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

by

Tze Leung Lai
Haipeng Xing

Technical Report No. 2004-21
September 2004

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
STRUCTURAL CHANGE AS AN ALTERNATIVE TO LONG MEMORY IN FINANCIAL TIME SERIES

by

Tze Leung Lai
Department of Statistics
Stanford University

Haipeng Xing
Department of Statistics
Stanford University

Technical Report No. 2004-21
September 2004

This research was supported in part by National Science Foundation grant DMS 0305749

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
STRUCTURAL CHANGE AS AN ALTERNATIVE TO LONG MEMORY IN FINANCIAL TIME SERIES

Tze Leung Lai and Haipeng Xing

ABSTRACT

This paper shows that volatility persistence in GARCH models and apparent 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.

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 $\omega, \alpha$ and $\beta$ 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 $\hat{\lambda} = \hat{\alpha} + \hat{\beta}$ close to 1. The constraint $\lambda < 1$ in the GARCH model (1), in which the $\epsilon_n$ are independent standard normal random variables that are not observable and the $y_n$ are the observations, enables the conditional variance $\sigma_n^2$ of $y_n$ given the past observations to be expressible as

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

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 $\omega$ in (1) not accommodating to shifts in monetary policy regimes might have led to an apparently integrated series of innovations. By assuming $\omega$ to be piecewise constant and allowing the possibility of jumps between evenly spaced time intervals, Lamoureux and Lastrapes (1990) obtained smaller estimates of $\lambda$, 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 $\omega$.

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 $\omega, \alpha, \beta$ 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 $\hat{\lambda}_N$ 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 apparent 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 behavior of the estimated autoregressive parameter. Some concluding remarks are given in Section V.

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 \( \epsilon_n \) 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 \texttt{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 \( \omega \) 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 \( \omega \) every 302 observations, thereby replacing \( \omega \) in (1) by \( \omega' + \delta_1 D_{1n} + \ldots + \delta_k D_{kn} \), where \( D_{1n}, \ldots, D_{kn} \) are dummy variables indicating the subsample to which \( n \) belongs. The
choice of 302, 604, 906, ... as the only possible jump points of $\omega$ while not allowing jumps in $\alpha$ and $\beta$ 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 $\mu$, $\hat{\rho}_1$ 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 $\hat{r}_n = \hat{\mu} + \hat{\rho}_1 r_{n-1} + \hat{\rho}_2 r_{n-2}$ 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 Eq. (20) in Section IIIB, in which $p$ is estimated by maximum likelihood as described in Section IIIC), 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 $\sigma_n$ undergoing periodic jumps according to a renewal process, unlike the continuously changing $\sigma_n$ in the GARCH model (1). As the
\( \sigma_n \) 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 \( \epsilon^*_n \) 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

Lai, Liu and Xing (2004) have recently introduced the following model to incorporate possible structural breaks in the AR(\( k \)) process

\[
X_n = \mu_n + \rho_{1n}X_{n-1} + \ldots + \rho_{kn}X_{n-k} + \sigma_n \epsilon_n, \quad n > k, \tag{2}
\]

where the \( \epsilon_n \) are i.i.d. unobservable standard normal random variables, and \( \theta_n = (\mu_n, \rho_{1n}, \ldots, \rho_{kn})^T \) and \( \sigma_n \) are piecewise constant parameters. Note that \( \sigma^2_n \) 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 \((\theta^T_t, \sigma_t)\) is assumed to form a discrete renewal process with parameter \( p \), or equivalently,

\[
I_t := 1_{\{\theta_t, \sigma_t \neq (\theta_{t-1}, \sigma_{t-1})\}} \text{ are i.i.d. Bernoulli random variables with } P(I_t = 1) = p \quad \tag{3}
\]

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

\[
(\theta^T_t, \tau_t) = (1 - I_t)(\theta^T_{t-1}, \tau_{t-1}) + I_t(Z^T_t, \gamma_t),
\]

where \((Z^T_t, \gamma_1), (Z^T_t, \gamma_2), \ldots\) are i.i.d. random vectors such that

\[
\gamma_t \sim \text{Gamma}(g, \lambda), \quad Z_t | \gamma_t \sim \text{Normal}(z, V/(2\gamma_t)). \quad \tag{4}
\]

#### A. Sequential Estimation of \( \sigma_n, \mu_n, \rho_{1,n}, \ldots, \rho_{k,n} : BCMIX Filters \)

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

\[
\tau_n \sim \text{Gamma}\left(g + \frac{n - J_n + 1}{2}, \frac{1}{a_{J_n,n}}\right), \quad \theta_n | \tau_n \sim \text{Normal}\left(z_{J_n,n}, \frac{1}{2\tau_n} V_{J_n,n}\right), \quad \tag{5}
\]
where for \( k < j \leq n \),
\[
V_{j,n} = \left( V^{-1} + \sum_{t=1}^{n} X_{t-k,t-1}X_{t-k,t-1}^T \right)^{-1}, \quad z_{j,n} = V_{j,n} \left( V^{-1} z + \sum_{t=1}^{n} X_{t-k,t-1}X_t \right),
\]
\[
a_{j,n} = \lambda^{-1} + z^TV^{-1}z + \sum_{t=j}^{n} X_t^2 - z_{j,n}^T V_{j,n}^{-1} z_{j,n}.
\]

Note that if \((2Y)^{-1}\) has a Gamma\((\tilde{g}, \tilde{\lambda})\) distribution, then \(Y\) has the inverse gamma 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
\[
E(\theta_n^T, \sigma_n^2 | X_{1,n}) = \sum_{j=k+1}^{n} p_{j,n} E(\theta_n^T, \sigma_n^2 | X_{1,n}, J_n = j) = \sum_{j=k+1}^{n} p_{j,n} \left( z_{j,n}^T \frac{a_{j,n}}{2g + n - j - 1} \right),
\]
where \(p_{j,n} = P(J_n = j | X_{1,n})\). Moreover, \(E(\sigma_n | X_{1,n}) = \Sigma_{j=k+1}^{n} p_{j,n} (a_{j,n}/2)^{1/2} \Gamma_{n-j}\), where \(\Gamma_i = \Gamma(g + i/2)/\Gamma(g + (i+1)/2)\).

The next step is to derive a recursive formula for \(p_{j,n}\). Denoting conditional densities by \(f(\cdot | \cdot)\), note that with the arrival of the new observation \(X_n\) at time \(n\), we have the updating formula
\[
f(\theta_n^T, \tau_n | X_{1,n}) = p_{n,n} f(\theta_n^T, \tau_n | X_{1,n}, J_n = n) + \sum_{j=k+1}^{n-1} p_{j,n} f(\theta_n^T, \tau_n | X_{1,n}, J_n = j),
\]
where
\[
p_{j,n} \propto p_{j,n}^* := \begin{cases} p f(X_n | J_n = j) & \text{if } j = n, \\ (1-p) p_{j,n-1} f(X_n | X_{j-k,n-1}, J_n = j) & \text{if } j \leq n-1. \end{cases}
\]

Since \(\Sigma_{i=k+1}^{n} p_{i,n} = 1\), \(p_{j,n}\) is given explicitly by \(p_{j,n} = p_{j,n}^* / \Sigma_{i=k+1}^{n} p_{i,n}^*\). Moreover, given \(J_n = j\) and \(X_{j-k,n-1}\), the conditional distribution of \((\theta_n^T, \tau_n)\) is
\[
\tau_n \sim \text{Gamma}(g + (n - j)/2, 1/a_{j,n-1}), \quad \theta_n | \tau_n \sim \text{Normal}(z_{j,n-1}, V_{j,n-1}/(2\tau_n)) \quad \text{if } j < n;
\]
\[
\tau_n \sim \text{Gamma}(g, \lambda), \quad \theta_n | \tau_n \sim \text{Normal}(z, V/(2\tau_n)) \quad \text{if } j = n;
\]
see Lemma 1 of Lai, Liu and Xing (2004) who also use this to show that
\[
f(\lambda_j^{-1}(X_n - z_{j,n-1}X_{j-k,n-1}) | J_n = j, X_{j-k,n-1}) = \text{Stud}(g_j)
\]
where \(\text{Stud}(g)\) denotes the Student-\(t\) density function with \(g\) degrees of freedom and
\[
g_j = 2g + n - j, \quad \lambda_j = a_{j,n} (1 + X_{n-k+1,n-1}^TVX_{n-k+1,n-1})/(2g + n - j) \quad \text{if } j < n;
\]
\[
g_j = 2g, \quad \lambda_j = (1 + X_{n-k+1,n-1}^TVX_{n-k+1,n-1})/(2\lambda g) \quad \text{if } j = n.
\]
Although the optimal 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 and computational and memory requirements in estimating \( \sigma_n \), and \( \theta_n \) as \( n \) keeps increasing. Lai, Liu and Xing (2004) 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 \( 0 \leq m_p < n_p \). Specifically, let \( \mathcal{K}_{n-1} \) denote the set of indices \( j \) such that the weight \( p_{j,n-1} \) is kept at stage \( n - 1 \); thus \( \mathcal{K}_{n-1} \supset \{n-1, \ldots, n-m_p\} \). At stage \( n \), define \( p_{j,n}^* \) by (8) for \( j \in \{n\} \cup \mathcal{K}_{n-1} \) and let \( i_n \) be the index not belonging to \( \{n, n-1, \ldots, n-m_p + 1\} \) such that

\[
p_{i_n,n}^* = \min \{ p_{j,n}^* : j \in \mathcal{K}_{n-1} \text{ and } j \leq n - m_p \},
\]

choosing \( i_n \) to be the one farthest from \( n \) if the minimizing set in (10) has more than one element. Define \( \mathcal{K}_n = \{n\} \cup (\mathcal{K}_{n-1} - \{i_n\}) \) and let \( p_{j,n} = p_{j,n}^*/\Sigma_{i \in \mathcal{K}_n} p_{j,n}^* \) for \( j \in \mathcal{K}_n \). For computational efficiency, Lai, Liu and Xing (2004) propose to use fast algorithms to evaluate the Student-\( t \) density (with possibly non-integer degrees of freedom) by taking logarithms and applying saddlepoint approximations when the number of degrees of freedom is large. Moreover, in view of (6), they propose to update \( V_{j,n} (j \leq n) \) recursively (in \( n \)) by

\[
V_{j,n} = V_{j,n-1} - V_{j,n-1} X_{n-k,n-1} X_{n-k,n-1}^T V_{j,n-1}/(1 + X_{n-k,n-1}^T V_{j,n-1} X_{n-k,n-1}), \quad \text{if } j < n,
\]

\[
V_{n,n} = V \cdot V X_{n-k,n-1} X_{n-k,n-1}^T V/(1 + X_{n-k,n-1}^T V X_{n-k,n-1}),
\]

\[
z_{j,n} = z_{j,n-1} + V_{j,n-1} X_{n-k,n-1} (X_n - X_{n-k,n-1} z_{j,n-1})/(1 + X_{n-k,n-1}^T V_{j,n-1} X_{n-k,n-1}),
\]

\[
a_{j,n} = a_{j,n-1} + (X_n - X_{n-k,n-1} z_{j,n-1})^2/(1 + X_{n-k,n-1}^T V_{j,n-1} X_{n-k,n-1}).
\]

(11)

In practice, the frequency of change-points in an observed time series is unknown and one has to estimate the hyperparameter \( p \) and possibly also other hyperparameters of the change-point autoregressive model from the data. Lai, Liu and Xing (2004) propose to estimate the hyperparameters by using Rissanen’s (1986) accumulated prediction error criterion in conjunction with the BCMIX algorithm that involves an unspecified hyperparameter \( \nu \) consisting of \( p, g, \lambda, z \) and \( V \). Specifically, the accumulated prediction error at time \( n \) is defined by

\[
\text{APE}_n(\nu) = \sum_{t=k+1}^n \{X_t - \hat{X}_{t|t-1}(\nu)\}^2,
\]

where \( \hat{X}_{t|t-1}(\nu) \) is the one-step-ahead predictor of \( X_t \) given by

\[
\hat{X}_{t|t-1}(\nu) = \{(1 - p) \hat{\theta}_{t-1}(\nu) + pz\}^T X_{t-k,t-1},
\]

(13)
in which $\hat{\theta}_t(\nu)$ is the BCMIX estimate (assuming the hyperparameter $\nu$) of $\theta$ based on $X_1, \ldots, X_t$. Let $H$ be a finite set of possible values of $\nu$ and let $\hat{\nu}_n$ be the minimizer of $\text{APE}_n(\nu)$ over $\nu \in H$. Lai, Liu and Xing (2004) propose to estimate $\theta_t$ and $\sigma_t^2$ by $\hat{\theta}_t(\hat{\nu}_{t-1})$ and $\hat{\sigma}_t^2(\hat{\nu}_{t-1})$, which they call the BCMIX-APE filter.

When $H$ has $h$ elements, we can compute the BCMIX-APE filter by $h$ parallel recursions. Specifically, for each $\nu \in H$, $\hat{\theta}_{t-1}(\nu)$ and $\hat{\sigma}_{t-1}^2(\nu)$ can be updated recursively in view of (8) and (11), and therefore we can easily update $\text{APE}_t(\nu)$ by $\text{APE}_t(\nu) = \text{APE}_{t-1}(\nu) + (X_t - \hat{X}_t|h-1(\nu))^2$. To choose a finite $H$, one can start by using prior information to come up with a range $(0 < \rho' \leq p \leq \rho'' < 1)$. Typically one has an upper bound $N$ for the length of past and future observations to be considered and can take $p' \geq 1/(2N)$, noting that the expected number of change-points for the series is $Np$. Lai, Liu and Xing (2004) replace the interval $[\rho', \rho'']$ by a finite set of points $2^{l}p'(0 \leq l \leq L)$ with $L = \max\{l : 2^{l}p' \leq p''\}$. Whereas the relative frequency $p$ of change-points is an essential feature of our structural change model, the other hyperparameters $g, \lambda, z$ and $V$ become significant only around change-points, where one does not have enough post-change observations and relies on prior information to estimate $\theta_t$ and $\sigma_t$. Accordingly, including a few plausible values of $(g, \lambda, z, V)$ should suffice. In particular, estimates of baseline values of $\theta$ and $\sigma^2$ (assumed not to be time-varying) from some historical data prior to $t = 0$ can provide estimated values of $(g, \lambda, z, V)$ via the method of moments. Thus the set $H$ chosen in this way typically has manageable cardinality for the implementation of parallel recursions.

B. Estimating Piecewise Constant Parameters from a Time Series: BCMIX Smoothers

Lai, Liu and Xing (2004) propose to evaluate the minimum variance estimate $E(\tau_t, \theta^T_t | X_1, \ldots, X_n)$, with $k + 1 \leq t \leq n$, by applying Bayes’ theorem to combine the forward filter that involves the conditional distribution of $(\tau_t, \theta^T_t)$ given $X_1, \ldots, X_t$ and the backward filter that involves the conditional distribution of $(\tau_t, \theta^T_t)$ given $X_{t+1}, \ldots, X_n$. Noting that the normal distribution for $\theta_t$ assigns positive probability to the explosive region \{ $\theta = (\mu, \rho_1, \ldots, \rho_k)^T : 1 - \rho_1 z - \cdots - \rho_k z^k \text{ has roots inside the unit circle}$ \}, they propose to replace the normal distribution in (4) by a truncated normal distribution that has support in some stability region $C$ such that

$$\inf_{|z| \leq 1} |1 - \rho_1 z - \cdots - \rho_k z^k| > 0 \quad \text{if } \theta = (\mu, \rho_1, \ldots, \rho_k)^T \in C. \tag{14}$$

Letting $T_{C}\text{Normal}(z, V)$ denote the conditional distribution of $Z$ given $Z \in C$, where $Z \sim$
Normal\((z, V)\) and \(C\) satisfies the stability condition \((14)\), they modify \((4)\) as

\[
\gamma_t \sim \text{Gamma}(g, \lambda), \quad Z_t|\gamma_t \sim TC\text{Normal}(z, V/(2\gamma_t)), \quad (15)
\]

and show that \((\tau_t, \theta^T_t, X_{t-k+1,t})\) has a stationary distribution under which \((\theta^T_t, \tau_t)\) has the same distribution as \((Z^T_t, \gamma_t)\) in \((15)\) and

\[
X_t|\theta^T_t, \tau_t \sim \text{Normal} \left( \mu_t/(1 - \rho_{1,t} - \cdots - \rho_{k,t}), (2\tau_t)^{-1} \nu_t \right),
\]

where \(\nu_t = \sum_{j=1}^{\infty} \beta_{j,t}^2\) and \(\beta_{j,t}\) are the coefficients in the power series representation of \(1/(1 - \alpha_{1,t}z - \cdots - \alpha_{k,t}z^k) = \sum_{j=0}^{\infty} \beta_{j,t}z^j\) for \(|z| \leq 1\). In addition, they show that the Markov chain \((\tau_t, \theta^T_t, X_{t-k+1,t})\) is reversible if it is initialized at the stationary distribution, and therefore the backward filter of \((\tau_t, \theta^T_t)\) based on \(X_n, \cdots, X_{t+1}\) 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

\[
f(\tau_t, \theta_t|X_{1,n}) \propto f(\tau_t, \theta_t|X_{1,t})f(\tau_t, \theta_t|X_{t+1,n})/\pi(\tau_t, \theta_t) \quad (16)
\]

where \(\pi\) denotes the stationary density function which is the same as that of \((\gamma_t, Z_t)\) given in \((15)\).

Because the truncated normal is used in lieu of the normal distribution in \((13)\), the conditional distribution of \(\theta_n\) given \(\tau_n\) needs to be replaced by \(\theta_n|\tau_n \sim TC \text{Normal}(z_{J,n}, V_{J,n}, (2\tau_n))\), while \((6)\) defining \(V_{j,n}, z_{j,n}\) and \(\alpha_{j,n}\) remains unchanged. For the implementation of the BCMIX filter, Lai, Liu and Xing (2004) propose to ignore the truncation and to apply the closed-form expressions in BCMIX, 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 \((\theta_t, \tau_t)\) given \(J_t = i, \bar{J}_t = j + 1\) and \(X_{i,j}\) for \(i \leq t < j \leq n\) can be described by

\[
\tau_t \sim \text{Gamma} \left( g + \frac{j - i + 1}{2}, \frac{1}{a_{i,j,t}} \right), \quad \theta_t|\tau_t \sim \text{Normal} \left( z_{i,j,t}, \frac{1}{2\tau_t} V_{i,j,t} \right) \quad (17)
\]

if we ignore the truncation in the truncated normal distribution in \((15)\), where

\[
V_{i,j,t} = (V_{i,t}^{-1} + \tilde{V}_{j,t}^{-1} - V^{-1})^{-1},
\]

\[
z_{i,j,t} = V_{i,j,t}(V_{i,j,t}^{-1} z_{i,t} + \tilde{V}_{j,t}^{-1} \tilde{z}_{j,t+1} - V^{-1} z),
\]

\[
a_{i,j,t} = \lambda^{-1} + z^T V^{-1} z + \sum_{l=i}^{j} X_l^2 - z_{i,j,t}^T V_{i,j,t}^{-1} z_{i,j,t},
\]
in which $V_{i,t}$, $z_{i,t}$ and $a_{i,t}$ are defined in (6) and $\tilde{V}_{j,t}$, $\tilde{z}_{j,t}$ and $\tilde{a}_{j,t}$ are defined similarly by reversing time. Let $|\cdot|$ denote the determinant of a matrix,

$$
    b_{i,j,t} = \left( \frac{|V_{i,t}||\tilde{V}_{j,t}|}{|V||\tilde{V}_{i,j,t}|} \right)^{-1/2} \left\{ \frac{\Gamma(g)\Gamma(g + \frac{1}{2}(j-i+1))}{\Gamma(g + \frac{1}{2}(t-i+1))\Gamma(g + \frac{1}{2}(j-t))} \right\} \frac{a_{i,t}^{g+(t-i+1)/2}a_{j,t}^{g+(j-t)/2}}{a_{i,j,t}^{g+(j-t-i+1)/2}},
$$

$$
    B_t = p + (1-p) \sum_{k+1 \leq i \leq j \leq n} p_{i,t}\tilde{p}_{j,t}b_{i,j,t}.
$$

Using (16) and (17), Lai, Liu and Xing (2004) have shown that analogous to (7),

$$
    E(\sigma^2_t|X_{1,n}) = \frac{p}{B_t} \sum_{i=k+1}^{t} p_{i,t}a_{i,t} + \frac{1-p}{B_t} \sum_{k+1 \leq i \leq j \leq n} p_{i,t}\tilde{p}_{j,t}b_{i,j,t}2g+j-i+1,
$$

$$
    E(\theta_t|X_{1,n}) = \frac{p}{B_t} \sum_{i=k+1}^{t} p_{i,t}z_{i,t} + \frac{1-p}{B_t} \sum_{k+1 \leq i \leq j \leq n} p_{i,t}\tilde{p}_{j,t}b_{i,j,t}z_{i,j,t},
$$

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

$$
    P\{I_t = 1|X_{1,n}\} = \sum_{i=k+1}^{t} P\{I_t = 1, J_{t-1} = i|X_{1,n}\} = \frac{p}{B_t},
$$

recalling that $P\{I_t = 1, J_{t-1} = i|X_{1,n}\} = pp_{i,1}$ and $\sum_{i=k+1}^{t} p_{i,t} = 1$.

The BCMIX smoother further approximates (18) by allowing at most $n_t$ weights $p_{i,t}$ and $n_p$ weights $\tilde{p}_{j,t}$ to be nonzero. When there are unspecified hyperparameters, extending the APE criterion (12) to the smoothing problem yields the usual cross validation criterion

$$
    CV(\nu) = \sum_{t=r+1}^{n-k} \{X_t - \hat{X}_{t,n}(\nu)\}^2,
$$

where $\nu$ is the hyperparameter vector and $\hat{X}_{t,n}(\nu)$ is a predictor of $X_t$ based on $(X_{1,t-1}, X_{t+1,n})$ obtained by combining the forward and backward BCMIX estimates. Lai, Liu and Xing (2004) propose to estimate $\nu$ by $\tilde{\nu}_n$ that minimizes $V(\nu)$ over $\nu \in H$. An alternative method to estimate $\nu$ is to use maximum likelihood, as in the next section.

**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 (19), in which the hyperparameters $p, z, V, g$ and $\lambda$ of the change-point autoregressive model (3)-(4) are determined as follows.

First we fit an AR(2) model to $\{r_t, r_{t-1}, \cdots, r_{t-19}\}$, $\max(k + 1, 20) \leq t \leq n$, 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 $\sigma_t^2$. The dotted curve in the bottom panel of Figure 4 is a plot of the $\tilde{\sigma}_t$ series. Then we estimate $z$ and $V$ by the sample mean $\bar{z}$ and the sample covariance matrix $\hat{V}$ 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 $\lambda$ of (4). Specifically, regarding $(2\tilde{\sigma}_t^2)^{-1}$ as a sample from the Gamma($g, \lambda$) 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 $\lambda$:

$$\hat{g} = 2.4479, \quad \hat{\lambda} = 0.03960,$$

$$\bar{z} = (0.1304 \quad 0.003722 \quad -0.008111)^T,$$

$$\hat{V} = \begin{pmatrix}
0.06096 & -0.009412 & 0.0003295 \\
-0.009412 & 0.01864 & 0.004521 \\
0.0003295 & 0.004521 & 0.01023
\end{pmatrix}.$$

With the hyperparameters $z, V, g$ and $\lambda$ thus chosen, we can estimate the hyperparameter $p$ by maximum likelihood, noting that

$$f(X_t|X_{1,t-1}) = pf(X_t|J_t = t) + (1 - p) \sum_{j=k+1}^{t-1} P(J_{t-1} = j|X_{1,t-1}) f(X_t|X_{1,t-1}, J_{t-1} = j)$$

$$= \sum_{j=k+1}^{t} p_{j,t}^*,$$

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

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

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

In this section we consider a simpler setting than those in the preceding sections, and demonstrate that apparent long memory can arise when the possibility of structural change is not incorporated. In order to derive explicit formulas, we assume an AR(1) model that has occasional jumps only in its mean \( \mu_n \) but not in its autoregressive and volatility parameters. Specifically, instead of (2), we now consider

\[
X_n = \mu_n + \rho X_{n-1} + \epsilon_n, \tag{20}
\]

in which \( \epsilon_n \) 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

\[
\mu_n = (1 - I_n) \mu_{n-1} + I_n Z_n, \tag{21}
\]

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

Without incorporating possible shifts in \( \mu \), 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,

\[
\hat{\rho} = \frac{\sum_{i=2}^{T} X_i X_{i-1} - \frac{1}{T-1} \left( \sum_{i=2}^{T} X_i \right) \left( \sum_{i=1}^{T-1} X_i \right)}{\sum_{i=1}^{T-1} X_i^2 - \frac{1}{T-1} \left( \sum_{i=1}^{T-1} X_i \right)^2}. \tag{22}
\]

Since \( |\rho| < 1 \), the Markov chain \( (\mu_n, X_n) \) defined by (20)-(21) has a stationary distribution \( \pi \) under which \( \mu_n \) has the \( N(\mu, \nu) \) distribution and \( (1 - \rho)X_n \) has the \( N(\mu, 1 + \nu) \) distribution. Hence

\[
E_\pi(X_n) = \mu/(1 - \rho), \quad E_\pi(X_n^2) = (1 + \nu + \mu^2)/(1 - \rho)^2, \quad E_\pi(\mu_n) = \mu. \tag{23}
\]

Making use of (23) and the stationary distribution of \( (\mu_n, X_n) \), 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 \( n \to \infty \) even though \( |\rho| < 1 \).

**Theorem:** Suppose that in (20)-(21), \( (1 - p)\rho = 1 - pv \). Then \( \hat{\rho} \to 1 \) 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 $\mu$ to $\{X_n, 1 \leq n \leq T\}$, in which $X_n$ is generated from (20) that allows periodic jumps in $\mu$. Unless stated otherwise, $T = 1000$. Each result in Table 2 is the average of 100 simulations from the change-point model (20), with the standard error included in parentheses. The table considers different values of the parameters $p$, $\rho$ and $\nu$ of the change-point model and also lists the corresponding values of the ratio $(1-p)\rho/(1-\nu\rho)$ 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 \texttt{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. CONCLUSION

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 and the asymptotic theory in Sections II and 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 $\theta_t$ and $\sigma_t$ undergoing Markovian jump dynamics, so that estimation of $\theta_t$ and $\sigma_t$ 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, Section III describes the BCMIX approximation, with relatively low computational complexity, to the optimal filters and smoothers. Another approach, introduced in Lai, Liu and Xing (2004), is based on Monte Carlo simulations and involves sequential importance sampling with resampling. In Section III C, 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.

ACKNOWLEDGMENT

12
This research was supported by the National Science Foundation.
REFERENCES


APPENDIX

The proof of the theorem makes use of (22), (23) and

\[ \text{Cov}_\pi(\mu_{n+1}, X_n) = (1 - p)v/\{1 - (1 - p)\rho\}. \]  \hspace{1cm} (24)

To prove (24), first note from (20) that

\[
\text{Cov}_\pi(\mu_n, X_n) = v + \rho \text{Cov}(\mu_n, X_{n-1})
\]
\[
= v + \rho \{E(\mu_n X_{n-1}) - E(\mu_n)E(X_{n-1})\}
\]
\[
= v + (1 - p)\rho \text{Cov}_\pi(\mu_{n-1}, X_{n-1}).
\]  \hspace{1cm} (25)

The last equality in (25) follows from (21) and that \(E_\pi(I_n Z_n) = p\mu\). The recursive representation of \(\text{Cov}_\pi(\mu_n, X_n)\) in (25) yields

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

Combining this with the first equality in (25) establishes (24). In view of (25) and the strong law for Markov random walks (cf. Theorem 17.1.7 of Meyn and Tweedie (1993)),

\[
\sum_{i=1}^{T-1} X_i/T \to E_\pi(X_n),
\]
\[
\left(\sum_{i=1}^{T-1} X_{i+1}X_i\right)/T = \left\{\sum_{i=1}^{T-1} \mu_{i+1}X_i + \rho \sum_{i=1}^{T-1} X_i^2 + \sum_{i=1}^{T-1} \epsilon_{i+1}X_i\right\}/T
\]
\[
\to E_\pi(\mu_{n+1}X_n) + \rho E_\pi(X_n^2) = \text{Cov}_\pi(\mu_{n+1}, X_n) + E_\pi(\mu_{n+1})E_\pi(X_n) + \rho E_\pi(X_n^2),
\]  \hspace{1cm} (26)

with probability 1, noting that \(E_\pi(X_n\epsilon_{n+1}) = 0\). Putting (26) and (23), (24) into (22) 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

<table>
<thead>
<tr>
<th>Period</th>
<th>$\mu$</th>
<th>$\rho_1$</th>
<th>$\rho_2$</th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov 19, 1984 – Jun 16, 1986</td>
<td>0.3328</td>
<td>0.3091</td>
<td>0.09838</td>
<td>1.7453</td>
<td>0.0364</td>
<td>0.0</td>
<td>0.0364</td>
</tr>
<tr>
<td>Jun 09, 1987 – Aug 15, 1990</td>
<td>0.1242</td>
<td>0.2068</td>
<td>0.0891</td>
<td>0.8364</td>
<td>0.4645</td>
<td>0.4784</td>
<td>0.9429</td>
</tr>
<tr>
<td>Aug 08, 1990 – Mar 13, 1998</td>
<td>0.3362</td>
<td>0.03874</td>
<td>0.1039</td>
<td>0.4560</td>
<td>0.08767</td>
<td>0.8008</td>
<td>0.8885</td>
</tr>
<tr>
<td>Mar 06, 1998 – Sep 15, 2003</td>
<td>0.2785</td>
<td>0.05612</td>
<td>0.04063</td>
<td>0.8777</td>
<td>0.2039</td>
<td>0.7819</td>
<td>0.9858</td>
</tr>
<tr>
<td>Nov 19, 1984 – Sep 15, 2003</td>
<td>0.3203</td>
<td>0.09241</td>
<td>0.07657</td>
<td>0.2335</td>
<td>0.2243</td>
<td>0.7720</td>
<td>0.9963</td>
</tr>
</tbody>
</table>
Table 2. Means and Standard Errors (in Parentheses) of Least Squares Estimates and P-Values of Unit Root Test

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

(T=2000, T=3000, 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.