [0.117,0.179] | 0.158 | [0.125, 0.195] | 0.159 | [0.129,0.197] | 0.151 | [0.1212,0.183] |
| $\sigma_\eta$ | 1.663 | [1.291,2.175] | 1.552 | [1.198, 2.007] | 1.615 | [1.253, 2.105] | 1.653 | [1.289,2.140] |
| | (c) Vary $\sigma_r^*$ | | | | (d) Vary $S_\sigma^*$ | | | |
| | $\sigma_r^* = 10$ | | $\sigma_r^* = 40$ | | $S_\sigma^* = 30$ | | $S_\sigma^* = 120$ | |
| $p$ | 0.00214 | [0.00104, 0.00362] | 0.00295 | [0.00128, 0.00545] | 0.00255 | [0.00114, 0.00474] | 0.00197 | [0.00095, 0.00330] |
| $\phi$ | 0.956 | [0.931, 0.973] | 0.950 | [0.922, 0.970] | 0.955 | [0.923, 0.973] | 0.960 | [0.938, 0.979] |
| $\sigma_\upsilon$ | 0.150 | [0.121, 0.184] | 0.159 | [0.127, 0.196] | 0.153 | [0.123, 0.192] | 0.151 | [0.119, 0.184] |
| $\sigma_\eta$ | 2.034 | [1.502, 2.766] | 1.219 | [1.001, 1.508] | 1.259 | [0.958, 1.681] | 2.134 | [1.668, 2.970] |

Note. The respective prior means are as follows. For $p$ in (a): 0.00621 and 0.09091; for $p$ in (b): 0.09091 and 0.00621; for $\sigma_\eta$ in (c): 2.739 and 1.257; for $\sigma_\eta$ in (d): 1.291 and 2.582.

Table 2. Results from estimating SV models using subsamples

| S&P 500 returns | | | | | |
|---|---|---|---|---|---|
| Sample period | #Obs | $\phi$ | | $\mu$ | |
| | | Posterior mean | 95% credible set | Posterior mean | 95% credible set |
| 1/2/80–10/9/87 | 1966 | 0.979 | [0.961,0.991] | -0.267 | [-0.499,-0.008] |
| Crash period | – | – | – | – | – |
| 1/18/88-2/4/92 | 1025 | 0.866 | [0.622,0.964] | -0.335 | [-0.503,-0.154] |
| 2/5/92-6/21/96 | 1108 | 0.912 | [0.760,0.976] | -1.181 | [-1.370,-0.992] |
| 6/24/96-8/28/98 | 552 | 0.950 | [0.885,0.987] | -0.176 | [-0.601,0.292] |
| 8/31/98-4/24/03 | 1168 | 0.965 | [0.934,0.987] | 0.574 | [0.280,0.883] |
| 4/25/03-2/26/07 | 966 | 0.977 | [0.950,0.993] | -0.802 | [-1.184,-0.434] |
| 2/27/07-9/3/08 | 384 | 0.966 | [0.917,0.992] | 0.220 | [-0.379,0.876] |
| Crash period | – | – | – | – | – |
| 6/3/09-12/2/10 | 381 | 0.954 | [0.896,0.989] | 0.076 | [-0.458,0.685] |
| NASDAQ returns | | | | | |
| Sample period | #Obs | $\phi$ | | $\mu$ | |
| | | Posterior mean | 95% credible set | Posterior mean | 95% credible set |
| 1/2/80-8/3/84 | 1162 | 0.926 | [0.869,0.967] | -0.620 | [-0.833, -0.395] |
| 8/6/84-10/8/87 | 803 | 0.854 | [0.737,0.938] | -1.313 | [-1.532,-1.109] |
| Crash period | – | – | – | – | – |
| 2/4/88-7/16/90 | 618 | 0.913 | [0.820,0.972] | -1.309 | [-1.576,-1.025] |
| 7/17/90-6/19/95 | 1246 | 0.928 | [0.859,0.972] | -0.540 | [-0.753,-0.330] |
| 6/20/95-8/20/98 | 801 | 0.917 | [0.847,0.964] | -0.011 | [-0.245,0.217] |
| 8/21/98-12/31/99 | 344 | 0.944 | [0.850,0.991] | 1.176 | [0.695,1.688] |
| 1/3/00-4/20/01 | 328 | 0.965 | [0.917,0.992] | 2.242 | [1.595,2.944] |
| 4/23/01-4/22/03 | 499 | 0.931 | [0.796,0.986] | 1.430 | [1.163,1.646] |
| 4/23/03-7/28/04 | 319 | 0.804 | [0.524,0.959] | 0.385 | [0.247,0.522] |
| 7/29/04-7/18/07 | 747 | 0.943 | [0.870,0.984] | -0.352 | [-0.552,-0.145] |
| 7/19/07-9/12/08 | 292 | 0.839 | [0.586,0.970] | 0.709 | [0.535,0.870] |
| Crash period | – | – | – | – | – |
| 6/2/09-12/2/10 | 381 | 0.950 | [0.882,0.988] | 0.240 | [-0.253,0.761] |

Table 3. Comovement between volatility components and business cycle indicators
(Sample period: 1992.2-2005.12)

| | $\mu_t$ | | $h_t$ | | $\mu_t + h_t$ | |
|---|---|---|---|---|---|---|
| | coefficient (t-stat) | $R^2$ | coefficient (t-stat) | $R^2$ | coefficient (t-stat) | $R^2$ |
| **Panel (a). S&P 500** | | | | | | |
| Interest spread | $-0.81^{**}$ $(-3.29)$ | 0.16 | $-0.02$ $(-0.06)$ | 0.00 | $-0.57$ $(-1.46)$ | 0.12 |
| Consumer sentiment | $6.71^{*}$ (1.83) | 0.23 | $-2.56$ $(-0.89)$ | 0.00 | 4.20 $(-1.54)$ | 0.12 |
| Money supply | $3.37^{**}$ (6.22) | 0.27 | $2.68^{**}$ (2.15) | 0.05 | $2.83^{**}$ (6.37) | 0.28 |
| Vendor performance | $-0.03^{*}$ $(-1.85)$ | 0.06 | 0.01 (0.35) | 0.00 | $-0.02^{*}$ $(-1.83)$ | 0.04 |
| New orders | $0.12^{*}$ (1.81) | 0.31 | 0.03 (0.79) | 0.00 | $0.09^{*}$ (1.84) | 0.24 |
| Work hours | $-0.00$ $(-0.72)$ | 0.03 | $-0.00$ $(-1.33)$ | 0.02 | $-0.00$ $(-0.91)$ | 0.04 |
| Building permits | 0.11 (1.53) | 0.15 | 0.04 (0.88) | 0.00 | 0.08 (1.38) | 0.12 |
| UI claims | $-0.01$ $(-0.25)$ | 0.00 | 0.04 (1.10) | 0.01 | $-0.00$ $(-0.09)$ | 0.00 |
| Coincident Index | $-0.14^{**}$ $(-4.26)$ | 0.23 | $-0.08$ $(-1.19)$ | 0.02 | $-0.11^{**}$ $(-3.87)$ | 0.21 |
| **Panel (b). NASDAQ** | | | | | | |
| | $\mu_t$ | | $h_t$ | | $\mu_t + h_t$ | |
| | coefficient (t-stat) | $R^2$ | coefficient (t-stat) | $R^2$ | coefficient (t-stat) | $R^2$ |
| Interest spread | $-0.56^{**}$ $(-2.80)$ | 0.15 | $-0.45$ $(-0.97)$ | 0.01 | $-0.49^{*}$ $(-1.86)$ | 0.14 |
| Consumer sentiment | $3.94^{*}$ (1.68) | 0.15 | 1.01 (0.35) | 0.00 | 3.33 (1.59) | 0.13 |
| Money supply | $2.38^{**}$ (6.50) | 0.26 | $2.27^{*}$ (1.70) | 0.06 | $2.15^{**}$ (6.10) | 0.25 |
| Vendor performance | $-0.22^{*}$ $(-1.88)$ | 0.06 | $-0.01$ $(-0.41)$ | 0.00 | $-0.02$ $(-1.46)$ | 0.06 |
| New orders | $0.09^{**}$ (2.38) | 0.30 | 0.01 (0.21) | 0.00 | $0.07^{**}$ (2.49) | 0.25 |
| Work hours | $-0.00$ $(-1.36)$ | 0.10 | 0.00 (0.11) | 0.00 | $-0.00$ $(-1.22)$ | 0.08 |
| Building permits | 0.08 (1.53) | 0.16 | $-0.00$ $(-0.06)$ | 0.00 | $0.07^{*}$ (1.70) | 0.13 |
| UI claims | $-0.00$ $(-0.12)$ | 0.00 | 0.00 (0.00) | 0.00 | $-0.00$ $(-0.08)$ | 0.00 |
| Coincident Index | $-0.13^{**}$ $(-4.21)$ | 0.39 | $-0.04$ $(-0.51)$ | 0.00 | $-0.12^{**}$ $(-3.87)$ | 0.34 |

Note. The t-stat is computed with HAC standard errors. The bandwidth is determined using Andrews' (1991) method with the quadratic spectral kernel. * and ** denote significance at 10 and 5 percent level, respectively.

Table 4. Comovement between volatility components and business cycle indicators
(Sample period: 1992.2-2010.12)

Panel (a). S&P 500

| | $\mu_t$ | | $h_t$ | | $\mu_t + h_t$ | |
|---|---|---|---|---|---|---|
| | coefficient (t-stat) | $R^2$ | coefficient (t-stat) | $R^2$ | coefficient (t-stat) | $R^2$ |
| Interest spread | −0.16 (−0.37) | 0.01 | 0.17 (0.51) | 0.00 | −0.09 (−0.28) | 0.00 |
| Consumer sentiment | −2.80 (−0.51) | 0.03 | −4.04 (−1.25) | 0.01 | −2.59 (−0.65) | 0.03 |
| Money supply | 2.72** (5.30) | 0.20 | 3.05** (2.83) | 0.05 | 2.39** (5.71) | 0.21 |
| Vendor performance | −0.03** (−2.64) | 0.09 | −0.00 (−0.10) | 0.00 | −0.02** (−2.11) | 0.07 |
| New orders | 0.07 (0.96) | 0.10 | 0.01 (0.33) | 0.00 | 0.05 (1.04) | 0.08 |
| Work hours | −0.01** (−1.97) | 0.15 | −0.00 (−1.27) | 0.01 | −0.01** (2.11) | 0.13 |
| Building permits | −0.14 (−0.85) | 0.09 | −0.02 (−0.21) | 0.00 | −0.11 (−0.93) | 0.07 |
| UI claims | 0.08 (1.37) | 0.13 | 0.05 (1.34) | 0.01 | 0.07 (1.58) | 0.12 |
| Coincident Index | −0.19** (−4.47) | 0.28 | −0.12* (−1.71) | 0.02 | −0.16** (−4.29) | 0.25 |

Panel (b). NASDAQ

| | $\mu_t$ | | $h_t$ | | $\mu_t + h_t$ | |
|---|---|---|---|---|---|---|
| | coefficient (t-stat) | $R^2$ | coefficient (t-stat) | $R^2$ | coefficient (t-stat) | $R^2$ |
| Interest spread | −0.26 (−0.71) | 0.03 | −0.34 (−0.85) | 0.00 | −0.24 (−0.75) | 0.03 |
| Consumer sentiment | 0.68 (0.18) | 0.00 | −2.12 (−0.66) | 0.00 | 0.36 (0.10) | 0.00 |
| Money supply | 2.46** (6.00) | 0.21 | 2.80** (2.50) | 0.03 | 2.23** (6.14) | 0.22 |
| Vendor performance | −0.03** (−2.61) | 0.10 | −0.01 (−0.59) | 0.00 | −0.02** (−2.06) | 0.09 |
| New orders | 0.05 (1.06) | 0.08 | 0.03 (0.65) | 0.00 | 0.05 (1.25) | 0.07 |
| Work hours | −0.01** (−2.10) | 0.16 | 0.00 (0.21) | 0.00 | −0.01* (−1.84) | 0.12 |
| Building permits | −0.01 (−0.10) | 0.00 | −0.04 (−0.43) | 0.00 | −0.01 (−0.14) | 0.00 |
| UI claims | 0.04 (0.77) | 0.04 | 0.01 (0.26) | 0.00 | 0.03 (0.78) | 0.03 |
| Coincident Index | −0.15** (−4.18) | 0.24 | −0.11 (−1.41) | 0.01 | −0.13** (−4.16) | 0.22 |

Note. The t-stat is computed with HAC standard errors. The bandwidth is determined using Andrews' (1991) method with the quadratic spectral kernel. * and ** denote significance at 10 and 5 percent level, respectively.

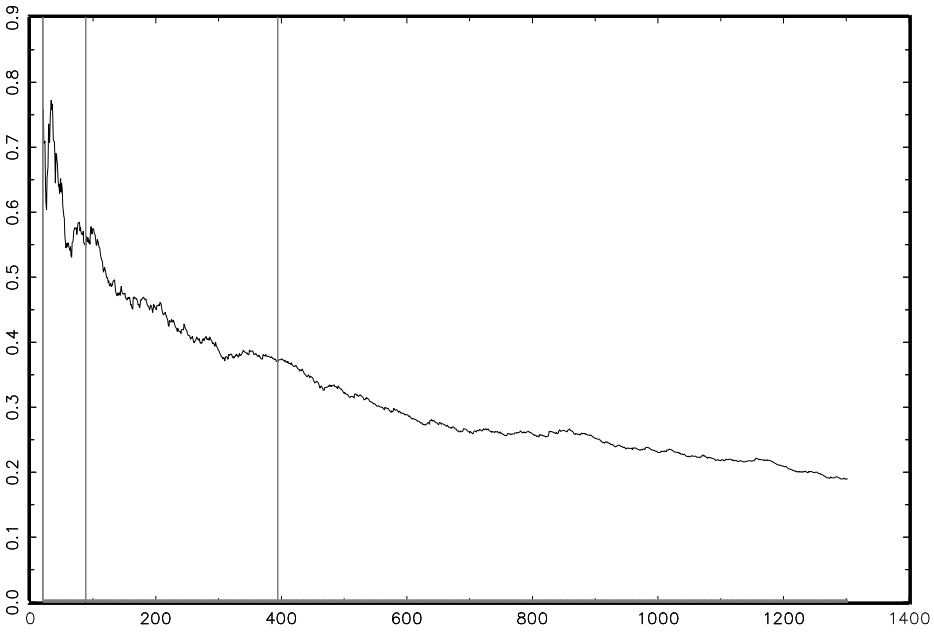
Table 5. Estimates from the forecasting regression (16)

| | | S&P 500 | | | | NASDAQ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $h$ | LS v.s. | $b_o$ | $b_1$ | $b_2$ | $\bar{R}^2$ | $b_o$ | $b_1$ | $b_2$ | $\bar{R}^2$ |
| 1 | SV-Rec | −0.21 (−0.57,0.16) | 1.11 (−2.12,4.33) | 0.13 (−3.54,3.80) | 0.30 | 0.33 (−0.05,0.70) | 1.00 (−0.42,1.58) | −0.33 (−0.94,0.28) | 0.25 |
| | SV-Rol | −0.20 (−0.56,0.17) | 1.10 (−2.14,4.35) | 0.12 (−3.29,3.52) | 0.30 | 0.32 (−0.09,0.72) | 0.74 (−0.28,1.51) | −0.02 (−0.87,0.83) | 0.25 |
| | FIG-Rec | −0.42 (−0.84,0.01) | 0.17 (−2.84,3.18) | 1.29 (−2.11,4.69) | 0.30 | 0.36 (−0.17,0.75) | 1.70 (0.72,2.67) | −1.17 (−2.20,−0.13) | 0.27 |
| | FIG-Rol | −0.51 (−0.88,−0.15) | −1.06 (−4.04,1.92) | 2.64 (−0.60,5.88) | 0.33 | 0.32 (−0.09,0.73) | 0.67 (−1.01,2.36) | 0.06 (−1.79,1.90) | 0.25 |
| 2 | SV-Rec | −0.05 (−0.53,0.43) | 0.85 (−0.44,2.14) | 0.18 (−1.48,1.83) | 0.33 | −0.35 (−1.22,0.53) | 1.13 (0.48,1.77) | −0.22 (−0.94,0.50) | 0.44 |
| | SV-Rol | 0.05 (−0.31,0.40) | 1.12 (−0.03,2.28) | −0.15 (−1.46,1.15) | 0.33 | −0.38 (−1.29,0.52) | 0.81 (0.21,1.42) | 0.16 (−0.57,0.89) | 0.44 |
| | FIG-Rec | 0.05 (−0.61,0.72) | 1.08 (−0.33,2.50) | −0.12 (−1.82,1.59) | 0.33 | −0.21 (−1.08,0.66) | 1.83 (0.90,2.76) | −1.08 (−2.12,−0.04) | 0.45 |
| | FIG-Rol | −0.09 (−0.77,0.59) | 0.56 (−1.56,2.68) | 0.49 (−1.93,0.33) | 0.33 | −0.39 (−1.25,0.47) | 1.22 (0.15,2.28) | −0.31 (−1.48,0.87) | 0.44 |
| 3 | SV-Rec | 0.23 (−0.44,0.90) | 0.85 (−0.10,1.80) | 0.45 (−1.24,1.33) | 0.37 | 0.01 (−1.07,1.08) | 0.94 (0.36,1.51) | −0.13 (−0.86,0.61) | 0.51 |
| | SV-Rol | 0.33 (−0.15,0.82) | 1.09 (0.19,2.00) | −0.24 (−1.32,0.83) | 0.38 | −0.05 (−1.13,1.04) | 0.65 (0.13,1.16) | 0.22 (−0.46,0.91) | 0.51 |
| | FIG-Rec | 0.38 (−0.46,1.21) | 1.08 (0.08,2.07) | −0.24 (−1.51,1.03) | 0.38 | 0.13 (−1.05,1.30) | 1.24 (0.41,2.06) | −0.50 (−1.55,0.55) | 0.52 |
| | FIG-Rol | 0.19 (−0.59,0.98) | 0.69 (−0.87,2.26) | 0.22 (−1.59,2.02) | 0.38 | −0.03 (−1.04,0.99) | 0.96 (0.09,1.83) | −0.15 (−1.14,0.84) | 0.51 |
| 4 | SV-Rec | 0.30 (−0.55,1.15) | 0.61 (−0.47,1.69) | 0.31 (−1.08,1.70) | 0.41 | −0.60 (−2.79,1.59) | 0.77 (0.35,1.20) | 0.22 (0.22,0.66) | 0.52 |
| | SV-Rol | 0.51 (−0.06,1.08) | 0.83 (−0.15,1.82) | 0.02 (−1.10,1.14) | 0.41 | −0.54 (−2.67,1.59) | 0.96 (−0.02,1.95) | −0.03 (−1.05,1.00) | 0.52 |
| | FIG-Rec | 0.60 (−0.26,1.45) | 0.95 (−0.15,2.04) | −0.12 (−1.41,1.18) | 0.41 | −0.61 (−2.84,1.62) | 0.82 (0.05,1.59) | 0.15 (−0.78,1.07) | 0.52 |
| | FIG-Rol | 0.57 (−0.09,1.24) | 0.99 (−0.27,2.24) | −0.15 (−1.50,1.20) | 0.41 | −0.54 (−2.69,1.61) | 0.93 (0.07,1.79) | 0.01 (−0.98,1.00) | 0.52 |
| 5 | SV-Rec | −0.14 (−0.88,0.59) | 0.30 (−0.40,1.01) | 0.73 (−0.19,1.65) | 0.50 | −0.35 (−2.77,2.08) | 0.76 (0.39,1.13) | 0.18 (−0.25,0.61) | 0.58 |
| | SV-Rol | 0.22 (−0.29,0.73) | 0.41 (0.23,1.05) | 0.53 (−0.20,1.26) | 0.50 | −0.27 (−2.61,2.06) | 0.91 (0.13,1.69) | −0.01 (−0.83,0.80) | 0.57 |
| | FIG-Rec | 0.09 (−0.65,0.84) | 0.45 (−0.24,1.13) | 0.51 (−0.30,1.33) | 0.50 | −0.56 (−2.96,1.85) | 0.55 (−0.05,1.15) | 0.44 (−0.23,1.12) | 0.58 |
| | FIG-Rol | 0.43 (−0.23,1.09) | 0.68 (−0.16,1.52) | 0.21 (−0.69,1.10) | 0.49 | −0.27 (−2.63,2.09) | 0.85 (0.03,1.68) | 0.05 (−1.00,1.09) | 0.57 |

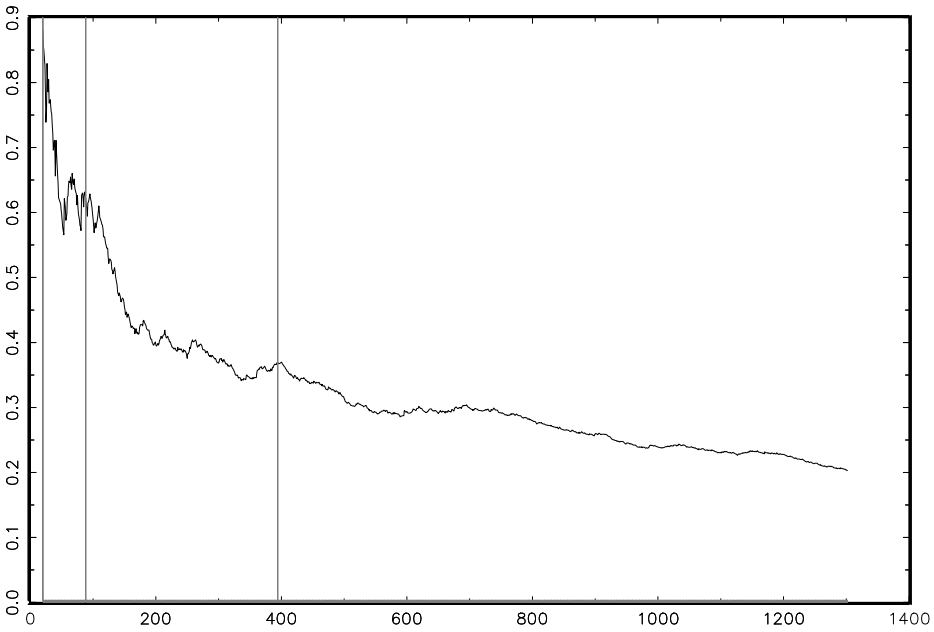
Note. The confidence interval is obtained using HAC standard errors.

Figure 1: The log-periodogram estimate of d as a function of m

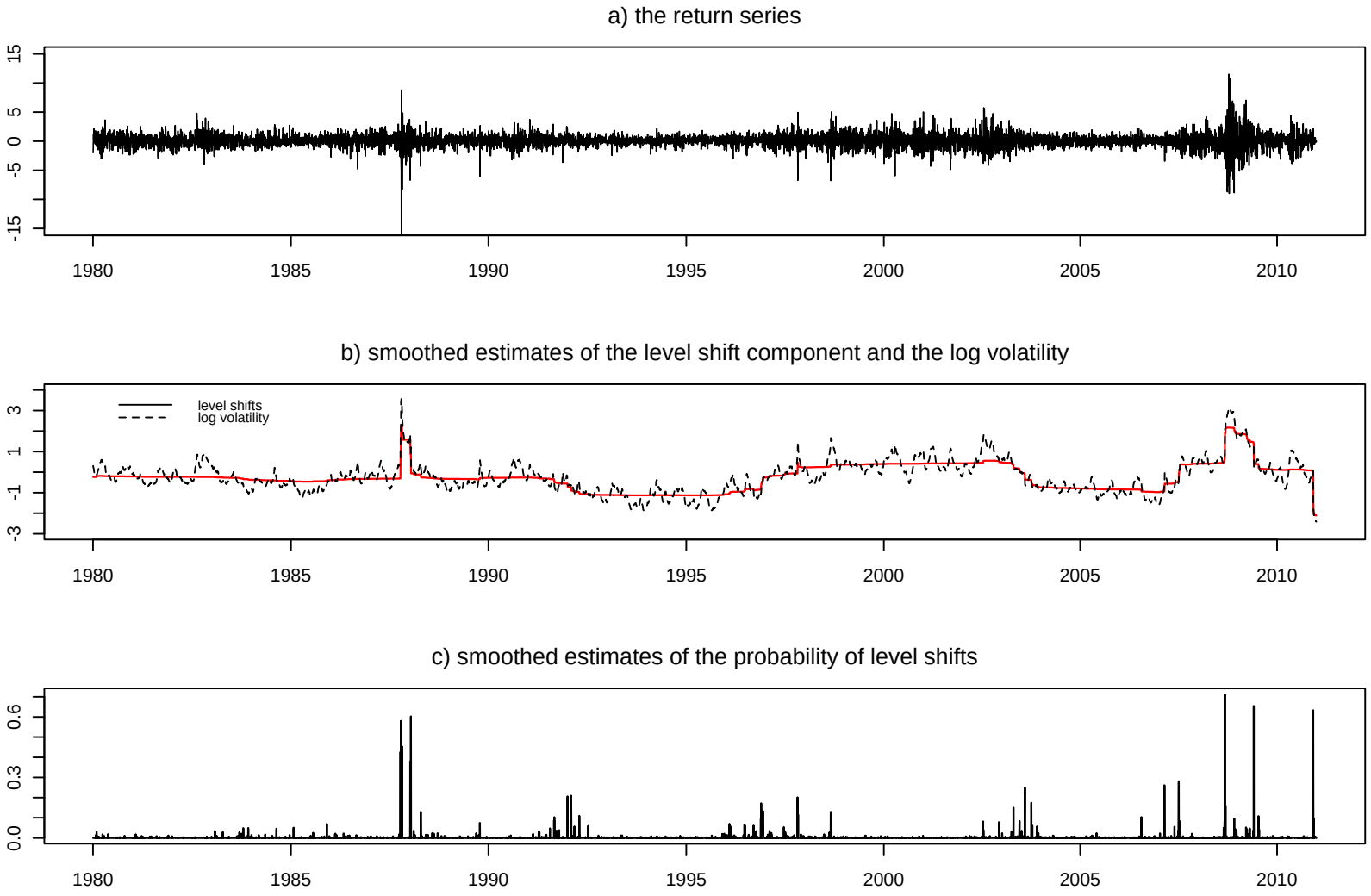
(a) Log squared S&P500 daily returns: 1980:1–2010:12



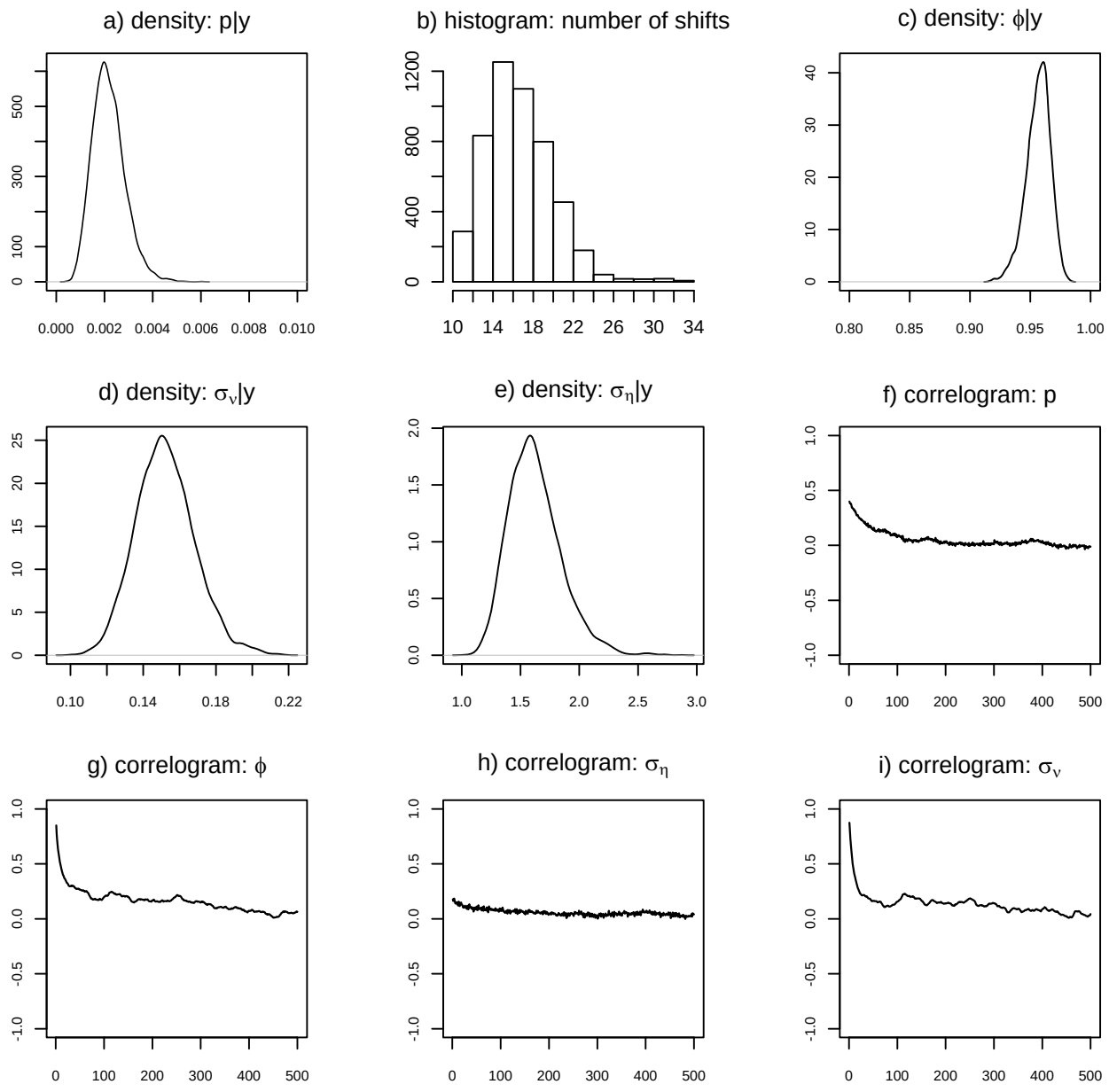

(a) Log squared NASDAQ daily returns: 1980:1–2010:12



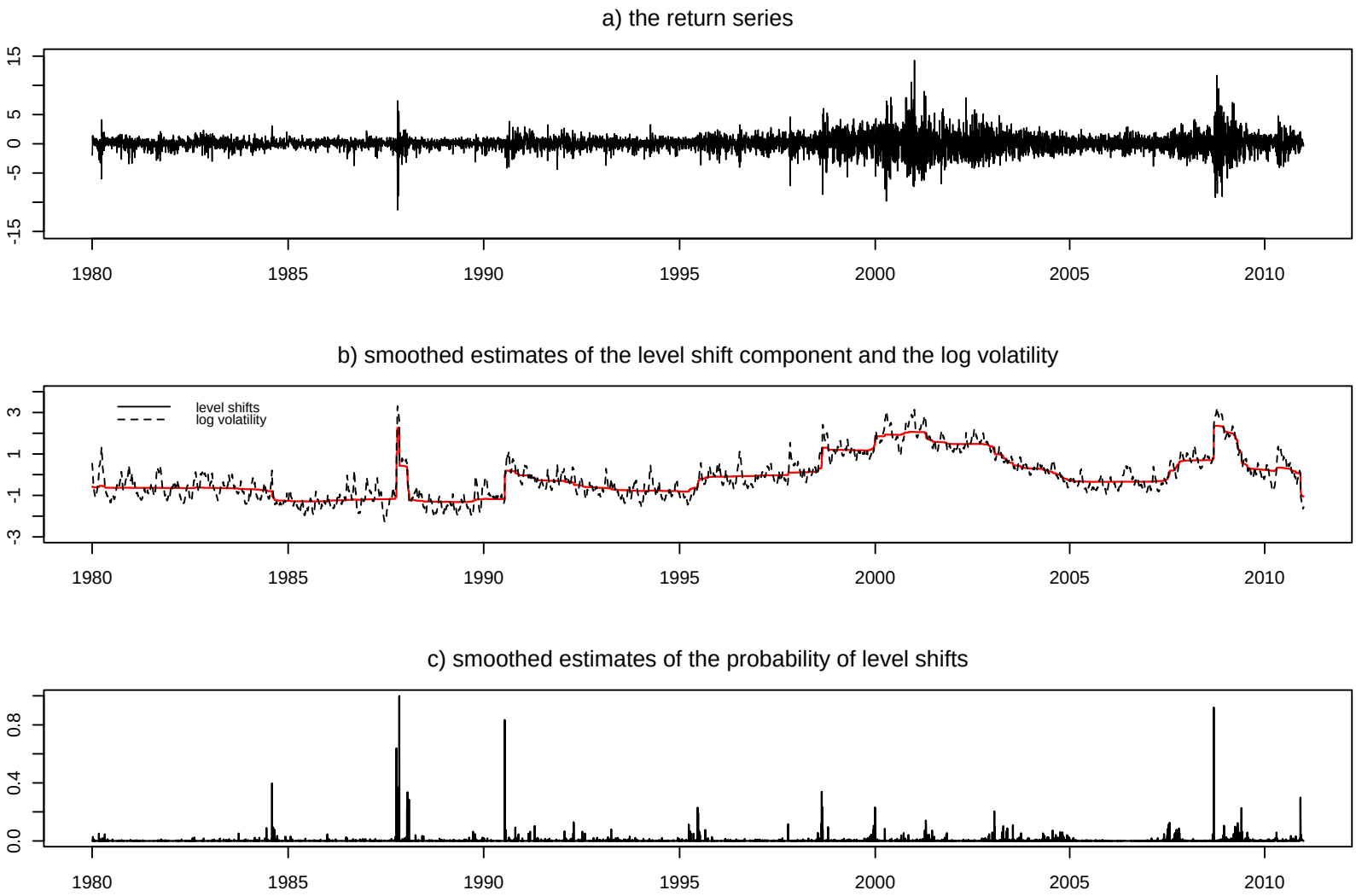
## Figure 2: Results for S&P500 volatility



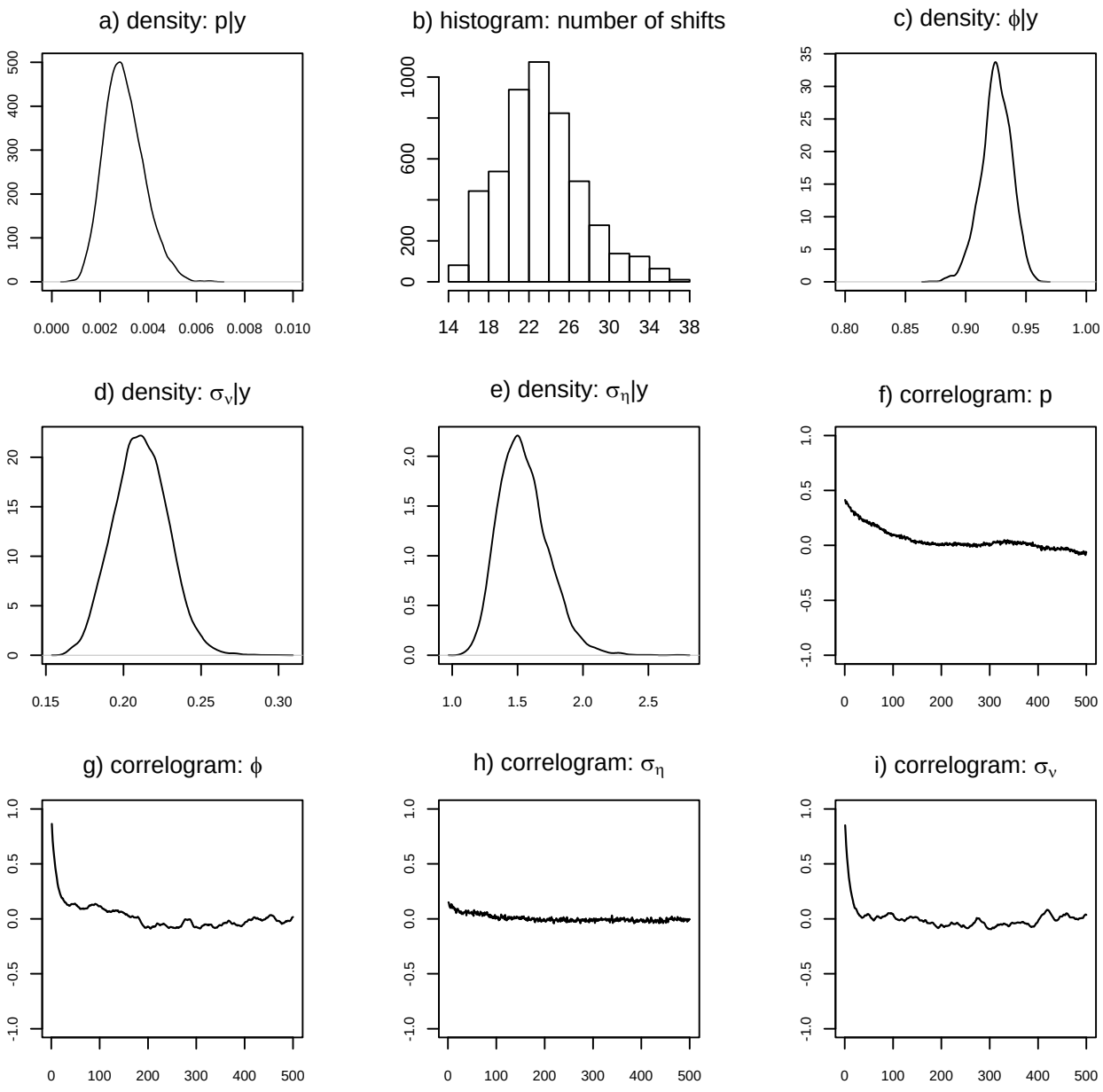
## Figure 3: Results for S&P500 volatility (cont'd)



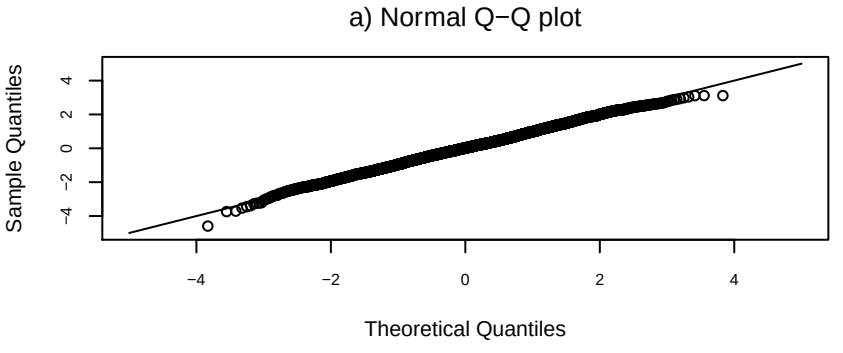
## Figure 4: Results for NASDAQ volatility

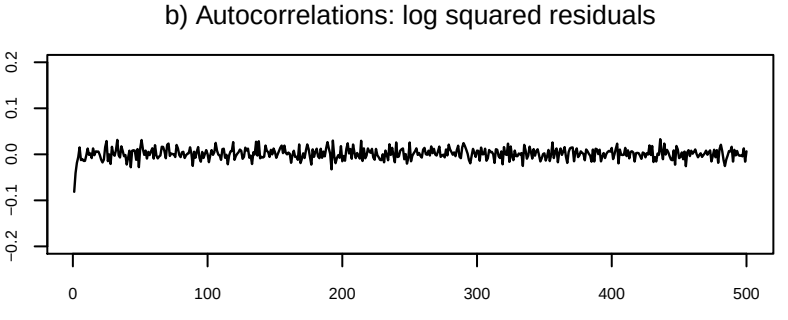


## Figure 5: Results for NASDAQ volatility (cont'd)



## Figure 6: Diagnostic results for S&P500



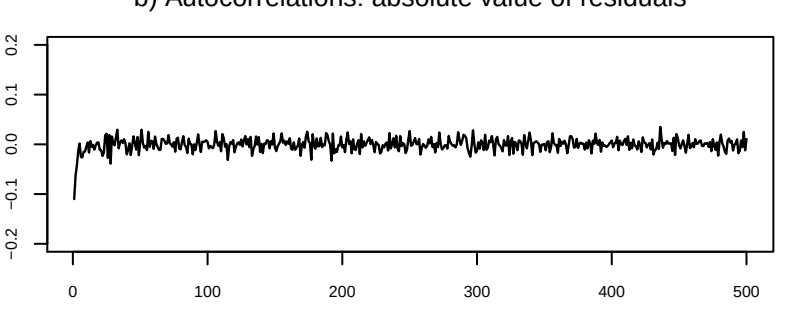



Figure 7: The mean of the log-periodogram estimate of d as a function of m for simulated log squared returns