AUTOREGRESSIVE MODELS WITH PIECEWISE CONSTANT VOLATILITY AND REGRESSION PARAMETERS

by

Tze Leung Lai
Haiyan Liu
Haipeng Xing

November 2003

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
AUTOREGRESSIVE MODELS WITH PIECEWISE CONSTANT VOLATILITY AND REGRESSIVE PARAMETERS

by

Tze Leung Lai
Statistics Department
Stanford University

Haiyan Liu
Statistics Department
Stanford University

Haipeng Xing
Statistics Department
Stanford University

November 2003

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
AUTOREGRESSIVE MODELS WITH PIECEWISE CONSTANT VOLATILITY AND REGRESSION PARAMETERS

Tze Leung Lai, Haiyan Liu and Haipeng Xing
Stanford University

Abstract: We introduce herein a new class of autoregressive models in which the regression parameters and error variances may undergo changes at unknown time points while staying constant between adjacent change-points. Assuming conjugate priors, we derive closed-form recursive Bayes estimates of the regression parameters and error variances. Approximations to the Bayes estimates are developed that have much lower computational complexity and yet are comparable to the Bayes estimates in statistical efficiency. We also address the problem of unknown hyperparameters and propose two practical methods for simultaneous estimation of the hyperparameters, regression parameters and error variances. Applications of the methodology to simulated and real data show that it provides a promising approach to modeling and forecasting econometric time series such as asset returns and their volatilities.

Key words and phrases: Bayesian inference, bounded complexity mixtures, change-point problems, filtering, sequential Monte Carlo, smoothing.
1. Introduction

The problem of modeling a time series whose parameters may undergo occasional changes arises in many engineering, econometric and biomedical applications, and has an extensive literature widely scattered in these fields besides statistics. In stochastic dynamical systems, if the parameters may change with time, it is more convenient to regard them as states. However, this requires specification or modeling of the dynamics of the parameters. A Bayesian approach to this problem is to take a stochastic process as the prior distribution for the time-varying parameters, whose posterior distribution then provides an estimate of the current parameter value given the current and past observations. This reduces the estimation problem to a filtering problem. Except for the special case in which the system and parameter dynamics can be represented by a linear Gaussian state-space model so that Kalman filtering can be applied, the optimal filter is typically nonlinear and infinite-dimensional. Even for the simple mean shift model $Y_t = \theta_t + \epsilon_t$, $t = 1, 2, \ldots$, in which (i) the $\epsilon_t$ are i.i.d. zero-mean normal random variables, (ii) the sequence of change-times of $\{\theta_t\}$ forms a discrete renewal process with geometric interarrival times with parameter $p$, and (iii) the post-change values of $\{\theta_t\}$ are i.i.d. normal, the unknown times of the occurrence of the mean shifts leads to great complexity of the Bayes estimate $\hat{\theta}_n = E(\theta_n|Y_1, \ldots, Y_n)$. Chernoff and Zacks (1964) gave a closed-form expression of $\hat{\theta}_n$ that requires $O(2^n)$ operations to compute. Yao (1984) later found another representation of the estimate that requires only $O(n^2)$ operations. By combining forward and backward (i.e., time-reversed) filters to solve the smoothing problem of estimating $\theta_t$ from $Y_1, \ldots, Y_n$ for $1 \leq t \leq n$, he also gave a formula for the Bayes estimate $E(\theta_t|Y_1, \ldots, Y_n)$ that requires $O(n^3)$ operations to compute.

A natural extension of the mean shift model is the regression model $Y_t = \theta_t^T X_t + \epsilon_t$, in which the regressors $X_t$ are random vectors that may depend on the past observations $(X_i, Y_i)$, $i \leq t - 1$. In this paper we consider the special case $X_t = (1, Y_{t-1}, \ldots, Y_{t-k})^T$ that corresponds to the AR($k$) model in the time series literature. Motivated by applications to financial econometrics, in which not only the levels but also the volatilities of asset returns are of basic interest, we consider a further extension to the case where changes in $\sigma_t^2 := \text{Var}(\epsilon_t)$ can also occur besides changes in $\theta_t$. These extensions, given in Section 2 for the filtering problem and in Section 3 for the smoothing problem, make use of certain formulas for Bayesian inference in normal populations, a comprehensive introduction to which can be found in Box and Tiao’s (1973) classic.

Even with a Yao-type algorithm, the complexity of the Bayes estimates of $\sigma_t$ and $\theta_t$ becomes unmanageable when $n$ is large. In Section 4 we develop two recursive approximations
that can be updated with a fixed number (not depending on \( n \)) of operations. One approximation, called BCMIX (bounded complexity mixture), uses only a fixed number of filters in the extensions of Yao’s algorithm in Sections 2 and 3. The other approximation, called SISR (sequential importance sampling with resampling), is a Monte Carlo approximation that uses a fixed and relatively small number of trajectories that are recursively simulated by importance sampling. Numerical results showing the efficiency of these approximations are provided.

Sections 2-4 assume \( p \), the hyperparameter of the Bayesian model, to be correctly specified. When \( p \) is unknown, one needs to estimate it from the available data. Section 5 describes some computationally convenient (when used in conjunction with BCMIX or SISR) estimators that have good statistical performance. Moreover, in practice the assumed change-point autoregressive model is only an approximation to the actual data generating mechanism of the observed time series, and the purpose of fitting the model is to derive forecasts of future values that are yet to be observed. In Section 5 we apply the BCMIX method (with estimated \( p \)) to sequentially update the \( \theta_t \) and \( \sigma_t \) for weekly returns of the NASDAQ stock index from November 1984 to September 2003, and thereby construct prediction quantiles for the following week’s return during this period as in VaR (value at risk) applications.

The results in Section 5 show that change-point autoregressive models indeed provide a rich and yet tractable class of models that can capture many key features of econometric time series. By using normal mixtures, they can adapt to various distributional forms of \( Y_t \). The AR\((k)\) model provides simple forecasts of future observations, and allowing the regression coefficients to change over time yields flexible non-linear predictors. Allowing also \( \sigma_t \) to change over time can account for relatively calm periods punctuated by highly volatile periods observed in stock price movements and other econometric time series. Markets respond to policy changes, announcements of companies’ earnings and a myriad of perceived or real changes of the economy, so it is useful to incorporate uncertain (random) change-points in modeling econometric time series. Box and Tiao’s (1975) seminal work on intervention analysis represents one of the major directions in this kind of modeling. Section 6 compares intervention analysis with change-point autoregression and gives some concluding remarks.

2. A Bayesian Change-point Model and Filters Estimating \( \theta_t \) and \( \sigma_t \)

An autoregressive model with piecewise constant volatility and regression parameters has the form

\[
Y_t = \mu_t + \alpha_{1,t}Y_{t-1} + \cdots + \alpha_{k,t}Y_{t-k} + \sigma_t\varepsilon_t, \quad t > k, \tag{2.1}
\]
where the $\epsilon_t$ are i.i.d. unobservable random disturbances with mean 0 and variance 1, and $\theta_t = (\mu_t, \alpha_{t1}, \cdots, \alpha_{tk})^T$ and $\sigma_t$ are piecewise constant parameters. A Bayesian modeling approach requires also specification of the distributions of $\epsilon_t$ and $\{((\theta^T_t, \sigma_t), t > k\}$. Following Yao (1984), we assume that the sequence of change-times of $(\theta^T_t, \sigma_t)$ forms a discrete renewal process with parameter $p$, or equivalently,

$$I_t := 1_{\{((\theta_t^T, \sigma_t) \neq (\theta_{t-1}^T, \sigma_{t-1}))\}}$$

are i.i.d. Bernoulli random variables with $P(I_t = 1) = p$ \hspace{1cm} (2.2)

for $t \geq k+2$ and $I_{k+1} = 1$. Another assumption underlying Yao’s closed-form expressions for Bayes estimates in the mean shift model is normal $\epsilon_t$ and $\theta_t$, resulting in normal mixtures for the posterior distributions of the conjugate prior. We generalize this idea by assuming at change-times an inverse gamma prior distribution for $\sigma_t^2$ and a normal prior distribution for $\theta_t$ given $\sigma_t$. Specifically, letting $\tau_t = (2\sigma_t^2)^{-1}$, we assume that

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

where $(Z^T_1, \gamma_1), (Z^T_2, \gamma_2), \cdots$ are i.i.d. random vectors such that

$$\gamma_t \sim \text{Gamma}(g, \lambda), \hspace{1cm} Z_t|\gamma_t \sim \text{Normal}(z, V/(2\gamma_t)). \hspace{1cm} (2.3)$$

When there is no change-point (i.e. $p = 0$), the posterior distribution of $\tau_t$ is still gamma while that of $\theta_t$ given $\tau_t$ is multivariate normal, and there are simple formulas for updating the shape and scale parameters of the gamma distribution and the normal mean and covariance matrices; see Section 2.7 of Box and Tiao (1973). In the presence of change-points, the posterior distribution of $(\theta^T_t, \tau_t)$ is a mixture of gamma-normal distributions and we now extend Yao’s (1984) algorithm to evaluate the parameters of the posterior distribution.

As in Yao’s algorithm, the most recent change-time $J_n := \max\{t \leq n : I_t = 1\}$ plays a basic role in computing the Bayes estimate $E\{(\theta^T_n, \sigma_n^2)|Y_1, \cdots, Y_n\}$. Let $Y_{t,n} = (1, Y_1, \ldots, Y_n)^T$. Recalling that $\tau_n = (2\sigma_n^2)^{-1}$, the conditional distribution of $(\theta^T_n, \tau_n)$ given $(J_n, Y_{J_n,n})$ can be described by

$$\tau_n \sim \text{Gamma}\left(g + \frac{n - J_n + 1}{2}, \frac{1}{a_{J_n,n}}\right), \hspace{1cm} \theta_n|\tau_n \sim \text{Normal}\left(z_{J_n,n}, \frac{1}{2\tau_n}V_{J_n,n}\right), \hspace{1cm} (2.4)$$

where for $k < j \leq n$,

$$V_{j,n} = \left(V^{-1} + \sum_{t=j}^{n} Y_{t-k,t-1}^T Y_{t-k,t-1}\right)^{-1}, \hspace{1cm} z_{j,n} = V_{j,n}(V^{-1}z + \sum_{t=j}^{n} Y_{t-k,t-1}Y_t),$$

$$a_{j,n} = \lambda^{-1} + z^TV^{-1}z + \sum_{t=j}^{n} Y_{t-k,t-1}Y_{j,n}^T V_{j,n}z_{j,n}. \hspace{1cm} (2.5)$$
Note that if \((2X)^{-1}\) has a Gamma\((\tilde{\gamma}, \tilde{\lambda})\) distribution, then \(X\) has the inverse gamma IG\((g, \lambda)\) distribution with \(g = \tilde{g}, \lambda = 2\tilde{\lambda}\), and that \(EX = \lambda^{-1}(g - 1)^{-1}\) when \(g > 1\) and \(E\sqrt{X} = \lambda^{-1/2}\Gamma(g - \frac{1}{2})/\Gamma(g)\). It then follows from (2.4) that

\[
E(\theta^n_T, \sigma^2_n|Y_{1,n}) = \sum_{j=k+1}^n p_{j,n}E(\theta^n_T, \sigma^2_n|Y_{1,n}, J_n = j) = \sum_{j=k+1}^n p_{j,n}\left(z_j^n, \frac{a_j^n}{2g + n - j - 1}\right),
\]

where \(p_{j,n} = P(J_n = j|Y_{1,n})\). Moreover, \(E(\sigma_n|Y_{1,n}) = \Sigma_{j=k+1}^n p_{j,n}(a_j^n/2)^{1/2}\Gamma_{n-j}\), where \(\Gamma_i = \Gamma(g + i/2)/\Gamma(g + (i + 1)/2)\).

The next step is to derive a recursive formula for \(p_{j,n}\), as in Yao (1984) for the mean shift problem. Denoting conditional densities by \(f(\cdot|\cdot)\), note that with the arrival of the new observation \(Y_n\) at time \(n\), we can update the posterior density of \((\theta^n_T, \tau_n)\) by

\[
f(\theta^n_T, \tau_n|Y_{1,n}) = p_{n,n}f(\theta^n_T, \tau_n|Y_{1,n}, J_n = n) + \sum_{j=k+1}^{n-1} p_{j,n}f(\theta^n_T, \tau_n|Y_{1,n}, J_n = j),
\]

where

\[
p_{j,n} = p_{j,n}^* := \begin{cases} 
p_j(Y_n|J_n = 1) & \text{if } j = n, \\
(1 - p)\prod_{i=k+1}^n f(Y_{i}|Y_{i-1}, J_{i-1} = j) & \text{if } j \leq n - 1.
\end{cases}
\]

Since \(\sum_{i=k+1}^n p_{i,n} = 1\), \(p_{j,n}\) is given explicitly by \(p_{j,n} = p_{j,n}^*/\sum_{i=k+1}^n p_{i,n}^*\). Moreover, noting that \(Y_n = \theta^n_T Y_{n-k,n-1} + \sigma_n e_n\), we can apply the following lemma to obtain an explicit formula for the conditional density function of \(Y_n\) given \(J_n = j\) and \(Y_{j,n-1}\).

**Lemma 1.** (i) Suppose that the conditional distribution of \(Y\) given \(\theta\) and \(\tau\) is Normal\((\theta^T \phi, (2\tau)^{-1})\), \(\tau \sim \text{Gamma}(\tilde{\gamma}, \tilde{\lambda})\) and that the conditional distribution of \(\theta\) given \(\tau\) is Normal\((\phi, V/(2\tau))\). Then

\[
(Y_i - z^T \phi)/\{(1 + \phi^T V \phi)/(2\tilde{g}\tilde{\lambda})\}^{1/2}
\]

has a Student-t distribution with 2\(\tilde{g}\) degrees of freedom.

(ii) Given \(J_n = j\) and \(Y_{j,n-1}\), the conditional distribution of \((\theta^n_T, \gamma_n)\) is

- \(\tau_n \sim \text{Gamma}(g + (n - j)/2, 1/a_{j,n-1})\), \(\theta^n_n|\tau_n \sim \text{Normal}(z_{j,n-1}, V_{j,n-1}/(2\tau_n))\) if \(j < n\);
- \(\tau_n \sim \text{Gamma}(g, \lambda)\), \(\theta^n_n|\tau_n \sim \text{Normal}(z, V/(2\tau_n))\) if \(j = n\).

**Proof.** (i) follows from \(f(y) = \int \int f(y|\theta, \tau)f(\theta|\tau)f(\tau)d\theta d\tau\) and using a change of variables to perform the integration. To prove(ii), apply (2.4) with \(n\) replaced by \(n - 1\) in the case \(J_n < n\) (so that \(J_n = J_{n-1}\)). When \(J_n = n\), \((\theta^n_T, \tau_n)\) has a jump at time \(n\) and its distribution follows (2.3).
3. Bayesian Smoothers for $\theta_t$ and $\sigma_t$

For the simple mean shift model described in Section 1 (which corresponds to $k = 0$ and $\sigma_t = \sigma$ known), Yao's (1984) algorithm for computing the Bayes estimate $E(\theta_t|Y_1, \cdots, Y_n)$ with $1 \leq t \leq n$ is based on combining the forward filter involving the posterior distribution of $\theta_t$ given $Y_1, \cdots, Y_t$ and the backward filter involving the conditional distribution of $\theta_t$ given $Y_{t+1}, \cdots, Y_n$. In view of the reversibility property that $(Y_1, \cdots, Y_n)$ has the same distribution as $(Y_n, \cdots, Y_1)$, the backward filter has the same structure as the forward predictor of $\theta_t$ based on the past $n-t$ observations. For the change-point autoregressive model (2.1), reversibility cannot hold because the normal distribution for $\theta_t$ gives positive probability to the explosive region $\{\theta = (\mu, \alpha_1, \ldots, \alpha_k)^T : 1 - \alpha_1 z - \cdots - \alpha_k z^k \text{ has roots inside the unit circle} \}$. On the other hand, if we replace the normal distribution in (2.3) by a truncated normal distribution that has support in some stability region $C$ such that

$$\inf_{|z| \leq 1} |1 - \alpha_1 z - \cdots - \alpha_k z^k| > 0 \quad \text{if} \quad \theta = (\mu, \alpha_1, \ldots, \alpha_k)^T \in C; \quad (3.1)$$

then the following theorem shows that the Markov chain $(\tau_t, \theta_t^T, Y_{t-k+1,t})$ has a stationary distribution and is reversible if it is initialized at the stationary distribution. We shall use the prefix $T_C$ to denote truncation of a distribution within the region $C$. In particular, $T_C \text{Normal}(z, V)$ denotes the conditional distribution of $Z$ given $Z \in C$, where $Z \sim \text{Normal}(z, V)$.

**Theorem 1.** Suppose (2.3) is modified as

$$\gamma_t \sim \text{Gamma}(\gamma, \lambda), \quad Z_t|\gamma_t \sim T_C \text{Normal}(z, V/(2\gamma_t)), \quad (3.2)$$

with the region $C$ satisfying the stability condition (3.1). Then $(\tau_t, \theta_t^T, Y_{t-k+1,t})$ has a stationary distribution under which $(\theta_t^T, \tau_t)$ has the same distribution as $(Z_t^T, \gamma_t)$ in (3.2) and

$$Y_t|\theta_t^T, \tau_t \sim \text{Normal}(\mu_t/(1 - \alpha_{1,t} z - \cdots - \alpha_{k,t} z^k), (2\tau_t)^{-1}v_t),$$

where $v_t = \sum_{j=0}^{\infty} \beta_{j,t}^2$ and $\beta_{j,t}$ are the coefficients in the power series representation of $1/(1 - \alpha_{1,t} z - \cdots - \alpha_{k,t} z^k) = \sum_{j=0}^{\infty} \beta_{j,t} z^j$ for $|z| \leq 1$. Moreover, the Markov chain $(\tau_t, \theta_t^T, Y_{t-k+1,t})$ is reversible if it is initialized at the stationary distribution.

The proof of Theorem 1 is given in Appendix A. Since $(\tau_t, \theta_t^T, Y_t)$ is reversible, the backward filter of $(\tau_t, \theta_t^T)$ based on $Y_n, \cdots, Y_{t+1}$ has the same structure as the forward
predictor based on the past \( n - t \) observations prior to \( t \). As in Proposition 4.2 of Yao (1984), we can apply Bayes’ theorem to combine the forward and backward filters, yielding

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

(3.3)

where \( \pi \) denotes the stationary density function which is the same as that of \((\gamma_t, Z_t)\) given in (3.2).

Because the truncated normal is used in lieu of the normal prior distribution in (3.1), the filtering formulas in Section 2 need to be modified somewhat. Specifically, the conditional distribution of \( \theta_n \) given \( \tau_n \) needs to be replaced by \( \theta_n|\tau_n \sim T_{C}\text{Normal}(z_{J_n,n}, V_{J_n,n}/(2\tau_n)) \), while (2.5) defining \( V_{j,n}, z_{j,n} \) and \( a_{j,n} \) remains unchanged. Moreover, Lemma 1 that gives the conditional density of \( Y_n \) given \( J_n \) and \( Y_{J_n,n-1} \) (which is used in the updating formula (2.7) for the weights \( p_{j,n} \)) needs to be modified as follows.

**Lemma 2** (i) Suppose that the conditional distribution of \( Y \) given \( \theta \) and \( \tau \) is \( \text{Normal}(\theta^T \phi, (2\tau)^{-1}) \), \( \tau \sim \text{Gamma}(\bar{\gamma}, \bar{\lambda}) \) and that the conditional distribution of \( \theta \) given \( \tau \) is \( T_{C}\text{Normal}(z, V/(2\tau)) \). Then (2.8) has density function \( f_C \) which approaches the Student-t density with \( 2\bar{\gamma} \) degrees of freedom as the truncation region \( C \) approaches the support \( R^{k+1} \) of the normal distribution.

(ii) Given \( J_n = j \) and \( Y_{J_n,n-1} \), the conditional distribution of \( (\theta_{n}^T, \tau_n) \) is

\[ \tau_n \sim \text{Gamma}(g + (n - j)/2, 1/a_{j,n-1}), \quad \theta_n|\tau_n \sim T_{C}\text{Normal}(z_{j,n-1}, V_{j,n-1}/(2\tau_n)) \text{ if } j < n; \]
\[ \tau_n \sim \text{Gamma}(g, \lambda), \quad \theta_n|\tau_n \sim T_{C}\text{Normal}(z, V/(2\tau_n)) \text{ if } j = n. \]

4. Bounded Complexity Approximations

For the mean shift model described in Section 1, Lai, Liu and Liu (2003) introduced a bounded complexity mixture (BCMIX) approximation to the Bayesian filter \( E(\theta_t|Y_{1,n}) \), while Chen and Lai (2003) developed sequential Monte Carlo approximations to both the filter and the smoother \( E(\theta_t|Y_{1,n}) \), \( 1 \leq t \leq n \), that involve a fixed and relatively small number of trajectories simulated by sequential importance sampling with resampling (SISR). In this section, we show how bounded complexity estimates of \( \theta_t \) and \( \sigma_t \) can be developed by extending BCMIX and SISR to the change-point autoregressive model.

4.1. BCMIX filters

Although the Bayes filter uses a recursive updating formula (2.7) 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. A simple idea to maintain bounded complexity is to keep only a fixed number $n_p$ of weights at each stage $n$ (which is tantamount to setting the other weights to be 0). Lai, Liu and Liu (2003) proposed to keep the most recent $m_p$ weights $p_{j,n}$ (with $n - m_p < j \leq n$) and the largest $n_p - m_p$ of the remaining weights, where $1 \leq m_p < n_p$. Specifically, the updating formula (2.7) for the weights $p_{j,n}$ is modified as follows. Let $\mathcal{K}_{n-1}$ denote the set of indices $j$ so that $p_{j,n-1}$ 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 (2.7) 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

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

(4.1)

choosing $i_n$ to be the one farthest from $n$ if the minimizing set in (4.1) has more than one element. Define $\mathcal{K}_n = \{n\} \cup (\mathcal{K}_{n-1} - \{i_n\})$ and let $p_{j,n} = p_{j,n}^*/\sum_{i \in \mathcal{K}_n} p_{i,n}^*$ for $j \in \mathcal{K}_n$.

For the implementation of the BCMIX filter when the prior distribution of $\theta_t$ is a truncated normal instead of normal, since the conditional density functions $f$ in (2.7) do not have simple closed-form expressions, we simply ignore the truncation and apply the formulas in Lemma 1 that are derived under the normal prior assumption. The constraint set $C$ only serves to generate non-explosive observations but has little effect on the values of the weights $p_{j,n}$ and on the performance of the BCMIX estimates that compute the posterior means via (2.6). In this connection, we also use some fast algorithms to compute the Student-$t$ density by taking logarithms and applying saddlepoint approximations when the number of degrees of freedom is large. Moreover, in view of (2.5), $V_{j,n}, z_{j,n}$ and $a_{j,n}$ ($j \leq n$) can be updated recursively (in $n$) by making use of the matrix inversion lemma (cf. Caines (1988, p. 824)):

$$
 V_{j,n} = V_{j,n-1} - V_{j,n-1}Y_{n-k,n-1}Y_{n-k,n-1}^T/(1 + Y_{n-k,n-1}^TV_{j,n-1}V_{j,n-1}Y_{n-k,n-1})^2, \text{ if } j < n,
$$

$$
 V_{n,n} = V - VY_{n-k,n-1}Y_{n-k,n-1}^T/(1 + Y_{n-k,n-1}^TVY_{n-k,n-1})^2,
$$

$$
 z_{j,n} = z_{j,n-1} + V_{j,n-1}Y_{n-k,n-1}(Y_{n} - Y_{n-k,n-1}^Tz_{j,n-1})/(1 + Y_{n-k,n-1}^TV_{j,n-1}Y_{n-k,n-1})^2,
$$

$$
 a_{j,n} = a_{j,n-1} + (Y_{n} - Y_{n-k,n-1}^Tz_{j,n-1})^2/(1 + Y_{n-k,n-1}^TV_{j,n-1}Y_{n-k,n-1})^2.
$$

(4.2)

4.2. SISR filters

The Bayes estimate (2.6) can be rewritten in the form

$$
 E(\theta_n^T, \sigma_n^2 | Y_{1,n}) = E\{(a_{j,n}^T, a_{j,n}/(2g + n - J_n - 1)| Y_{1,n} \},
$$

(4.3)

which can be computed by Monte Carlo simulations using the conditional distribution of $J_n$ given $Y_{1,n}$. Let $I_t = (I_{k+1}, \ldots, I_t)$. It is difficult to sample $I_n$ directly from its con-
ditional distribution given $\mathbf{Y}_{1,n}$. As in the mean shift model considered by Chen and Lai (2003), a basic idea behind sequential importance sampling is to sample $I_1, \ldots, I_n$ sequentially from an alternative distribution $Q$ under which $I_t | I_{t-1}$ has the same distribution as $P(I_t = i | I_{t-1}, \mathbf{Y}_{1,t})$, which is Bernoulli assuming the values 1 and 0 with probabilities in the proportion

$$pf(Y_t | I_t = 1) : (1 - p) f(Y_t | I_{t-1}, Y_{1,t-1}, I_t = 0), \tag{4.4}$$

in which the conditional densities are Student-$t$ densities given by Lemma 1. Letting $a_t(p)$ and $b_t(p)$ denote the two terms in (4.4), note that $f(Y_t | I_{t-1}, Y_{t-1}) = a_t(p) + b_t(p)$. As in Example 3 of Chen and Lai (2003), we can rewrite (4.3) as

$$E(\theta_n^T, \sigma_n^2 | \mathbf{Y}_{1,n}) = E_Q \{ w_n(x_{n,n}^T, a_{n,n}/(2g + n - J_n - 1)) / E_Q(w_n) \}, \tag{4.5}$$

where the importance weights can be generated recursively by

$$w_t = w_{t-1} \{ a_t(p) + b_t(p) \}, \quad t \geq k + 2; \quad w_{k+1} = 1. \tag{4.6}$$

When $p$ is small, change-points occur very infrequently. Chen and Lai (2003) propose to pick up more change-points by sampling instead from $Q'$ in which $p$ in (4.4) is replaced by $p' > p$. With $Q$ replaced by $Q'$ in (4.5), the weights $w_t$ in (4.6) are changed to

$$w_t = \begin{cases} 
  w_{t-1} \{ a_t(p') + b_t(p') \} a_t(p) / a_t(p') & \text{if } I_t = 1, \\
  w_{t-1} \{ a_t(p') + b_t(p') \} b_t(p) / b_t(p') & \text{if } I_t = 0.
\end{cases} \tag{4.7}$$

The weight $w_n$ defined recursively by (4.6) (or (4.7) if $Q'$ is used in lieu of $Q$) tends to have a large coefficient of variation for large $n$. To overcome this difficulty, the SISR filter also incorporates occasional resampling (hence the symbol R) to keep the coefficient of variation (cv) within certain bounds. Specifically it draws $m$ samples $I_n^{(i)}$ sequentially from the proposal distribution $Q$ and updates the importance weights $w_t^{(i)}$, $i = 1, \ldots, n$, by the following procedure, starting with $m$ samples $I_{t-1}^{(1)}, \ldots, I_{t-1}^{(m)}$ having weights $w_{t-1}^{(1)}, \ldots, w_{t-1}^{(m)}$ at time $t - 1$:

(a) Draw $\tilde{I}_t^{(j)}$ from (4.4) and update the weight $w_t^{(j)}$ by (4.6), $j = 1, \ldots, m$.

(b) If the $cv^2$ of $\{w_t^{(1)}, \ldots, w_t^{(m)}\}$ exceeds or equals a certain bound, resample from $\{\tilde{I}_t^{(1)}, \ldots, \tilde{I}_t^{(m)}\}$ with probabilities proportional to $\{w_t^{(1)}, \ldots, w_t^{(m)}\}$ to produce a random sample $\{I_t^{(1)}, \ldots, I_t^{(m)}\}$ with equal weights. Otherwise let $\{I_t^{(1)}, \ldots, I_t^{(m)}\} = \{\tilde{I}_t^{(1)}, \ldots, \tilde{I}_t^{(m)}\}$ and return to step (a).
If \( Q' \) is used as the proposal distribution instead, replace \( p \) in (4.4) by \( p' \) and (4.6) by (4.7) in step (a). The theory of the resampling and the choice of the \( \alpha_2 \) bound are discussed in Section 3 of Chen and Lai (2003), where it is shown that the SISR filter with suitable choice of the \( \alpha_2 \) bound can provide good approximations to the Bayes filter with as few as \( m = 50 \) simulated trajectories.

### 4.3. Bounded complexity smoothers

As pointed out in Section 3, if the prior normal distribution for \( \theta_t \) is truncated within a stability region \( C \), then \((\tau_t, \theta_t^T, Y_{t-k+1,t})\) is reversible and we can form the Bayesian smoother by combining the forward and backward filters. We shall ignore the truncation in implementing these forward and backward filters. Specifically, let \( \tilde{I}_n = 1, \tilde{I}_t = 1, t \notin (\theta_t^T, \sigma_2) \notin (\theta_{t+1}, \sigma_{2+1}) \) and \( \tilde{J}_t = \min\{j > t|\tilde{J}_j = 1\} \). Let \( p_{t,i,t} = P(J_t = i|Y_{1,t}), \tilde{p}_{j,t} = P(\tilde{J}_t = j|Y_{j+1,n}) \), and note that
\[
\sum_{t=k+1}^n p_{t,t} = 1 = \sum_{j=t+1}^n \tilde{p}_{j,t}.
\]
The backward weights \( \tilde{p}_{j,t} \) can be determined by backward induction on \( t \) using an analog of (2.7).

Analogous to (2.4)-(2.5), it is shown in Appendix B that for \( i \leq t < j \leq n \), the conditional distribution of \((\theta_t, \tau_t)\) given \( J_t = i, \tilde{J}_t = j \) and \( Y_{i,j} \) can be described by
\[
\tau_t \sim \text{Gamma}\( g + \frac{n - i - j + 2}{2}, \frac{1}{2a_{i,j,t}}\), \quad \theta_t|\tau_t \sim \text{Normal}\( z_{i,j,t}, \frac{1}{2\tau_t}V_{i,j,t} \),
\]
if we ignore the truncation in the truncated normal distribution in (3.2), where
\[
V_{i,j,t} = (V_{i,t} - V_{j,t})^{-1},
\]
\[
a_{i,j,t} = a_{i,t} + \tilde{a}_{j,t+1} - \lambda^{-1} + z_{i,t}^\top V_{i,t}^{-1} z_{i,t} + \tilde{z}_{j,t+1}^\top \tilde{V}_{j,t+1} \tilde{z}_{j,t+1} - z^\top V^{-1} z
\]
\[
- (V_{i,t}^{-1} z_{i,t} + \tilde{V}_{j,t+1} \tilde{z}_{j,t+1} - V^{-1} z)^\top V_{i,j,t} (V_{i,j,t}^{-1} z_{i,t} + \tilde{V}_{j,t+1} \tilde{z}_{j,t+1} - V^{-1} z),
\]
\[
z_{i,j,t} = V_{i,j,t} (V_{i,t}^{-1} z_{i,t} + \tilde{V}_{j,t+1} \tilde{z}_{j,t+1} - V^{-1} z),
\]
where \( V_{i,t}, z_{i,t} \) and \( a_{i,t} \) are defined in (2.5) and \( \tilde{V}_{j,t}, \tilde{z}_{j,t} \) and \( \tilde{a}_{j,t} \) are defined similarly by reversing time. Using (3.3) and (4.8), it is shown in Appendix B that analogous to (2.6),
\[
E(\sigma_t^2|Y_{1,n}) = p \sum_{i=k+1}^t \frac{a_{i,t}}{2g + t - i - 1} + (1 - p) \sum_{i=k+1}^t \sum_{j=i+1}^n p_{i,j} \tilde{p}_{j,t} a_{i,j,t} / 2g + n - i - j,
\]
\[
E(\theta_t|Y_{1,n}) = p \sum_{i=k+1}^t z_{i,t} + (1 - p) \sum_{i=k+1}^t \sum_{j=t+1}^n p_{i,j} \tilde{p}_{j,t} z_{i,j,t},
\]
in which the approximation ignores truncation within \( C \). The BCMIX smoother further approximates (4.7) by allowing at most \( n_p \) weights \( p_{i,t} \) and \( n_p \) weights \( \tilde{p}_{j,t} \) to be nonzero, as described in Section 4.4.
The SISR smoother can be formed in the similar way. Define $I_t$ and $Q$ as in Section 4.2. Let $\mathbf{I}_t = (\tilde{I}_t, \ldots, \tilde{I}_n)$ and define $\tilde{Q}$ similarly so that $I_t|\mathbf{I}_{t+1}$ has the Bernoulli distribution assuming the value 1 and 0 with probabilities in the proportion

$$pf(Y_t|\tilde{I}_t = 1) : (1 - p) f(Y_t|Y_{t+1}, \tilde{I}_t, \tilde{I}_t = 0). \quad (4.10)$$

The SISR forward filter is sampled from $Q$ and the backward filter from $\tilde{Q}$, yielding importance weights $w_t$ defined recursively by (4.6) and $\tilde{w}_t$ defined similarly by backward induction, with occasional resampling to keep the coefficients of variation of $w_t$ and $\tilde{w}_t$ within certain bounds. With $m$ forward and $m$ backward trajectories simulated in this way, the SISR smoother for $(\theta_t, \sigma_t^2)$ can be expressed as

$$p \sum_{i=1}^{m} w_t^{(i)} \left( \begin{array}{c} \mathbf{z}_{j_t}^{(i)} \frac{a_{j_t}^{(i),t}}{2g + t - J_t^{(i)} - 1} \end{array} \right) / \left( \sum_{i=1}^{m} w_t^{(i)} \right)$$

$$+ (1 - p) \sum_{i=1}^{m} \sum_{j=1}^{m} w_t^{(i)} \tilde{w}_t^{(j)} \left( \begin{array}{c} \mathbf{z}_{j_t}^{(i),\tilde{J}_t^{(i)}} \frac{a_{j_t}^{(i),\tilde{J}_t^{(i)}}}{2g + n - \tilde{J}_t^{(i)} - \tilde{J}_t^{(j)}} \end{array} \right) / \left( \sum_{i=1}^{m} w_t^{(i)} \right) \left( \sum_{j=1}^{m} w_t^{(j)} \right), \quad (4.11)$$

which is analogous to (4.9); see Appendix B for the derivation.

4.4. Simulation studies

The top panel of Figure 1 plots a time series of $n = 3000$ observations generated from the change-point AR(2) model with

$$p = 0.001, \gamma_t \sim \text{Gamma (3, 4)}, \mathbf{Z}_t | \gamma_t \sim \mathbf{T}_C \text{ Normal (0, I)}, \quad (4.12)$$

where $C = \{ (\mu, \alpha_1, \alpha_2)^T : |\alpha_1| + |\alpha_2| < 1 \}$. There are 2 change-times in the dataset and the piecewise constant parameter values are

$$(\mu_t, \sigma_t, \alpha_{1,t}, \alpha_{2,t}) = \begin{cases} 
(0.5019, -0.2171, -0.8360, 0.0629) & \text{if } 1 \leq t < 943, \\
(0.8723, 1.0373, -0.0328, 0.2855) & \text{if } 943 \leq t < 1623, \\
(0.5970, 0.1043, -0.1115, 0.4333) & \text{if } 1623 \leq t \leq 3000.
\end{cases} \quad (4.13)$$

The Bayes estimates, $E(\sigma_t^2 | Y_{1:t})$ and $E(\mu_t + \alpha_{1,t} Y_{t-1} + \alpha_{2,t} Y_{t-2} | Y_{1:t})$, of the variance and regression function are also plotted in the middle and bottom panels, together with the corresponding BCMIX estimates (with $n_p = 25$, $m_p = 10$) and SISR estimates (based on $m = 100$ SISR trajectories).
Table 1(a) reports simulation results on the sum of squared errors

$$\text{SSE} := \sum_{t=3}^{n} \{(1, Y_{t-1}, Y_{t-2})(\widehat{\theta}_t - \theta_t)\}^2$$

(4.14)

and the Kullback-Leibler divergence between the true and estimated parameter values, defined by

$$\text{KL} = \sum_{t=3}^{n} \left\{ \frac{[1, Y_{t-1}, Y_{t-2})(\widehat{\theta}_t - \theta_t)]^2}{\sigma_t^2} + \left( \frac{\sigma_t^2}{\sigma_t^2} - 1 - \log \frac{\sigma_t^2}{\sigma_t^2} \right) \right\},$$

(4.15)

with $(\widehat{\theta}_t, \widehat{\sigma}_t)$ being the Bayes, BCMIX and SISR filters in the respective columns. Note that the quantity inside the curly brackets of (4.15) is $E[\log\{f_{\theta_t, \sigma_t}(Y^*_t)/f_{\widehat{\theta}_t, \widehat{\sigma}_t}(Y^*_t)\}]$, where the expectation is conditional on $(\widehat{\theta}_t, \widehat{\sigma}_t)$ and taken under the true parameter $(\theta_t, \sigma_t)$, and $Y^*_t$ has density function $f_{\theta_t, \sigma_t}$ and is independent of $(\widehat{\theta}_t, \widehat{\sigma}_t)$. Each result in Table 1(a) is based on 100 simulations generated from the change-point AR(2) model whose parameters are given in (4.12) and also for several other values of $p$ listed in the table. For $p = 0.0005$ we choose $n = 10,000$, whereas for the larger values of $p$ we take $n = 5,000$, noting that the expected number of change-points in each simulated sequence is $np$. Note that although SSE for BCMIX can be more than twice that for the Bayes filter, the KL for BCMIX remains within 1.2 of that for the Bayes filter. Moreover, the SISR filter has a smaller SSE but larger KL than BCMIX. Whereas the Bayes filter at time $t$ consists of a mixture of $t$ components and $t$ can increase up to 5000 (or 10000), the BCMIX filter involves a mixture of only $np = 25$ components. If we use a larger $np$ for smaller value of $p$, then the performance of BCMIX can be improved markedly. An asymptotic theory showing that $np$ should be chosen at a slightly larger order of magnitude than $\log(1/p)$ as $p \to 0$ is given in Lai, Liu and Liu (2003).

It is also worth noting that when the SSE of BCMIX exceeds twice that of the Bayes filter in Table 1, the SSE values for both the Bayes and BCMIX filters are small (< 0.15) when divided by $n - 2$ (yielding SSE per observation).

**INSERT TABLE 1 ABOUT HERE**

While Table 1(a) considers the Bayes risks of various estimators of $\theta_t$ and $\sigma_t^2$ and generates each simulation from the Bayesian model, Table 1(b) considers the frequentist risks for three fixed piecewise constant specifications of $(\theta_t, \sigma_t)$. The first specification, called Case 1, is the same as that in Figure 1; see (4.13). The other two specifications allow unit-root
nonstationarity in the AR(2) model. Specifically, Case 2 (or 3) has the same $\theta_t, \sigma_t$ values as in Case 1 except that $(\mu_t, \alpha_{1,t}, \alpha_{2,t})$ takes the value $(0, 1, 0)$ for $943 \leq t < 1623$ in Case 2 (or for $t \geq 1623$ in Case 3). The KL of BCMIX ranges between 1.05 to 1.30 times that of the Bayes filter, although BCMIX has a markedly larger SSE. For Case 3, both the KL and the SSE of SISR are over two times those of the Bayes filter. To handle the unit-root nonstationarity for $t \geq 1623$, it seems necessary to increase the number of simulated SISR trajectories by ten or more times the number 100 used in Table 1.

We next consider the performance of the bounded complexity smoothers. For $n$ of the size considered in Table 1, it is difficult to compute the Bayesian smoother $E(\theta_t^T, \sigma_t^2|Y_{1,n})$ of Section 3. Instead of this computationally prohibitive benchmark for comparison with BCMIX smoothers, we consider a much simpler benchmark in which the change-points are known so that the Bayes estimates of $\theta_t, \sigma_t^2$ between two change-points are given by the standard Bayesian formulas for normal populations (cf. Section 2.7 of Box and Tiao (1973)). Table 2 compares this "fictitious Bayes" smoother with the BCMIX smoother in terms of the SSE and KL (for which the sums in (4.14) and (4.15) are now replaced by $\sum_{t=3}^{n-2}$ to allow for backward filtering). Comparison of Table 2 with Table 1 shows the substantially smaller SSE and KL for BCMIX smoothers than the BCMIX filters.

INSERT TABLE 2 ABOUT HERE

We have not included SISR smoothers in the simulation study of Table 2 because of the computational cost of many simulated replicates of such smoothers, which involve another layer of SISR simulations. Simulating SISR smoothers for some particular cases yields results similar to those for BCMIX smoothers. In particular, for the simulated data in Figure 1, we evaluated both the BCMIX and SISR (with $m = 100$) smoothers and the results for the estimates of $\sigma_t^2$ are shown in Figure 2. Figure 2 also plots another SISR smoother, denoted by SISR* and to be described in Section 5, that puts a Beta distribution on $p$ instead of assuming it to be known. All smoothers are close to each other, and they are visually indistinguishable for the regression function estimates, which are therefore omitted from Figure 2.

INSERT FIGURE 2 ABOUT HERE

5. Choice of Hyperparameters and a Forecasting Application

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
Bayesian change-point autoregressive model from the data. For the mean shift model in Section 1, Yao (1984) considered the maximum likelihood approach. He developed an algorithm to compute the likelihood function with \(O(n^2)\) operations but found it to be impractical to be used in conjunction with iterative search for the maximum likelihood estimate. Accordingly he developed an approximation to the likelihood function by collapsing the mixture distribution for \(\theta_t | Y_{1:t-1}\) to a single normal and maximized this pseudo-likelihood instead. In this section we describe two methods to estimate the hyperparameters, one based on the accumulated prediction error (APE) criterion to be used in conjunction with the BCMIX algorithm, and the other assuming a prior distribution on the hyperparameters to be used in conjunction with the SISR estimates. Not only do both methods have relatively low computational costs in comparison with likelihood methods to estimate the hyperparameters, but they also lead to good estimates of \(\sigma_t\) and \(\theta_t\), as illustrated by Table 3 below. Section 5.3 applies the BCMIX method, with \(p\) determined by the APE criterion, to construct one-step-ahead forecasts of weekly returns of the NASDAQ index based on 100 stocks using the National Association of Securities Dealers Automatic Quotations Service.

5.1. APE criterion and BCMIX estimates

The accumulated prediction error (APE) criterion was introduced by Rissanen (1986) and applied to order determination in classical time series models by Hemmerly and Davis (1989), Wei (1992) and Lai and Lee (1997). Because APE\((\nu)\) can be computed recursively, it is well suited to recursive estimators in time series models that involve an unspecified order \(\nu\) to be determined from the data, and to BCMIX filters in change-point autoregressive models involving an unspecified hyperparameter \(\nu\) that consists of \(p, g, \lambda, z\) and \(V\).

For the change-point AR\((k)\) model, the accumulated prediction error at time \(n\) is defined by

\[
APE_n(\nu) = \sum_{t=k+1}^{n} \{Y_t - \hat{Y}_{t|t-1}(\nu)\}^2,
\]

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

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

in which \(\hat{\theta}_t(\nu)\) is the BCMIX estimate (assuming the hyperparameter \(\nu\)) of \(\theta\) based on \(Y_1, \ldots, Y_t\). For the mean shift model described in Section 1, Lai, Liu and Liu (2003) proposed to use the APE criterion to choose the hyperparameter at every stage \(n\) from a given set \(\mathcal{H}\) of possible values and provided an asymptotic theory for this approach. Extending
their approach to the current setting, let $\hat{\nu}_n$ be the minimizer of $\text{APE}_n(\nu)$ over $\nu \in \mathcal{H}$ and estimate $\theta_t$ and $\sigma_t^2$ by $\hat{\theta}_t(\hat{\nu}_{t-1})$ and $\hat{\sigma}_t^2(\hat{\nu}_{t-1})$. These estimates will be called BCMIX-APE and compared with BCMIX(\nu) that assumes the hyperparameter $\nu$ to be known.

When $\mathcal{H}$ is a finite set, we can compute BCMIX-APE by $h$ parallel recursions. Specifically, for each $\nu \in \mathcal{H}$, $\hat{\theta}_{t-1}(\nu)$ and $\hat{\sigma}^2_{t-1}(\nu)$ can be updated recursively in view of (2.7) and (4.2), and therefore we can easily update $\text{APE}_t(\nu)$ by $\text{APE}_t(\nu) = \text{APE}_{t-1}(\nu) + \{Y_t - \hat{Y}_{t|t-1}(\nu)\}^2$. To choose a finite $\mathcal{H}$, one can start by using prior information to come up with a range $(0 <) p' \leq p \leq p''(< 1)$. Typically one has an upper bound (horizon) $N$ 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 Liu (2003) propose to replace the interval $[p', p'']$ by a finite set of points $2^\ell p'$ ($0 \leq \ell \leq L$) with $L = \max\{\ell : 2^\ell p' \leq p''\}$. Whereas the relative frequency $p$ of change-points is an essential feature of our Bayesian 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. Thus the set $\mathcal{H}$ chosen in this way typically has manageable cardinality.

As an illustration, we consider the change-point AR(2) model whose hyperparameters are given in (4.12) and also for several other values of $p$, as in Table 1(a). For each value of $p$, we let $p' = p/10$ and $p'' = 10p$ and define $\mathcal{H}$ with cardinality $2L$, where $L$ is defined above and each $(p, g, \lambda, z, V) \in \mathcal{H}$ has the following form: $p = 2^{\ell} p'$ and either (i) $g = 4, \lambda = \frac{1}{10}, z = \frac{1}{2} 1, V = \frac{3}{2} I$, or (ii) $g = \frac{5}{2}, \lambda = \frac{1}{3}, z = -\frac{1}{3} 1, V = 2I$, where $I$ is the identity matrix and $1$ is a vector of 1's. Note that $\sigma^2_t$ has a prior mean of $5/3$ under (i) and 1 under (ii). The BCMIX-APE estimate of $(\theta^*_t, \sigma^*_t)$ constructed in this way is compared with BCMIX(\nu) in terms of the Kullback-Leibler divergence (4.15) in Table 3, each result of which is based on 100 simulations.

INSERT TABLE 3 ABOUT HERE

Extending the APE criterion to the smoothing problem yields the usual cross validation criterion

$$CV(\nu) = \sum_{t=3}^{n-2} \{Y_t - \hat{Y}_{t,n}(\nu)\}^2,$$

where $\hat{Y}_{t,n}(\nu)$ is a predictor of $Y_t$ based on $(Y_{1,t-1}, Y_{t+1,n})$ obtained by combining the forward and backward BCMIX estimates in a way similar to that in Section 4.3; see Appendix C for
details. The hyperparameter $\nu$ can be estimated for BCMIX smoothers by $\tilde{\nu}_n$ that minimizes $\text{CV}(\nu)$ over $\nu \in \mathcal{H}$.

5.2. Conjugate hyperprior distributions and SISR estimates

Suppose that in the change-point AR($k$) model the probability $p$ of change is unknown and is specified by a prior Beta($a, b$) distribution with mean $a/(a + b)$, where $a$ and $b$ are positive integers. Using the proposal distribution $Q$ for which $I_t|I_{t-1}$ has the same distribution as $P(I_t = 1|I_{t-1}, Y_{1:t})$, it can be shown that $I_t|I_{t-1}$ is Bernoulli assuming the values 1 and 0 with probabilities in the proportion

$$
\left( \frac{n_{t-1,1} + a}{t - 1 + a + b} \right) f(Y_t|I_t = 1) : \left( \frac{n_{t-1,0} + b}{t - 1 + a + b} \right) f(Y_t|Y_{I_{t-1}, t-1}, I_t = 0),
$$

(5.4)

where the conditional densities are Student-$t$ densities given by Lemma 1, and $n_{t-1,1}$ and $n_{t-1,0}$ are the number of 1’s and 0’s in $\{I_1, I_2, \ldots, I_{t-1}\}$ as noted by Chen and Lai (2003) for the simpler mean shift model. Letting $a_t$ and $b_t$ denote the two terms in (5.3), note that $f(Y_t|I_{t-1}, Y_{t-1}) = a_t + b_t$ and that the importance weights can be generated recursively by $w_t = w_{t-1}(a_t + b_t)$, as in (4.6).

By combining these adaptive forward filters with the corresponding backward filters, we obtain adaptive smoothers; see Section 5.2 of Chen and Lai (2003) for details in a similar model. In Figure 2, the SISR* smoother is computed by this method. Note how close it is to the BCMIX and SISR smoothers that assume $p$ to be known.

5.3. Forecasting NASDAQ weekly returns

Figure 3 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 $r_t = \log(P_t/P_{t-1})$. The data consisting of 982 observations are available at http://finance.yahoo.com. We model the time series as a change-point AR(2) model and use the BCMIX-APE procedure to estimate the parameters $\mu_t, \alpha_{1,t}, \alpha_{2,t}$ and $\sigma_t^2$. The BCMIX filter is applied with $n_p = 50$ and $m_p = 20$. The hyperparameter $p$ is chosen from the subset $2p'$ in the range $p' = .001 \leq p \leq p'' = .05$, and we only use one choice $(g, \lambda, z, V) = (2.4, 2000, 0, 1)$. This choice is based on some rough estimate of the volatility based on weekly returns in the year before the start date of the time series.

Figure 3 considers application of the BCMIX-APE filter $\hat{\Theta}_t = (\hat{\mu}_t, \hat{\alpha}_{1,t}, \hat{\alpha}_{2,t})^T$ and the APE estimate $\hat{p}_t$ of $p$, based on data up to week $k$, to forecast next week’s return $Y_{t+1}$, which
is modeled by $Y_{t+1} = \theta_{t+1}^T Y_{t-1,t} + \sigma_{t+1} \epsilon_{t+1}$. The one-week-ahead forecast is given by

$$\hat{Y}_{t+1|t} = (1 - \hat{\rho}_t) \theta_t^T Y_{t-1,t}$$  \hspace{1cm} (5.5)$$

and is plotted in Figure 3, recalling that $z = 0$. In risk management applications, one is interested in certain quantiles of $Y_{t+1}$; see Section 7.1 of Tsay (2002). In particular, the one-week 99\% VaR (value at risk) of an investment on the NASDAQ index involves the estimated first percentile of $Y_{t+1}$, i.e. $\hat{Y}_{t+1|t} - 2.33\hat{\sigma}_{t+1|t}$, where $\hat{\sigma}^2_{t+1|t}$ is the BCMIX-APE estimate of $\sigma^2_{t+1}$ (cf. (2.6)). Besides this estimated first percentile, Figures 3 also plots the estimated 99th percentile $\hat{Y}_{t+1|t} + 2.33\hat{\sigma}_{t+1|t}$. Note that the actual returns lie within these upper and lower quantiles of one-week-ahead forecasts, even during the highly volatile periods around October 1987 and after September 11, 2001.

INSERT FIGURE 3 ABOUT HERE

6. Conclusion

The change-point AR($k$) model introduced herein is a simple Bayesian model that captures 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, we have been able to develop two approximations, with relatively low computational complexity, to the Bayesian filters and smoothers. The first is the simulation-based SISR which uses recent advances in sequential Monte Carlo methods. The second is BCMIX, which is developed from explicit formulas for the Bayesian filters.

The Bayesian model has certain hyperparameters among which is the relative frequency $p$ of change-points in the time series. We have described two general approaches to determining hyperparameters from the data. Omitted from the discussion in the preceding sections is that one often has for a particular application some external information, and the strength of Bayesian modeling is that one can conveniently incorporate it into the Bayesian model. For example, since $p$ is a hyperparameter that is sequentially determined from the data in Section 5, we can incorporate external information besides using a criterion like APE (which depends solely on the observed time series) to determine $p$. Such external information plays a fundamental role in the intervention analysis of Box and Tiao (1975). They noted that certain external events (e.g., the diversion of traffic by the opening of a new freeway, or a
new law on the allowable proportion of reactive hydrocarbons in gasoline) might produce structural changes in an observed time series (such as hourly readings of oxidant pollution level in a town). Their approach is to model a known intervention at a certain time point $t_0$ by an input of the form $1_{t \geq t_0}$ in an ARMAX model. The ARMAX model, however, only involves the dynamics of the level of $Y_t$ but not its volatility. If we use the change-point AR$(k)$ model instead to model both the level and volatility, we can adjust the hyperparameter $p$ at different time points to incorporate knowledge of external events such as interventions at these times.

Acknowledgment

This research was supported by the National Science Foundation.

Appendix A: Proof of Theorem 1

Lemma 3. (i) If $\{Y_t, 1 \leq t \leq n\}$ is a stationary Gaussian sequence, then it is reversible, i.e. it has the same distribution as the time-reversed sequence $\tilde{Y}_t := Y_{n+1-t}$.

(ii) Let $Y_t = \mu + \alpha_1 Y_{t-1} + \cdots + \alpha_k Y_{t-k} + \sigma \epsilon_t$, $t > k$, in which the $\epsilon_t$ are i.i.d. standard normal and $(\alpha_1, \alpha_2, \cdots, \alpha_k)$ satisfies the stability assumption (3.1). Then $\{Y_{t-k:t-1}, t > k\}$ is a geometrically ergodic Markov chain having a normal stationary distribution. In particular, $Y_n$ has a limiting normal distribution with mean $\mu/(1 - \alpha_1 - \cdots - \alpha_k)$ and variance $\nu = \sigma^2 \sum_{j=0}^{\infty} \beta_j^2$, where $\beta_j$ are the coefficients in the power series representation of $1/(1 - \alpha_1 z - \cdots - \alpha_k z^k) = \sum_{j=0}^{\infty} \beta_j z^j$ for $|z| \leq 1$.

(iii) For the AR$(k)$ model in (ii), if $Y_{1:k}$ is initialized at the stationary distribution, then $\{Y_t, k < t < n-k\}$ is reversible.

Proof. (i) follows easily from the covariance function of the stationary Gaussian sequence. To prove (ii), note that $Y_{t-k+1:t} = A Y_{t-k:t-1} + (0, \mu + \sigma \epsilon_t, 0, \cdots, 0)^T$, where the first row of $A$ is $(1, 0, \cdots, 0)$, the second row is $(0, \alpha_1, \cdots, \alpha_k)$, the third row is $(0, 1, 0, \cdots, 0)$, etc. It is a geometrically ergodic Markov chain (cf. Meyn and Tweedie (1993)). If the chain is initialized at its stationary distribution, then $Y_n$ can be written as an infinite moving average $\sum_{j=0}^{\infty} \beta_j (\mu + \sigma \epsilon_{n-j})$, so (iii) follows from (i).

Proof of Theorem 1. First note that the probability measure $Q$, under which $(\tau_{k+2}, \theta_{k+2}^T)$ has the same distribution as (3.2) and is independent of the Bernoulli random variable $I_{k+2}$, is an invariant measure (stationary distribution) for the Markov chain $\{(\tau_t, \theta_t^T, I_t), t \geq k+2\}$. Moreover, the chain clearly satisfies the geometric drift condition (V4) of Meyn and Tweedie (1993, p.367) and is reversible (since the $I_t$ are i.i.d.). Combining this with parts (ii) and
Appendix B: Proof of (4.8), (4.9) and (4.11)

From (3.3) it follows that

\[
\begin{align*}
\frac{f(\theta_t^T, \sigma_t^2 | Y_{1:n})}{f(\theta_t^T, \sigma_t^2)} & \propto \frac{f(\theta_t^T, \sigma_t^2 | Y_{1:t})f(\theta_t^T, \sigma_t^2 | Y_{t+1:n})}{f(\theta_t^T, \sigma_t^2)} \\
& \propto \frac{f(\theta_t^T, \sigma_t^2 | Y_{1:t})}{f(\theta_t^T, \sigma_t^2)} \left[ f(\theta_t^T, \sigma_t^2 | \tilde{I}_t = 1)P(\tilde{I}_t = 1) + f(\theta_t^T, \sigma_t^2 | Y_{t+1:n}, \tilde{I}_t = 0)P(\tilde{I}_t = 0) \right] \\
& \propto pf(\theta_t^T, \sigma_t^2 | Y_{1:t}) + (1-p)f(\theta_t^T, \sigma_t^2 | Y_{1:t}, Y_{t+1:n}, \tilde{I}_t = 0)
\end{align*}
\]

(B.1)

The second term in the last line of (B.1) can be expressed as \((1-p)\sum_{i=k+1}^{t} \sum_{j=t+1}^{n} p_{i,t} \tilde{p}_{j,t} e_{i,j,t}\), where for \(i \leq t < j \leq n\),

\[
\begin{align*}
e_{i,j,t} &= \frac{f(\theta_t^T, \sigma_t^2 | Y_{1:t}, J_t = i)f(\theta_t^T, \sigma_t^2 | Y_{t+1:n}, \tilde{J}_t = j)}{f(\theta_t^T, \sigma_t^2)} \\
&= \frac{\text{Normal}(z_{i,t}, V_{i,t}/(2\tau_t))\text{Normal}(\tilde{z}_{j,t}, V_{j,t}/(2\tau_t))}{\text{Normal}(z, V/(2\tau_t))} \\
& \quad \times \frac{\text{Gamma}(g + (t - i + 1)/2, 1/a_{i,t})\text{Gamma}(g + (n - t - j + 1)/2, 1/a_{j,t})}{\text{Gamma}(g, \lambda)} \\
&= \text{Normal}(z_{i,j,t}, V_{i,j,t}/(2\tau_t)) \times \text{Gamma}(g + \frac{n - i - j}{2}, \frac{1}{a_{i,j,t}}).
\end{align*}
\]

(B.2)

In (B.2), we have used \text{Gamma}(\cdot, \cdot) to denote the gamma density function of \(\tau_t := (2\sigma_t^2)^{-1}\), with the indicated shape and scale parameters, and \text{Normal}(\cdot, \cdot) to denote the normal density function of \(\theta_t\) given \(\tau_t\), with the indicated mean and covariance matrices. Note that by Bayes’ theorem, (B.1) also gives the conditional distribution of \((\theta_t, \tau_t)\) given \(J_t = i, \tilde{J}_t = j\) and \(Y_{i,j}\), thus proving (4.8). Combining (B.1) with (B.2) and (2.6) yields (4.9). For the SISR smoother, changing \(p_{i,t}\) and \(\tilde{p}_{j,t}\) to \(\tilde{w}_i^{(i)}\) and \(\tilde{w}_i^{(j)}\) in (4.9) gives (4.11).

Appendix C: Derivation of \(\hat{Y}_{t,n}(\nu)\) in (5.3)

Analogous to (3.3), Bayes’ theorem yields

\[
\begin{align*}
f(Y_t | Y_{1:t-1}, Y_{t+1:n}) & \propto f(Y_t | Y_{1:t-1})f(Y_t | Y_{t+1:n})/f(Y_t) \\
& = f(Y_{1:t} | Y_t, Y_{t+1:n})/f(Y_t).
\end{align*}
\]

(C.1)
Let $\tilde{Y}_t := Y_{n+1-t}$,

$$
\begin{align*}
\mu_{i,t} &= \begin{cases} 
    Z_{i,t-1}^T Y_{t-k,t-1} & \text{if } i \leq t - 1, \\
    Z^T \tilde{Y}_{n-t-k+1,n-t} & \text{if } j \geq t + 1,
\end{cases} \\
\tilde{\mu}_{j,t} &= \begin{cases} 
    \tilde{Z}^T_{j,t} \tilde{Y}_{n-t-k+1,n-t} & \text{if } j \geq t + 1, \\
    Z^T \tilde{Y}_{n-t-k+1,n-t} & \text{if } j = t,
\end{cases} \\
\sigma^2_{i,t} &= \begin{cases} 
    a_{i,t-1}(1 + Y_{t-k,t-1}^T V_{i,t-1} Y_{t-k,t-1}) / (2g + t - i) & \text{if } i \leq t - 1, \\
    (1 + Y_{t-k,t-1}^T V Y_{t-k,t-1}) / (2g) & \text{if } i = t,
\end{cases} \\
\sigma^2_{j,t} &= \begin{cases} 
    \tilde{a}_{j,t+1}(1 + \tilde{Y}_{n-t-k+1,n-t}^T \tilde{V}_{j,t+1} \tilde{Y}_{n-t-k+1,n-t}) / (2g + n - t - j) & \text{if } j \geq t + 1, \\
    (1 + \tilde{Y}_{n-t-k+1,n-t}^T V \tilde{Y}_{n-t-k+1,n-t}) / (2g) & \text{if } j = t.
\end{cases}
\end{align*}
$$

Then by Lemma 1(i), conditional on $J_t = i$ and $Y_{i,t-1}$,

$$
(Y_t - \mu_{i,t}) / \sigma_{i,t} \sim \text{Student-t} (2g + t - i)
$$

(C.2)

in which the quantity $2g + t - i$ in parentheses denotes the degrees of freedom. Using time reversal, it follows similarly from Lemma 1(i) that conditional on $\tilde{J}_t = j$ and $Y_{t+1,j}$,

$$
(Y_t - \tilde{\mu}_{j,t}) / \tilde{\sigma}_{j,t} \sim \text{Student-t} (2g + n - t - j).
$$

(C.3)

Although both factors in the numerator of (C.1) have the simple forms (C.2) and (C.3), the denominator is considerably more complicated. We shall ignore the denominator as our goal is to combine the forward and backward predictors of $Y_t$ in (C.2) and (C.3) in a simple way. Using this approximation in (C.1), we can regard $Y_{1,t-1}$ and $Y_{t+1,n}$ as two independent sources of information on $Y_t$ since the numerator in the right hand side of (C.1) factorizes into these two components. Accordingly we weight the forward and backward predictors given by (C.2) and (C.3) by their respective variances, leading to the estimate

$$
\tilde{E}(Y_t | J_t = i, \tilde{J}_t = j, Y_{i,t-1}, Y_{t+1,j}) = \left\{ \frac{\mu_{i,t}}{\sigma^2_{i,t}} + \frac{\tilde{\mu}_{j,t}}{\tilde{\sigma}^2_{j,t}} \right\} / \left\{ \frac{1}{\sigma^2_{i,t}} + \frac{1}{\tilde{\sigma}^2_{j,t}} \right\}. \quad \text{(C.4)}
$$

There is little loss of information in ignoring the denominator in (C.1) unless $\max(t-i, j-t)$ is small, but $(i,j)$ pairs with both $i$ and $j$ near $t$ do not have much predictive value for $Y_t$. We therefore have for any fixed value of the hyperparameter $\nu = (p, g, \lambda, z, V)$ the estimate

$$
\hat{Y}_{t,n}(\nu) = \sum_{i=1}^{t} \sum_{j=t}^{n} \beta_{i,t} \beta_{j,t} \left\{ \frac{\mu_{i,t}}{\sigma^2_{i,t}} + \frac{\tilde{\mu}_{j,t}}{\tilde{\sigma}^2_{j,t}} \right\} / \left\{ \frac{1}{\sigma^2_{i,t}} + \frac{1}{\tilde{\sigma}^2_{j,t}} \right\}.
$$
REFERENCES


Table 1. Sum of squared errors (SSE) and Kullback-Leibler divergence (KL) for Bayes, BCMIX and SISR filters. Standard errors are given in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>SSE</th>
<th></th>
<th></th>
<th></th>
<th>KL</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>p</td>
<td>Bayes</td>
<td>BCMIX</td>
<td>SISR</td>
<td>Bayes</td>
<td>BCMIX</td>
<td>SISR</td>
</tr>
<tr>
<td>(a)</td>
<td>Bayesian</td>
<td>setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>.0005</td>
<td></td>
<td>137.6</td>
<td>233.4</td>
<td>397.9</td>
<td>94.4</td>
<td>108.9</td>
<td>262.6</td>
</tr>
<tr>
<td></td>
<td>(7.72)</td>
<td></td>
<td>(43.83)</td>
<td>(24.90)</td>
<td>(3.67)</td>
<td>(4.75)</td>
<td>(10.60)</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>0.001</td>
<td></td>
<td>122.4</td>
<td>305.8</td>
<td>213.0</td>
<td>84.6</td>
<td>97.0</td>
<td>146.6</td>
</tr>
<tr>
<td></td>
<td>(7.10)</td>
<td></td>
<td>(104.00)</td>
<td>(13.59)</td>
<td>(3.29)</td>
<td>(4.23)</td>
<td>(7.30)</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>0.003</td>
<td></td>
<td>290.3</td>
<td>746.8</td>
<td>432.3</td>
<td>190.6</td>
<td>216.8</td>
<td>283.7</td>
</tr>
<tr>
<td></td>
<td>(11.29)</td>
<td></td>
<td>(260.47)</td>
<td>(16.71)</td>
<td>(4.25)</td>
<td>(5.00)</td>
<td>(6.38)</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>0.01</td>
<td></td>
<td>659.3</td>
<td>1082.0</td>
<td>812.0</td>
<td>437.1</td>
<td>493.3</td>
<td>524.7</td>
</tr>
<tr>
<td></td>
<td>(12.22)</td>
<td></td>
<td>(98.91)</td>
<td>(15.26)</td>
<td>(5.46)</td>
<td>(6.37)</td>
<td>(6.28)</td>
<td></td>
</tr>
<tr>
<td>5000</td>
<td>0.02</td>
<td></td>
<td>1044.7</td>
<td>1614.1</td>
<td>1319.8</td>
<td>693.2</td>
<td>761.9</td>
<td>842.5</td>
</tr>
<tr>
<td></td>
<td>(15.64)</td>
<td></td>
<td>(75.04)</td>
<td>(19.88)</td>
<td>(6.57)</td>
<td>(6.41)</td>
<td>(8.40)</td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td>Frequentist</td>
<td>setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 3000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case 1</td>
<td></td>
<td></td>
<td>32.6</td>
<td>34.2</td>
<td>54.8</td>
<td>41.2</td>
<td>42.7</td>
<td>70.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.72)</td>
<td>(0.77)</td>
<td>(1.62)</td>
<td>(0.77)</td>
<td>(0.83)</td>
<td>(2.02)</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td></td>
<td>27.5</td>
<td>82.6</td>
<td>39.05</td>
<td>36.5</td>
<td>47.4</td>
<td>48.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.89)</td>
<td>(6.95)</td>
<td>(3.61)</td>
<td>(0.94)</td>
<td>(0.98)</td>
<td>(2.53)</td>
</tr>
<tr>
<td>Case 3</td>
<td></td>
<td></td>
<td>32.2</td>
<td>34.2</td>
<td>77.02</td>
<td>41.4</td>
<td>43.6</td>
<td>99.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.70)</td>
<td>(0.73)</td>
<td>(4.86)</td>
<td>(0.77)</td>
<td>(0.85)</td>
<td>(6.15)</td>
</tr>
</tbody>
</table>
Table 2. Sum of squared errors (SSE) and Kullback-Leibler divergence (KL) for “fictitious Bayes” (fBayes) and BCMIX smoothers. Standard errors are given in parentheses.

<table>
<thead>
<tr>
<th>n</th>
<th>p</th>
<th>fBayes</th>
<th>BCMIX</th>
<th>fBayes</th>
<th>BCMIX</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>0.005</td>
<td>77.7</td>
<td>164.7</td>
<td>44.4</td>
<td>55.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(30.55)</td>
<td>(41.96)</td>
<td>(7.96)</td>
<td>(4.34)</td>
</tr>
<tr>
<td>5000</td>
<td>0.001</td>
<td>80.8</td>
<td>158.6</td>
<td>65.7</td>
<td>47.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(31.96)</td>
<td>(32.67)</td>
<td>(16.76)</td>
<td>(3.61)</td>
</tr>
<tr>
<td>5000</td>
<td>0.003</td>
<td>229.7</td>
<td>661.6</td>
<td>142.1</td>
<td>130.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(49.12)</td>
<td>(194.39)</td>
<td>(24.26)</td>
<td>(4.70)</td>
</tr>
<tr>
<td>5000</td>
<td>0.01</td>
<td>526.2</td>
<td>1390.7</td>
<td>304.0</td>
<td>372.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(36.05)</td>
<td>(126.4)</td>
<td>(12.89)</td>
<td>(5.97)</td>
</tr>
<tr>
<td>5000</td>
<td>0.02</td>
<td>1347.5</td>
<td>2382.0</td>
<td>467.0</td>
<td>641.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(528.36)</td>
<td>(419.14)</td>
<td>(9.58)</td>
<td>(8.10)</td>
</tr>
</tbody>
</table>

Table 3. Kullback-Leibler divergence of BCMIX(\nu) and BCMIX-APE filters with n_p = 35. Standard errors are given in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>m_p = 5</th>
<th></th>
<th>m_p = 0</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>BCMIX(\nu)</td>
<td>BCMIX-APE</td>
<td>BCMIX(\nu)</td>
<td>BCMIX-APE</td>
</tr>
<tr>
<td>0.0005</td>
<td>96.0 (3.68)</td>
<td>105.8 (4.24)</td>
<td>96.1 (3.68)</td>
<td>106.8 (4.56)</td>
</tr>
<tr>
<td>0.001</td>
<td>86.1 (3.33)</td>
<td>94.6 (3.52)</td>
<td>86.2 (3.34)</td>
<td>97.8 (4.03)</td>
</tr>
<tr>
<td>0.003</td>
<td>192.0 (4.30)</td>
<td>209.8 (4.51)</td>
<td>192.2 (4.31)</td>
<td>210.4 (4.51)</td>
</tr>
<tr>
<td>0.01</td>
<td>439.2 (5.51)</td>
<td>478.1 (6.09)</td>
<td>439.3 (5.50)</td>
<td>478.3 (6.12)</td>
</tr>
<tr>
<td>0.02</td>
<td>696.1 (6.57)</td>
<td>770.5 (7.25)</td>
<td>696.2 (6.57)</td>
<td>770.9 (7.25)</td>
</tr>
</tbody>
</table>
Figure 1. Top panel: time series of 3000 observations generated from a change-point AR(2) model. Middle panel: true and estimated values of $\sigma_t^2$. Bottom panel: true value of $\mu_t + \alpha_{1,t}Y_{t-1} + \alpha_{2,t}Y_{t-2}$ and its Bayes estimate, which are visually indistinguishable at most places where only the Bayes estimates is shown. Not shown in the bottom panel are the BCMIX and SISR estimates which are visually indistinguishable from the Bayes estimate.
Figure 2. True value of $\sigma_t^2$ and its BCMIX, SISR and SISR* estimates based on 3000 observations.
Figure 3. Weekly returns $Y_t$ and one-week-ahead forecasts of NASDAQ index. Also shown are the 1st and 99th percentiles of $Y_{t+1}$ estimated from $\{Y_i, i \leq t\}$. 

25