

STRUCTURAL CHANGE AS AN ALTERNATIVE TO LONG MEMORY IN FINANCIAL TIME SERIES

Tze Leung Lai and Haipeng Xing

Stanford University and Columbia University

ABSTRACT

This paper shows that volatility persistence in GARCH models and spurious long memory in autoregressive models may arise if the possibility of structural changes is not incorporated in the time series model. It also describes a tractable hidden Markov model in which regression parameters and error variances may undergo abrupt changes at unknown time points, while staying constant between adjacent change-points. Applications to real and simulated financial time series are given to illustrate the issues and methods.

JEL Classifications: C11, C13, C22.

Corresponding author's (Lai's) address: Department of Statistics, Stanford University, Stanford, CA 94305-4065

Email: lait@stat.stanford.edu

Fax: 650-725-8977

I. INTRODUCTION

Volatility modeling is a cornerstone of empirical finance, as portfolio theory, asset pricing and hedging all involve volatilities. Since the seminal works of Engle (1982) and Bollerslev (1986), generalized autoregressive conditionally heteroskedastic (GARCH) models have been widely used to model and forecast volatilities of financial time series. In many empirical studies of stock returns and exchange rates, estimation of the parameters ω , α and β in the GARCH(1,1) model

$$y_n = \sigma_n \epsilon_n, \quad \sigma_n^2 = \omega + \alpha y_{n-1}^2 + \beta \sigma_{n-1}^2 \quad (1)$$

reveals high volatility persistence, with the maximum likelihood estimate $\widehat{\lambda} = \widehat{\alpha} + \widehat{\beta}$ close to 1. The constraint $\lambda < 1$ in the GARCH model (1), in which the ϵ_n are independent standard normal random variables that are not observable and the y_n are the observations, enables the conditional variance σ_n^2 of y_n given the past observations to be expressible as

$$\sigma_n^2 = (1 - \lambda)^{-1} \omega + \alpha \{(y_{n-1}^2 - \sigma_{n-1}^2) + \lambda(y_{n-2}^2 - \sigma_{n-2}^2) + \lambda^2(y_{n-3}^2 - \sigma_{n-3}^2) + \dots\}.$$

For $\lambda = 1$, the contributions of the past innovations $y_{n-t}^2 - \sigma_{n-t}^2$ to the conditional variance do not decay over time but are “integrated” instead, yielding the IGARCH model of Engle and Bollerslev (1986). Baille et al. (1996) introduced fractional integration in their FIGARCH models, with a slow hyperbolic rate of decay for the influence of the past innovations, to quantify the long memory of exchange rate volatilities.

In his comments on Engle and Bollerslev (1986), Diebold (1986) noted with respect to interest rate data that the choice of a constant term ω in (1) not accommodating to shifts in monetary policy regimes might have led to an apparently integrated series of innovations. By assuming ω to be piecewise constant and allowing the possibility of jumps between evenly spaced time intervals, Lamoureux and Lastrapes (1990) obtained smaller estimates of λ , from the daily returns of 30 stocks during the period 1963-1979, than those showing strong persistence based on the usual GARCH model with constant ω .

In Section II we carry out another empirical study of volatility persistence in the weekly returns of the NASDAQ index by updating the estimates of the parameters ω , α , β in the GARCH model (1) as new data arrive during the period January 1986 to September 2003, starting with an initial estimate based on the period November 1984 to December 1985. These sequential estimates show that the parameters changed over time and that $\widehat{\lambda}_N$ eventually became very close to 1 as the sample size N increased with accumulating data. The empirical study therefore suggests using a model that incorporates the possibility of structural

changes for these data. In Section III we describe a structural change model which allows changes in the volatility and regression parameters at unknown times and with unknown change magnitudes. It is a hidden Markov model, for which the volatility and regression parameters at time t based on data up to time t (or up to time $n > t$) can be estimated by recursive filters (or smoothers).

Although we have been concerned with volatility persistence in GARCH models so far, the issue of spurious long memory when the possibility of structural change is not incorporated into the time series model also arises in autoregressive and other models. In Section IV, by making use of the stationary distribution of a variant of the structural change model in Section III, we compute the asymptotic properties of the least squares estimates in a nominal AR(1) model that assumes constant parameters, thereby showing near unit root behavior of the estimated autoregressive parameter. Section V gives some concluding remarks and discusses related literature on structural breaks and the advantages of our approach.

II. VOLATILITY PERSISTENCE IN NASDAQ WEEKLY RETURNS

Figure 1, top panel, plots the weekly returns of the NASDAQ index, from the week starting on November 19, 1984 to the week starting on September 15, 2003. The series r_t is constructed from the closing price P_t on the last day of the week via $r_t = 100 \log(P_t/P_{t-1})$. The data consisting of 982 observations are available at <http://finance.yahoo.com>. Similar to Lamoureux and Lastrapes (1990, p. 227) who use lagged dependent variables to account for nonsynchronous trading, we fit an AR(2) model $r_n = \mu + \rho_1 r_{n-1} + \rho_2 r_{n-2} + y_n$ to the return series. The y_n are assumed to follow a GARCH(1,1) process so that the ϵ_n in (1) is standard normal and independent of $\{(\epsilon_i, y_i, r_i), i \leq n - 1\}$. The parameters of this AR-GARCH model can be estimated by maximum likelihood using `garchfit` in MATLAB. Besides using the full dataset, we also estimated these parameters sequentially from January 1986 to September 2003, starting with an initial estimate based on the period November 1984 to December 1985. These sequential estimates are plotted in Figure 2. Note that the estimates on the last date, September 15, 2003, corresponds to those based on the full dataset; see the last row of Table 1.

INSERT FIGURES 1 AND 2 ABOUT HERE

Lamoureux and Lastrapes (1990) proposed the following strategy to incorporate possible time variations in ω in fitting GARCH(1,1) models to a time series of $N = 4228$ observations of daily returns of each of 30 stocks during the period 1963-1979. They allow for possible jumps in ω every 302 observations, thereby replacing ω in (1) by $\omega' + \delta_1 D_{1n} + \dots + \delta_k D_{kn}$,

where D_{1n}, \dots, D_{kn} are dummy variables indicating the subsample to which n belongs. The choice of 302, 604, 906, ... as the only possible jump points of ω while not allowing jumps in α and β appears to be overly restrictive. In Section IIIC we fit a hidden Markov model of structural change to the observed time series and use the fitted model to estimate the probability of changes in both the autoregressive and volatility parameters at any time point. On the basis of these estimated probabilities, which are plotted in the bottom panel of Figure 1, we divide the time series of NASDAQ index returns into 4 segments of different lengths; see Section IIIC for further details. The estimates of the AR-GARCH parameters in each segment are compared with those based on the entire dataset in Table 1. Consecutive segments in Table 1 have slight overlap because we try not to initialize the returns with highly volatile data for a segment. Table 1 shows substantial differences in the estimated AR and GARCH parameters among the different time periods. In particular, $\hat{\lambda}$ is as small as 0.0364 in the first segment but as high as 0.9858 in the last segment. Although this seems to suggest volatility persistence during the last period of $5\frac{1}{2}$ years, we give an alternative viewpoint at the end of this section.

INSERT TABLE 1 ABOUT HERE

Using the estimates $\hat{\mu}$ of μ , $\hat{\rho}_1$ and $\hat{\rho}_2$ of the autoregressive parameters and $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\beta}$ of the GARCH parameters, we can compute the estimated mean return $\hat{\mu} + \hat{\rho}_1 r_{n-1} + \hat{\rho}_2 r_{n-2}$ at time n , the residual $\hat{y}_n = r_n - (\hat{\mu} + \hat{\rho}_1 r_{n-1} + \hat{\rho}_2 r_{n-2})$, and the volatility $\hat{\sigma}_n = (\hat{\omega} + \hat{\alpha} \hat{y}_{n-1}^2 + \hat{\beta} \hat{\sigma}_{n-1}^2)^{1/2}$ recursively. The results are plotted in Figures 3 and 4, with the top panel using parameter estimates based on the entire dataset and the middle panel using different parameter estimates for different segments of the data. The bottom panel of Figure 3 plots the estimated mean returns based on a hidden Markov model of structural change (see Section IIIB). The bottom panel of Figure 4 plots (i) the estimated volatilities (solid curve) based on the structural change model (see Section IIIB), and (ii) standard deviations (dotted curve) of the residuals in fitting an AR(2) model to the current and previous nineteen observations (see Section IIIC for further details).

INSERT FIGURES 3 AND 4 ABOUT HERE

Consider the last $5\frac{1}{2}$ year period in Table 1 that gives $\hat{\lambda} = 0.9858$. An alternative to the GARCH model, which is used for the bottom panel of Figure 4, is the structural change model which assumes that $y_n = \sigma_n \epsilon_n$ with σ_n undergoing periodic jumps according to a renewal process, unlike the continuously changing σ_n in the GARCH model (1). As the

σ_n are unknown, we replace them by their estimates $\hat{\sigma}_n$ (given by the solid curve in the bottom panel of Figure 4) and consider a simulated set of innovations $y_n^* = \hat{\sigma}_n \epsilon_n^*$ for this last segment of the data, in which ϵ_n^* are i.i.d. standard normal random variables. Fitting a nominal GARCH (1,1) model to these simulated innovations yielded a value $\hat{\lambda} = 0.9843$, which is also close to 1, and differs little from value of 0.9858 for the last period in Table 1.

III. AUTOREGRESSIVE MODELS WITH STRUCTURAL BREAKS IN VOLATILITY AND REGRESSION PARAMETERS

Bayesian modeling of structural breaks in autoregressive time series has been an active research topic in the last two decades. Hamilton (1989) proposed a regime-switching model which treats the autoregressive parameters as the hidden states of a finite-state Markov chain. Noting that volatility persistence of stock returns may be overestimated when one ignores possible structural changes in the time series of stock returns, Hamilton and Susmel (1994) extended regime switching for autoregression to a Markov-switching ARCH model, which allows the parameters of Engle's (1982) ARCH model of stock returns to oscillate among a few unspecified regimes, with transitions between regimes governed by a Markov chain. Albert and Chib (1993) considered autoregressive models with exogenous variables whose autoregressive parameters and error variances are subject to regime shifts determined by a two-state Markov chain with unknown transition probabilities. Wang and Zivot (2000) introduced a Bayesian time series model which assumes the number of the change-points to be known and in which multiple change points in level, trend and error variance are modeled using conjugate priors. Similar Bayesian autoregressive models allowing structural changes were introduced by Carlin, Gelfand and Smith (1992), McCulloch and Tsay (1993), Chib (1998) and Chib, Nardari and Shepard (2002). However, except for Hamilton's (1989) seminal regime-switching model, Markov Chain Monte Carlo techniques are needed in the statistical analysis of these models because of their analytic intractability.

Lai, Liu and Xing (2005) recently introduced the following model to incorporate possible structural breaks in the AR(k) process

$$X_n = \mu_n + \rho_{1n}X_{n-1} + \dots + \rho_{kn}X_{n-k} + \sigma_n \epsilon_n, \quad n > k, \quad (2)$$

where the ϵ_n are i.i.d. unobservable standard normal random variables, and $\boldsymbol{\theta}_n = (\mu_n, \rho_{1n}, \dots, \rho_{kn})^T$ and σ_n are piecewise constant parameters. Note that σ_n^2 is assumed to be piecewise constant in this model, unlike the continuous change specified by the linear difference equation (1). The sequence of change-times of $(\boldsymbol{\theta}_t^T, \sigma_t)$ is assumed to form a discrete renewal process with

parameter p , or equivalently,

$$I_t := 1_{\{(\boldsymbol{\theta}_t, \sigma_t) \neq (\boldsymbol{\theta}_{t-1}, \sigma_{t-1})\}} \text{ are i.i.d. Bernoulli random variables with } P(I_t = 1) = p \quad (3)$$

for $t \geq k + 2$ and $I_{k+1} = 1$. In addition, letting $\tau_t = (2\sigma_t^2)^{-1}$, it is assumed that

$$(\boldsymbol{\theta}_t^T, \tau_t) = (1 - I_t)(\boldsymbol{\theta}_{t-1}^T, \tau_{t-1}) + I_t(\mathbf{Z}_t^T, \gamma_t),$$

where $(\mathbf{Z}_1^T, \gamma_1), (\mathbf{Z}_2^T, \gamma_2), \dots$ are i.i.d. random vectors such that

$$\gamma_t \sim \text{Gamma}(g, \lambda), \quad \mathbf{Z}_t | \gamma_t \sim \text{Normal}(\mathbf{z}, \mathbf{V}/(2\gamma_t)). \quad (4)$$

An important advantage of this Bayesian model of structural breaks is its analytic and computational tractability. Explicit recursive formulas are available for estimating σ_n , μ_n , $\rho_{1,n}, \dots$, and $\rho_{k,n}$ in the Bayesian model (2)-(4). As we now proceed to show, these formulas also have an intuitively appealing interpretation as certain weighted averages of the estimates based on different segments of the data and assuming that there is no change-point in each segment.

A. Sequential Estimation of $\sigma_n, \mu_n, \rho_{1,n}, \dots, \rho_{k,n}$ via Recursive Filters

To estimate $(\boldsymbol{\theta}_n^T, \sigma_n^2)$ from current and past observations X_1, \dots, X_n , let $\mathbf{X}_{t,n} = (1, X_n, \dots, X_t)^T$ and consider the most recent change-time $J_n := \max\{t \leq n : I_t = 1\}$. Recalling that $\tau_n = (2\sigma_n^2)^{-1}$, the conditional distribution of $(\boldsymbol{\theta}_n^T, \tau_n)$ given $(J_n, \mathbf{X}_{J_n,n})$ can be described by

$$\tau_n \sim \text{Gamma}\left(g + \frac{n - J_n + 1}{2}, \frac{1}{a_{J_n,n}}\right), \quad \boldsymbol{\theta}_n | \tau_n \sim \text{Normal}\left(\mathbf{z}_{J_n,n}, \frac{1}{2\tau_n} \mathbf{V}_{J_n,n}\right), \quad (5)$$

where for $k < j \leq n$,

$$\begin{aligned} \mathbf{V}_{j,n} &= \left(\mathbf{V}^{-1} + \sum_{t=1}^n \mathbf{X}_{t-k,t-1} \mathbf{X}_{t-k,t-1}^T\right)^{-1}, \quad \mathbf{z}_{j,n} = \mathbf{V}_{j,n} \left(\mathbf{V}^{-1} \mathbf{z} + \sum_{t=1}^n \mathbf{X}_{t-k,t-1} X_t\right), \\ a_{j,n} &= \lambda^{-1} + \mathbf{z}^T \mathbf{V}^{-1} \mathbf{z} + \sum_{t=j}^n X_t^2 - \mathbf{z}_{j,n}^T \mathbf{V}_{j,n}^{-1} \mathbf{z}_{j,n}. \end{aligned} \quad (6)$$

Note that if $(2Y)^{-1}$ has a $\text{Gamma}(\tilde{g}, \tilde{\lambda})$ distribution, then Y has the inverse gamma $\text{IG}(g, \lambda)$ distribution with $g = \tilde{g}$, $\lambda = 2\tilde{\lambda}$, and that $EY = \lambda^{-1}(g - 1)^{-1}$ when $g > 1$ and $E\sqrt{Y} = \lambda^{-1/2}\Gamma(g - \frac{1}{2})/\Gamma(g)$. It then follows from (5) that

$$E(\boldsymbol{\theta}_n^T, \sigma_n^2 | \mathbf{X}_{1,n}) = \sum_{j=k+1}^n p_{j,n} E(\boldsymbol{\theta}_n^T, \sigma_n^2 | \mathbf{X}_{1,n}, J_n = j) = \sum_{j=k+1}^n p_{j,n} \left(\mathbf{z}_{j,n}^T, \frac{a_{j,n}}{2g + n - j - 1}\right), \quad (7)$$

where $p_{j,n} = P(J_n = j | \mathbf{X}_{1,n})$; see Lai, Liu and Xing (2005, p.282). Moreover, $E(\sigma_n | \mathbf{X}_{1,n}) = \sum_{j=k+1}^n p_{j,n} (a_{j,n}/2)^{1/2} \Gamma_{n-j}$, where $\Gamma_i = \Gamma(g + i/2)/\Gamma(g + (i + 1)/2)$. Denoting conditional densities by $f(\cdot | \cdot)$, the weights $p_{j,n}$ can be determined recursively by

$$p_{j,n} \propto p_{j,n}^* := \begin{cases} pf(X_n | J_n = j) & \text{if } j = n, \\ (1-p)p_{j,n-1}f(X_n | \mathbf{X}_{j-k,n-1}, J_n = j) & \text{if } j \leq n-1. \end{cases} \quad (8)$$

Since $\sum_{i=k+1}^n p_{i,n} = 1$, $p_{j,n} = p_{j,n}^* / \sum_{i=k+1}^n p_{i,n}^*$. Moreover, as shown in Lemma 1 of Lai, Liu and Xing (2005),

$$f(\lambda_j^{-1}(X_n - \mathbf{z}_{j,n-1}^T \mathbf{X}_{j-k,n-1}) | J_n = j, \mathbf{X}_{j-k,n-1}) = \text{Stud}(g_j) \quad (9)$$

where $\text{Stud}(g)$ denotes the Student- t density function with g degrees of freedom and

$$\begin{aligned} g_j &= 2g + n - j, \quad \lambda_j = a_{j,n}(1 + \mathbf{X}_{n-k+1,n-1}^T \mathbf{V} \mathbf{X}_{n-k+1,n-1}) / (2g + n - j) \quad \text{if } j < n; \\ g_j &= 2g, \quad \lambda_j = (1 + \mathbf{X}_{n-k+1,n-1}^T \mathbf{V} \mathbf{X}_{n-k+1,n-1}) / (2\lambda g) \quad \text{if } j = n. \end{aligned}$$

B. Estimating Piecewise Constant Parameters from a Time Series via Bayes Smoothers

Lai, Liu and Xing (2005) evaluate the minimum variance estimate $E(\tau_t, \boldsymbol{\theta}_t^T | X_1, \dots, X_n)$, with $k+1 \leq t \leq n$, by applying Bayes' theorem to combine the forward filter that involves the conditional distribution of $(\tau_t, \boldsymbol{\theta}_t^T)$ given X_1, \dots, X_t and the backward filter that involves the conditional distribution of $(\tau_t, \boldsymbol{\theta}_t^T)$ given X_{t+1}, \dots, X_n . Noting that the normal distribution for $\boldsymbol{\theta}_t$ assigns positive probability to the explosive region $\{\boldsymbol{\theta} = (\mu, \rho_1, \dots, \rho_k)^T : 1 - \rho_1 z - \dots - \rho_k z^k \text{ has roots inside the unit circle}\}$, they propose to replace the normal distribution in (4) by a truncated normal distribution that has support in some stability region C such that

$$\inf_{|z| \leq 1} |1 - \rho_1 z - \dots - \rho_k z^k| > 0 \quad \text{if } \boldsymbol{\theta} = (\mu, \rho_1, \dots, \rho_k)^T \in C. \quad (10)$$

Letting $\mathbf{T}_C \text{Normal}(\mathbf{z}, \mathbf{V})$ denote the conditional distribution of \mathbf{Z} given $\mathbf{Z} \in C$, where $\mathbf{Z} \sim \text{Normal}(\mathbf{z}, \mathbf{V})$ and C satisfies the stability condition (10), they modify (4) as

$$\gamma_t \sim \text{Gamma}(g, \lambda), \quad \mathbf{Z}_t | \gamma_t \sim \mathbf{T}_C \text{Normal}(\mathbf{z}, \mathbf{V} / (2\gamma_t)), \quad (11)$$

and show that $(\tau_t, \boldsymbol{\theta}_t^T, \mathbf{X}_{t-k+1,t})$ has a stationary distribution under which $(\boldsymbol{\theta}_t^T, \tau_t)$ has the same distribution as $(\mathbf{Z}_t^T, \gamma_t)$ in (11) and

$$X_t | (\boldsymbol{\theta}_t^T, \tau_t) \sim \text{Normal}(\mu_t / (1 - \rho_{1,t} - \dots - \rho_{k,t}), (2\tau_t)^{-1} v_t),$$

where $v_t = \sum_{j=0}^{\infty} \beta_{j,t}^2$ and $\beta_{j,t}$ are the coefficients in the power series representation of $1/(1 - \alpha_{1,t}z - \dots - \alpha_{k,t}z^k) = \sum_{j=0}^{\infty} \beta_{j,t}z^j$ for $|z| \leq 1$. In addition, they show that the Markov chain $(\tau_t, \boldsymbol{\theta}_t^T, \mathbf{X}_{t-k+1,t})$ is reversible if it is initialized at the stationary distribution, and therefore the backward filter of $(\tau_t, \boldsymbol{\theta}_t^T)$ based on X_n, \dots, X_{t+1} has the same structure as the forward predictor based on the past $n - t$ observations prior to t . Application of Bayes' theorem shows that the forward and backward filters can be combined to yield

$$f(\tau_t, \boldsymbol{\theta}_t | \mathbf{X}_{1,n}) \propto f(\tau_t, \boldsymbol{\theta}_t | \mathbf{X}_{1,t}) f(\tau_t, \boldsymbol{\theta}_t | \mathbf{X}_{t+1,n}) / \pi(\tau_t, \boldsymbol{\theta}_t) \quad (12)$$

where π denotes the stationary density function which is the same as that of (γ_t, \mathbf{Z}_t) given in (11); see Lai, Liu and Xing (2005, p.284).

Because the truncated normal is used in lieu of the normal distribution in (11), the conditional distribution of $\boldsymbol{\theta}_n$ given τ_n needs to be replaced by $\boldsymbol{\theta}_n | \tau_n \sim \mathbf{T}_C \text{Normal}(\mathbf{z}_{J_n,n}, \mathbf{V}_{J_n,n}/(2\tau_n))$, while (6) defining $\mathbf{V}_{j,n}, \mathbf{z}_{j,n}$ and $\alpha_{j,n}$ remains unchanged. For the implementation of the forward or backward filter, Lai, Liu and Xing (2005) propose to ignore the truncation, as the constraint set C only serves to generate non-explosive observations but has little effect on the values of the weights $p_{j,n}$. Analogous to (5)-(6), the conditional distribution of $(\boldsymbol{\theta}_t, \tau_t)$ given $J_t = i, \tilde{J}_t = j + 1$ and $\mathbf{X}_{i,j}$ for $i \leq t < j \leq n$ can be described by

$$\tau_t \sim \text{Gamma}\left(g + \frac{j - i + 1}{2}, \frac{1}{a_{i,j,t}}\right), \quad \boldsymbol{\theta}_t | \tau_t \sim \text{Normal}\left(\mathbf{z}_{i,j,t}, \frac{1}{2\tau_t} \mathbf{V}_{i,j,t}\right) \quad (13)$$

if we ignore the truncation in the truncated normal distribution in (11), where

$$\begin{aligned} \mathbf{V}_{i,j,t} &= (\mathbf{V}_{i,t}^{-1} + \tilde{\mathbf{V}}_{j,t}^{-1} - \mathbf{V}^{-1})^{-1}, \\ \mathbf{z}_{i,j,t} &= \mathbf{V}_{i,j,t} (\mathbf{V}_{i,j,t}^{-1} \mathbf{z}_{i,t} + \tilde{\mathbf{V}}_{j,t}^{-1} \tilde{\mathbf{z}}_{j,t+1} - \mathbf{V}^{-1} \mathbf{z}), \\ a_{i,j,t} &= \lambda^{-1} + \mathbf{z}^T \mathbf{V}^{-1} \mathbf{z} + \sum_{l=i}^j X_l^2 - \mathbf{z}_{i,j,t}^T \mathbf{V}_{i,j,t}^{-1} \mathbf{z}_{i,j,t}, \end{aligned}$$

in which $\mathbf{V}_{i,t}, \mathbf{z}_{i,t}$ and $a_{i,t}$ are defined in (6) and $\tilde{\mathbf{V}}_{j,t}, \tilde{\mathbf{z}}_{j,t}$ and $\tilde{a}_{j,t}$ are defined similarly by reversing time. Let $|\cdot|$ denote the determinant of a matrix,

$$\begin{aligned} b_{i,j,t} &= \left(\frac{|\mathbf{V}_{i,t}| |\tilde{\mathbf{V}}_{j,t}|}{|\mathbf{V}| |\mathbf{V}_{i,j,t}|} \right)^{-1/2} \left\{ \frac{\Gamma(g) \Gamma(g + \frac{1}{2}(j - i + 1))}{\Gamma(g + \frac{1}{2}(t - i + 1)) \Gamma(g + \frac{1}{2}(j - t))} \right\} \frac{a_{i,t}^{g+(t-i+1)/2} \tilde{a}_{j,t}^{g+(j-t)/2}}{a^g a_{i,j,t}^{g+(j-i+1)/2}}, \\ B_t &= p + (1 - p) \sum_{k+1 \leq i \leq t < j \leq n} p_{i,t} \tilde{p}_{j,t} b_{i,j,t}. \end{aligned}$$

Using (12) and (13), Lai, Liu and Xing (2005, p.288) have shown that analogous to (7),

$$\begin{aligned}
E(\sigma_t^2|\mathbf{X}_{1,n}) &\doteq \frac{p}{B_t} \sum_{i=k+1}^t \frac{p_{i,t} a_{i,t}}{2g+t-i-1} + \frac{1-p}{B_t} \sum_{k+1 \leq i \leq t < j \leq n} p_{i,t} \tilde{p}_{j,t} b_{i,j,t} \frac{a_{i,j,t}}{2g+j-i+1}, \\
E(\boldsymbol{\theta}_t|\mathbf{X}_{1,n}) &\doteq \frac{p}{B_t} \sum_{i=k+1}^t p_{i,t} \mathbf{z}_{i,t} + \frac{1-p}{B_t} \sum_{k+1 \leq i \leq t < j \leq n} p_{i,t} \tilde{p}_{j,t} b_{i,j,t} \mathbf{z}_{i,j,t},
\end{aligned} \tag{14}$$

in which the approximation ignores truncation within C . Moreover, the conditional probability of a structural break at time t ($\leq n$) given $\mathbf{X}_{1,n}$ is

$$P\{I_t = 1|\mathbf{X}_{1,n}\} = \sum_{i=k+1}^t P\{I_t = 1, J_{t-1} = i|\mathbf{X}_{1,n}\} \doteq p/B_t. \tag{15}$$

Although the Bayes filter uses a recursive updating formula (8) for the weights $p_{j,n}$ ($k < j \leq n$), the number of weights increases with n , resulting in unbounded computational and memory requirements in estimating σ_n and $\boldsymbol{\theta}_n$ as n keeps increasing. Lai, Liu and Xing (2005) propose a bounded complexity mixture (BCMIX) approximation to the optimal filter (7) by keeping the most recent m_p weights together with the largest $n_p - m_p$ of the remaining weights at every stage n (which is tantamount to setting the other $n - n_p$ weights to be 0), where $0 \leq m_p < n_p$. For the forward and backward filters in the Bayes smoothers, we can again use these BCMIX approximations, yielding a BCMIX smoother that approximates (14) by allowing at most n_p weights $p_{i,t}$ and n_p weights $\tilde{p}_{j,t}$ to be nonzero. In particular, the Bayes estimates of the time-varying mean returns and volatilities in the bottom panels of Figure 3 and 4 use $n_p = 40$ and $m_p = 25$.

C. Segmentation via Estimated Probabilities of Structural Breaks

The plot of the estimated probabilities of structural breaks over time in the bottom panel of Figure 1 for the NASDAQ weekly returns provides a natural data segmentation procedure that locates possible change-points at times when these estimated probabilities are relatively high. These probabilities are computed via (15), in which the hyperparameters $p, \mathbf{z}, \mathbf{V}, g$ and λ of the change-point autoregressive model (2)-(4) are determined as follows.

First we fit an AR(2) model to $\{r_t, r_{t-1}, \dots, r_{t-19}\}$, $\max(k+1, 20) \leq t \leq n$, yielding $\tilde{\mu}_t, \tilde{\rho}_{1t}, \tilde{\rho}_{2t}$ and $\tilde{\sigma}_t^2 = t^{-1} \sum_{j=t-19}^t (r_j - \tilde{\mu}_t - \tilde{\rho}_{1t} r_{j-1} - \tilde{\rho}_{2t} r_{j-2})^2$, which are the ‘‘moving window estimates’’ (with window size 20) of $\mu_t, \rho_{1t}, \rho_{2t}$ and σ_t^2 . The dotted curve in the bottom panel of Figure 4 is a plot of the $\tilde{\sigma}_t$ series. Then we estimate \mathbf{z} and \mathbf{V} by the sample mean $\hat{\mathbf{z}}$ and the sample covariance matrix $\hat{\mathbf{V}}$ of $\{(\tilde{\mu}_t, \tilde{\rho}_{1t}, \tilde{\rho}_{2t})^T, \max(k+1, 20) \leq t \leq n\}$, and

apply the method of moments to estimate g and λ of (4). Specifically, regarding $(2\tilde{\sigma}_t^2)^{-1}$ as a sample from the Gamma(g, λ) distribution, $E(\tilde{\sigma}_t^2) = (2\lambda)^{-1}(g-1)^{-1}$ and $\text{Var}(\tilde{\sigma}_t^2) = (2\lambda)^{-2}(g-1)^{-2}(g-2)^{-1}$, and using the sample mean and the sample variance of the $\tilde{\sigma}_t^2$ to replace their population counterparts yields the following estimates of g and λ :

$$\begin{aligned}\hat{g} &= 2.4479, & \hat{\lambda} &= 0.03960, \\ \hat{\mathbf{z}} &= (0.1304 \quad 0.003722 \quad -0.008111)^T, \\ \hat{\mathbf{V}} &= \begin{pmatrix} 0.06096 & -0.009412 & 0.0003295 \\ -0.009412 & 0.01864 & 0.004521 \\ 0.0003295 & 0.004521 & 0.01023 \end{pmatrix}.\end{aligned}$$

With the hyperparameters \mathbf{z} , \mathbf{V} , g and λ thus chosen, we can estimate the hyperparameter p by maximum likelihood, noting that

$$\begin{aligned}f(X_t|\mathbf{X}_{1,t-1}) &= pf(X_t|J_t = t) + (1-p) \sum_{j=k+1}^{t-1} P(J_{t-1} = j|\mathbf{X}_{1,t-1})f(X_t|\mathbf{X}_{1,t-1}, J_{t-1} = j) \\ &= \sum_{j=k+1}^t p_{j,t}^*,\end{aligned}$$

in which $p_{j,t}^*$ is defined by (8) and is a function of p . Hence the log-likelihood function is

$$l(p) = \log \left(\prod_{t=k+1}^n f(X_t|\mathbf{X}_{1,t-1}) \right) = \sum_{t=k+1}^n \log \left(\sum_{j=k+1}^t p_{j,t}^* \right).$$

Figure 5 plots the log-likelihood function thus defined for the NASDAQ weekly returns from November 19, 1984 to September 15, 2003, giving the maximum likelihood estimate $\hat{p} = 0.027$ of the hyperparameter p .

INSERT FIGURE 5 ABOUT HERE

IV. NEAR UNIT ROOT BEHAVIOR OF LEAST SQUARES ESTIMATES IN A CHANGE-POINT AUTOREGRESSIVE MODEL

We have shown in Section II that volatility persistence in GARCH models may arise if the possibility of structural changes is ignored. Similarly, spurious unit root in autoregressive models may also arise if one omits the possibility of structural changes. There is an extensive literature on the interplay between unit root (or more general long memory) and structural breaks; see Perron (1989), Perron and Vogelsang (1992), Zivot and Andrews (1992), Lumsdaine and Papell (1997), Maddala et al. (1999), Granger and Hyund (1999), Diebold and

Inoue (2001) and the special issues of Journal of Business and Economic Statistics (1992) and Journal of Econometrics (1996).

In this section we use the hidden Markov model of structural changes similar to that in Section III to demonstrate that spurious long memory can arise when the possibility of structural change is not incorporated. Consider the autoregressive model with occasional mean shifts

$$X_n = \mu_n + \rho X_{n-1} + \epsilon_n, \quad (16)$$

where ϵ_n are i.i.d. unobservable standard normal random variables and are independent of $I_n := 1_{\{\mu_n \neq \mu_{n-1}\}}$ which are i.i.d. Bernoulli random variables with $P(I_n = 1) = p$, as in (3), and

$$\mu_n = (1 - I_n)\mu_{n-1} + I_n Z_n, \quad (17)$$

with i.i.d. normal Z_n that have common mean μ and variance ν . The autoregressive parameter ρ is assumed to be less than 1 in absolute value but is unknown. The methods in Section IIIA or B can be readily modified to estimate μ_n and ρ , either sequentially or from a given set of data $\{X_n, 1 \leq n \leq T\}$.

Without incorporating possible shifts in μ , suppose we fit an AR(1) model $X_n = \mu + \rho X_{n-1} + \epsilon_n$ to $\{X_n, 1 \leq n \leq T\}$. The maximum likelihood estimate $(\hat{\mu}, \hat{\rho})$ is the same as the least squares estimate; in particular,

$$\hat{\rho} = \frac{\sum_{i=2}^T X_i X_{i-1} - \frac{1}{T-1} \left(\sum_{i=2}^T X_i \right) \left(\sum_{i=1}^{T-1} X_i \right)}{\sum_{i=1}^{T-1} X_i^2 - \frac{1}{T-1} \left(\sum_{i=1}^{T-1} X_i \right)^2}. \quad (18)$$

Since $|\rho| < 1$, the Markov chain (μ_n, X_n) defined by (16)-(17) has a stationary distribution π under which μ_n has the $N(\mu, \nu)$ distribution and $(1 - \rho)X_n$ has the $N(\mu, 1 + \nu)$ distribution. Hence

$$E_\pi(X_n) = \mu/(1 - \rho), \quad E_\pi(X_n^2) = (1 + \nu + \mu^2)/(1 - \rho)^2, \quad E_\pi(\mu_n) = \mu. \quad (19)$$

Making use of (19) and the stationary distribution of (μ_n, X_n) , we prove in the Appendix the following result which shows that the least squares $\hat{\rho}$ in the nominal model that ignores structural breaks can approach 1 as $n \rightarrow \infty$ even though $|\rho| < 1$.

Theorem: *Suppose that in (16)-(17), $(1 - p)\rho = 1 - p\nu$. Then $\hat{\rho} \rightarrow 1$ with probability 1.*

Table 2 reports the results of a simulation study on the behavior of the least squares estimate $\hat{\rho}$ in fitting an AR(1) model with constant μ to $\{X_n, 1 \leq n \leq T\}$, in which X_n is

generated from (16) that allows periodic jumps in μ . Unless stated otherwise, $T = 1000$. Each result in Table 2 is the average of 100 simulations from the change-point model (16), with the standard error included in parentheses. The table considers different values of the parameters p , ρ and v of the change-point model and also lists the corresponding values of the ratio $(1-p)\rho/(1-pv)$ in the preceding theorem. Table 2 shows that $\hat{\rho}$ is close to 1 when this ratio is 1, in agreement with the theorem. It also shows that $\hat{\rho}$ can be quite close to 1 even when this ratio differs substantially from 1. Also given in Table 2 are the p-values (computed by `PP.test` in R) of the Phillips-Perron test of the null hypothesis that $\rho = 1$; see Perron (1988). These results show that near unit root behavior of $\hat{\rho}$ occurs quite frequently if the possibility of structural change is not taken into consideration.

INSERT TABLE 2 ABOUT HERE

V. CONCLUDING REMARKS

Financial time series modeling should incorporate the possibility of structural changes, especially when the series stretches over a long period of time. Volatility persistence or long memory behavior of the parameter estimates typically occurs if one ignores possible structural changes, as shown by the empirical study in Section II and the asymptotic theory in Section IV. Another way to look at such long memory behavior is through “multiresolution analysis” or “multiple time scales”, in which the much slower time scale for parameter changes (causing “long memory”) is overlaid on the short-run fluctuations of the time series.

The structural change model in Section III provides a powerful and tractable method to capture structural changes in both the volatility and regression parameters. It is a hidden Markov model (HMM), with the unknown regression and volatility parameters θ_t and σ_t undergoing Markovian jump dynamics, so that estimation of θ_t and σ_t can be treated as filtering and smoothing problems in the HMM. Making use of the special structure of the HMM that involves gamma-normal conjugate priors, there are explicit recursive formulas for the optimal filters and smoothers. Besides the BCMIX approximation described in Section IIIB, another approach, introduced in Lai, Liu and Xing (2005), is based on Monte Carlo simulations and involves sequential importance sampling with resampling. In Section IIIC, we have shown how the optimal structural change model can be applied to segment financial time series by making use of the estimated probabilities of structural breaks.

There is an extensive literature on tests for structural breaks in static regression models. Quandt (1960) and Kim and Siegmund (1989) considered likelihood ratio tests to detect a change-point in simple linear regression, and described approximations to the significance

level. The issue of multiple change-points was addressed by Bai (1999), Bai and Perron (1998), and Bai et. al (1998). The computational complexity of the likelihood statistic and the associated significance level increase rapidly with the prescribed maximum number of change-points. Extension from static to dynamic regression models entails substantially greater complexity in the literature on tests for structural breaks in autoregressive models involving unit root and lagged variables; see Kramer et al. (1988), Banerjee, et al. (1992), Dufour and Kiviet (1996), Harvey et al. (2004) and the references therein. The posterior probabilities of change-points given in Section IIIB for the Bayesian model (2)-(4), which incorporates changes not only in the autoregressive parameters but also in the error variances, provides a much more tractable Bayes test for structural breaks, allowing multiple change-points. The frequentist properties of the test are given in Lai and Xing (2005), where it is shown that the Bayes test provides a tractable approximation to the likelihood ratio test.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation.

REFERENCES

- Albert, J.H., Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* 11, 1-15.
- Bai, J., 1999. Likelihood ratio tests for multiple structural changes. *Journal of Econometrics* 91, 299-323.
- Bai, J., Lumsdaine, R.L., Stock, J.H., 1998. Testing for and dating common breaks in multivariate time series. *Review of Economic Studies* 65, 395-432.
- Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47-78.
- Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3-30.
- Banerjee, A., Lumsdaine, R.L., Stock, J.H., 1992. Recursive and sequential tests of the unit-root and trend-break hypothesis: Theory and international evidence. *Journal of Business and Economic Statistics* 10, 271-287.

- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.
- Carlin, B.P., Gelfand, A. E., Smith, A. F. M., 1992. Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics* 41, 389-405.
- Chib, S., 1998. Estimation and comparison of multiple change-point models. *Journal of Econometrics* 86, 221-241.
- Chib, S., Nardari, F., Shepard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108, 281-316.
- Diebold, F.X., 1986. Comment on modeling the persistence of conditional variance. *Econometric Reviews* 5, 51-56.
- Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. *Journal of Econometrics* 105, 131-159.
- Ding, A., Granger, C.W.J., 1996. Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics* 73, 185-211.
- Doufour, J.-M., and Kivlet, J.F., 1996. Exact tests for structural change in first-order dynamic models. *Journal of Econometrics* 70, 39-68.
- Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1007.
- Engle, R.F., Bollerslev, T., 1986. Modeling the persistence of conditional variances. *Econometric Reviews* 5, 1-50.
- Granger, C.W.J., Hyung, N., 1999. Occasional structural breaks and long memory. UCSD Discussion Paper 99-14.
- Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357-384.
- Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307-333.
- Harvey, D.L., Leybourne, S.J., and Newbold, P., 2004. Tests for a break in level when the order of integration is unknown. *Oxford Bulletin of Economics and Statistics*, 66, 133-146.

- Kim, H., Siegmund, O.D., 1989. The likelihood ratio test for a change-point in simple linear regression. *Biometrika* 76, 409-423.
- Kramer, W., Plogerger, W., Alt, R., 1988. Testing for structural change in dynamic models. *Econometrica* 56, 1355-1369.
- Lai, T.L., Liu, H., Xing, H., 2005. Autoregressive models with piecewise constant volatility and regression parameters. *Statistica Sinica* 15, 279-301.
- Lai, T.L., Xing, H., 2005. Hidden Markov models for structural changes and stochastic cycles in regression and time series. Working paper, Department of Statistics, Stanford University.
- Lamoureux, C.G., Lastrapas, W.D., 1990. Persistence in variance, structural change and the GARCH model. *Journal of Business and Economic Statistics* 8, 225-234.
- Lumsdaine, R.L., Papell, D.H., 1997. Multiple trend breaks and the unit root hypothesis. *Review of Economics and Statistics*, 79, 212-218.
- Maddala, G.S., Kim, I., Phillips, P.C.B., 1999. *Unit Roots, Cointegration, and Structural Change*. Cambridge, New York.
- McCulloch, R. E., Tsay, R. S., 1993. Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association* 88, 968-978.
- Meyn, S.P., Tweedie, R.L., 1993. *Markov Chains and Stochastic Stability*. Springer-Verlag, New York.
- Perron, P., 1988. Trends and random walks in macroeconomic time series. *Journal of Economic Dynamics and Control* 12, 297-332.
- Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57, 1361-1401.
- Perron, P., and Vogelsang, T.J., 1992. Nonstationarity and level shifts with an application to purchasing power parity. *Journal of Business and Economic Statistics*, 10, 301-320.
- Quandt, R.E., 1960. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association* 55, 324-330.
- Wang, J., Zivot, E., 2000. A Bayesian time series model of multiple structural changes in level, trend and variance. *Journal of Business and Economic Statistics* 18, 374-386.

Zivot, E., Andrews, D.W.K., 1992. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics* 10, 251-270.

APPENDIX

The proof of the theorem makes use of (18), (19) and

$$\text{Cov}_\pi(\mu_{n+1}, X_n) = (1-p)v/\{1-(1-p)\rho\}. \quad (20)$$

To prove (20), first note from (16) that

$$\begin{aligned} \text{Cov}_\pi(\mu_n, X_n) &= v + \rho \text{Cov}(\mu_n, X_{n-1}) \\ &= v + \rho\{E(\mu_n X_{n-1}) - E(\mu_n)E(X_{n-1})\} \\ &= v + (1-p)\rho \text{Cov}_\pi(\mu_{n-1}, X_{n-1}). \end{aligned} \quad (21)$$

The last equality in (21) follows from (17) and that $E_\pi(I_n Z_n) = p\mu$. The recursive representation of $\text{Cov}_\pi(\mu_n, X_n)$ in (21) yields

$$\text{Cov}_\pi(\mu_n, X_n) = v/\{1-(1-p)\rho\}.$$

Combining this with the first equality in (21) establishes (20). In view of (21) and the strong law for Markov random walks (cf. Theorem 17.1.7 of Meyn and Tweedie (1993)),

$$\begin{aligned} \sum_{i=1}^{T-1} X_i/T &\rightarrow E_\pi(X_n), \\ \left(\sum_{i=1}^{T-1} X_{i+1}X_i \right) / T &= \left\{ \sum_{i=1}^{T-1} \mu_{i+1}X_i + \rho \sum_{i=1}^{T-1} X_i^2 + \sum_{i=1}^{T-1} \epsilon_{i+1}X_i \right\} / T \\ &\rightarrow E_\pi(\mu_{n+1}X_n) + \rho E_\pi(X_n^2) = \text{Cov}_\pi(\mu_{n+1}, X_n) + E_\pi(\mu_{n+1})E_\pi(X_n) + \rho E_\pi(X_n^2), \end{aligned} \quad (22)$$

with probability 1, noting that $E_\pi(X_n \epsilon_{n+1}) = 0$. Putting (22) and (19), (20) into (18) shows that with probability 1, $\hat{\rho}$ converges to $\rho + \{v(1-p)(1-\rho)^2\}/\{(1+v)[1-(1-p)\rho]\}$, which is equal to 1 if $\rho = (1-pv)/(1-p)$.

Table 1. Estimated AR(2)-GARCH(1,1) Parameters for Different Periods

Period	μ	ρ_1	ρ_2	ω	α	β	λ
Nov 19, 1984 – Jun 16, 1987	0.3328	0.3091	0.09838	1.7453	0.0364	0.0	0.0364
Jun 09, 1987 – Aug 15, 1990	0.1242	0.2068	0.0891	0.8364	0.4645	0.4784	0.9429
Aug 08, 1990 – Mar 13, 1998	0.3362	0.03874	0.1039	0.4560	0.08767	0.8008	0.8885
Mar 06, 1998 – Sep 15, 2003	0.2785	0.05612	0.04063	0.8777	0.2039	0.7819	0.9858
Nov 19, 1984 – Sep 15, 2003	0.3203	0.09241	0.07657	0.2335	0.2243	0.7720	0.9963

Table 2. Means and Standard Errors (in Parentheses) of Least Squares Estimates and P-Values of Unit Root Test

p	v	ρ	$\frac{(1-p)\rho}{1-pv}$	$\hat{\rho}$	p-value
0.002	4	0.50	0.5030	0.8628 (0.0180)	0.1023 (0.0174)
0.002	8	0.20	0.2028	0.7591 (0.0320)	0.1595 (0.0240)
0.002	8	0.84	0.8520	0.9730 (0.0052)	0.4443 (0.0371)
0.002	16	0.50	0.5155	0.9134 (0.0174)	0.3818 (0.0329)
0.004	8	0.50	0.5145	0.9505 (0.0126)	0.3505 (0.0318)
0.008	8	0.50	0.5299	0.9891 (0.0007)	0.3514 (0.0252)
0.01	16	84/99	1	0.9978 (0.0002)	0.6110 (0.0252)
0.01	16	84/99	1	0.9986 (0.0001)	0.4241 (0.0201) (T=2000)
0.01	16	84/99	1	0.9988 (0.0001)	0.2585 (0.0193) (T=3000)
0.01	16	84/99	1	0.9988 (0.0000)	0.0811 (0.0090) (T=4000)

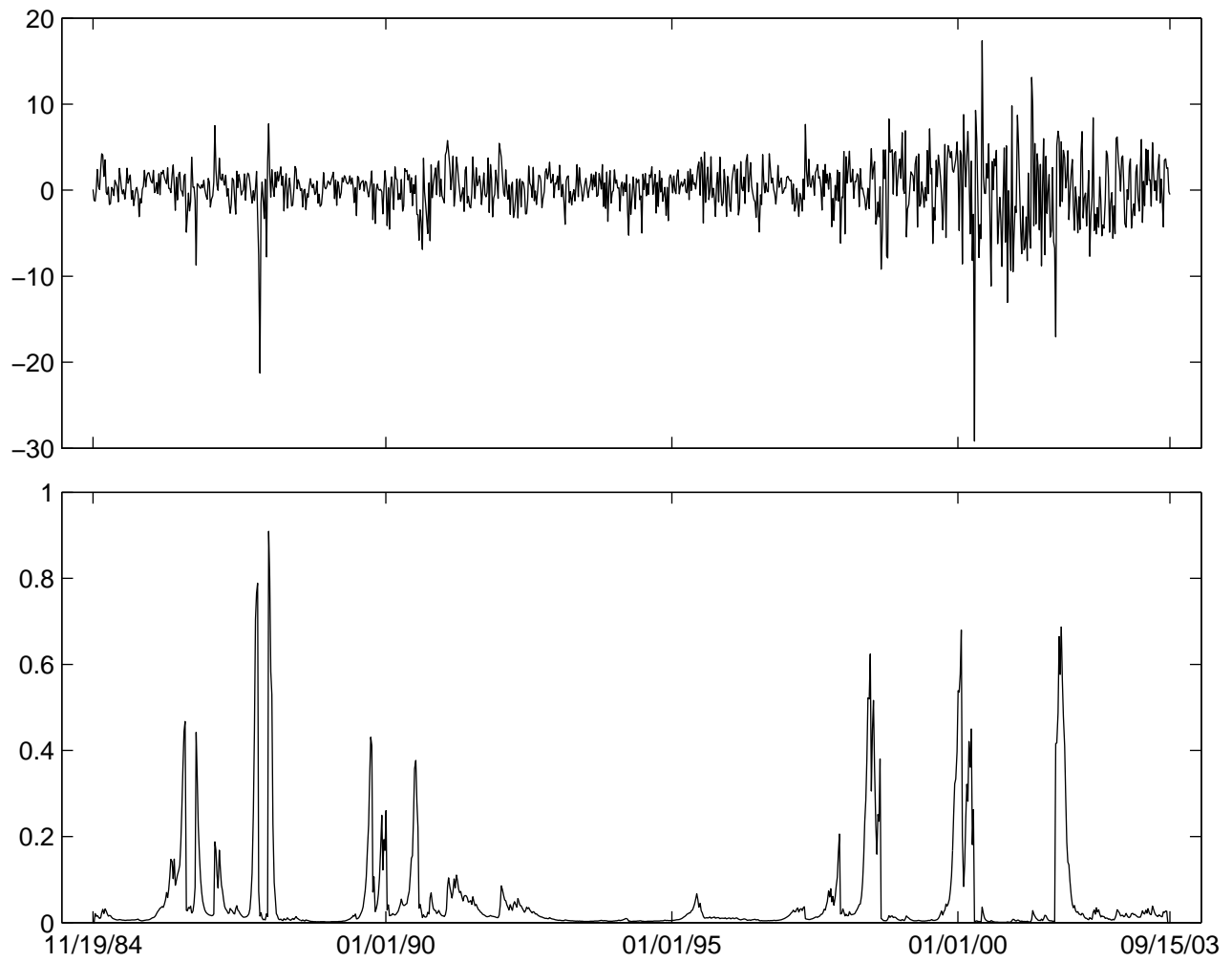


Fig. 1. Weekly Returns (top panel) of NASDAQ Index and Estimated Probabilities of Structural Change (bottom panel).

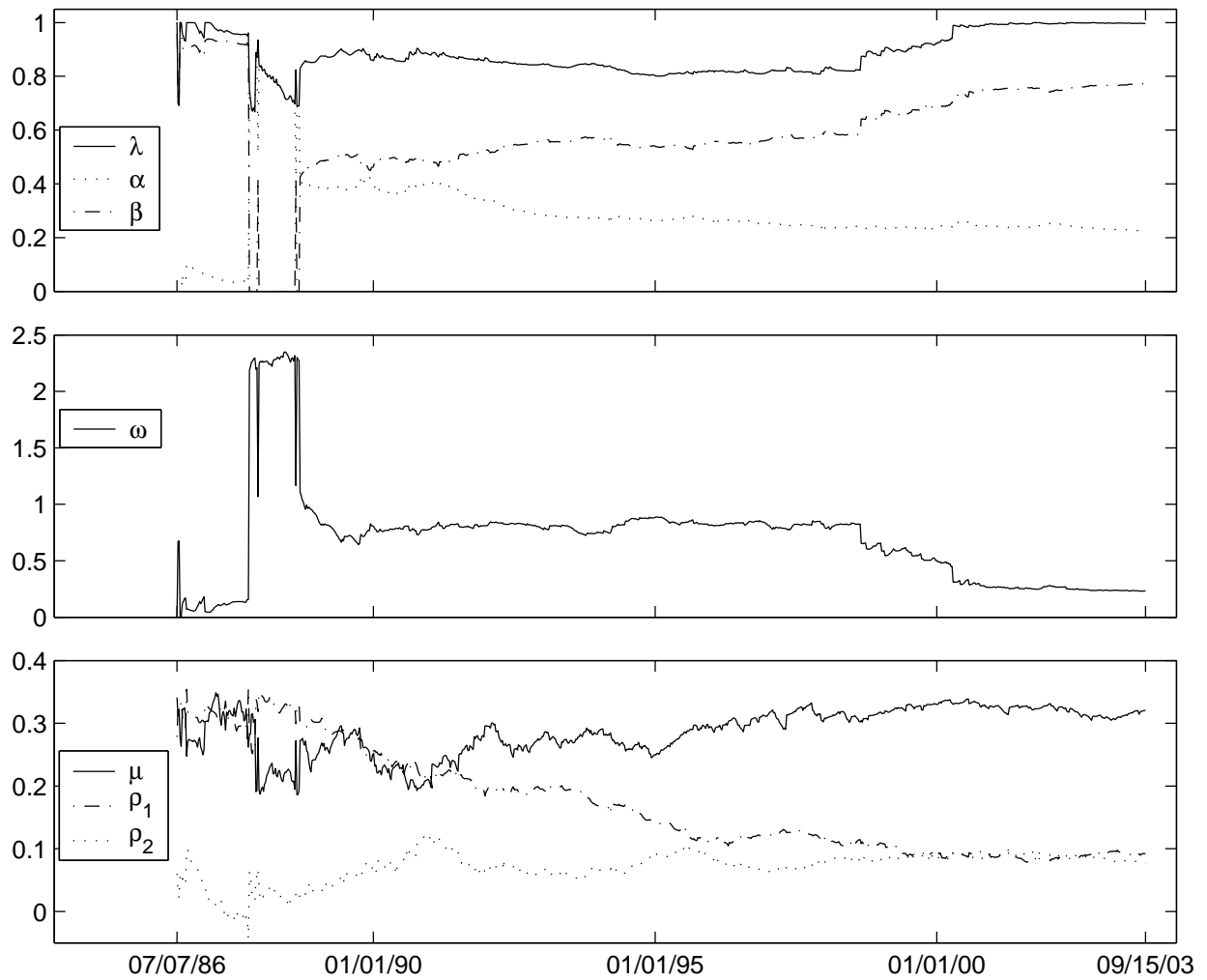


Fig. 2. Sequential Estimates of AR-GARCH Parameters.

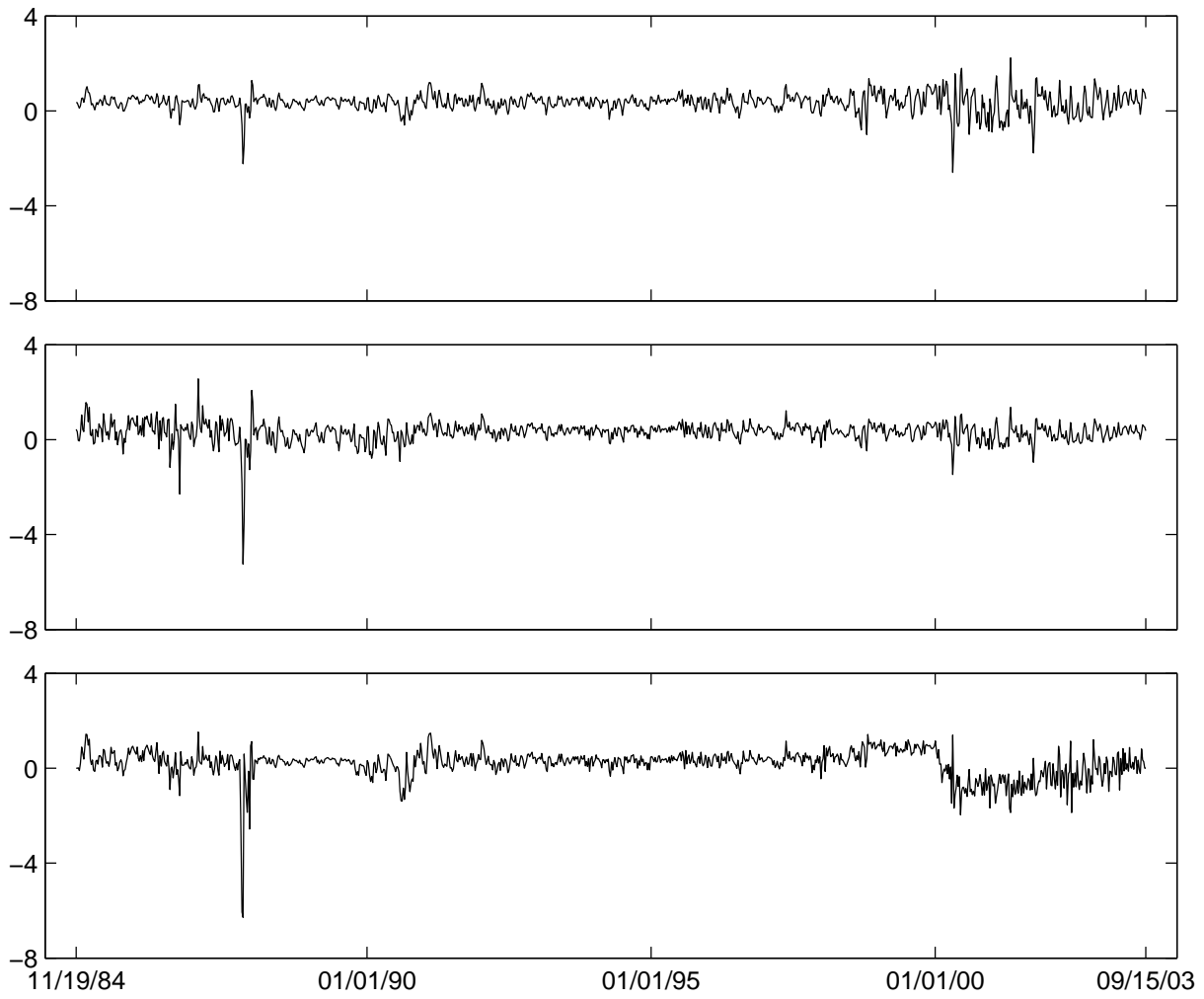


Fig. 3. Mean Return Estimates Using Three Different Methods.

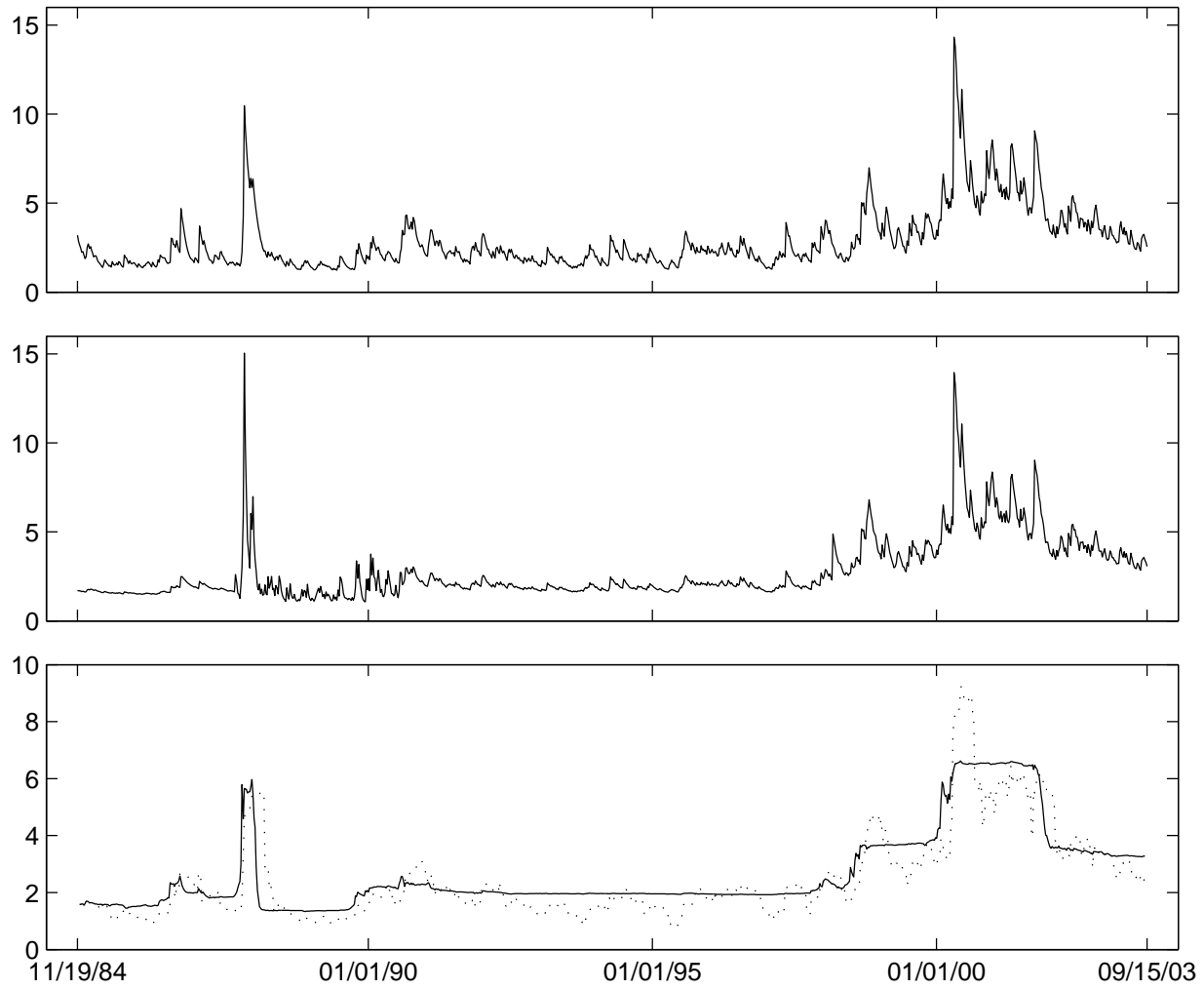


Fig. 4. Volatility Estimates Using Three Different Methods.

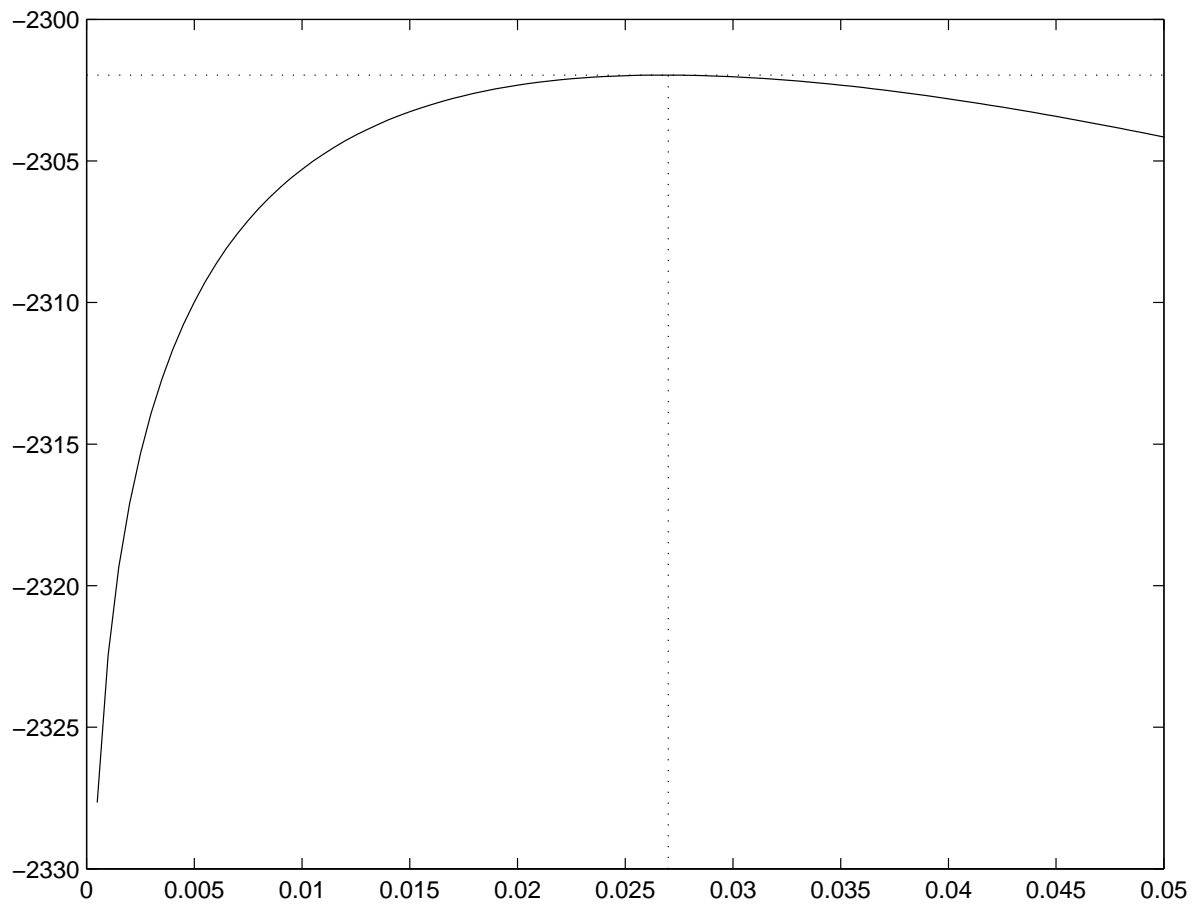


Fig. 5. Log-likelihood Function of Hyperparameter p in Bayesian Change-point Model.