<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No:</strong> SS-2020-0439</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
<tr>
<td><strong>Notice:</strong></td>
</tr>
</tbody>
</table>
HYBRID RESAMPLING CONFIDENCE INTERVALS FOR CHANGE-POINT OR STATIONARY HIGH-DIMENSIONAL STOCHASTIC REGRESSION MODELS

Wei Dai and Ka Wai Tsang

Abstract: Herein we show how hybrid resampling, introduced by Chuang and Lai twenty years ago in this journal, can be applied to address (a) the long-standing problem of inference on change-times and changed parameters in change-point ARX-GARCH models, and (b) the challenging problem of valid confidence intervals, after variable selection under sparsity assumptions, for the parameters in linear regression models with high-dimensional stochastic regressors and asymptotically stationary noise. For the latter problem, we introduce consistent estimators of the selected parameters and a resampling approach to overcome the inherent difficulties of post-selection confidence intervals. For the former problem, we use sequential Monte Carlo for the latent states (representing the change-times and changed parameters) of a hidden Markov model. Asymptotic efficiency theory together with simulation and empirical studies are provided to demonstrate and illustrate the advantages of the proposed methods.

Key words and phrases: change-point ARX-GARCH models, coverage probabil-
ity of credible and confidence intervals, double block bootstrap, hidden Markov models and particle filters, sequential Monte Carlo, sparsity and variable selection

1. Introduction and Background

Consider the linear regression model

$$y_t = \beta_0 + \sum_{j=1}^{p_n} \beta_j x_{tj} + \varepsilon_t, \quad t = 1, \ldots, n,$$

(1.1)

with $p_n$ predictor variables $x_{t1}, \ldots, x_{tp_n}$ that are uncorrelated with the mean-zero random disturbances $\varepsilon_t$. By centering $y_t$ and $x_{tj}$ at their respective means, it will be assumed without loss of generality that $\beta_0 = 0$ throughout the sequel. In the big data era, $p_n$ is often larger than $n$, and one relies on sparsity assumptions on the regression coefficients to carry out least squares or penalized least squares in conjunction with greedy variable selection algorithms such as Lasso, $L_2$-boosting, orthogonal matching pursuit (also called orthogonal greedy algorithm or OGA), sure independence screening, high-dimensional information criterion (HDIC), as reviewed by Ing and Lai (2011) who also derive (a) the convergence rate of OGA under weak sparsity assumptions, (b) variable selection consistency of OGA+HDIC under strong sparsity, and (c) the oracle property of
OGA+HDIC+Trimming (i.e., elimination of irrelevant variables) that the
resultant regression estimate is equivalent to least squares regression on an
asymptotically minimal set of relevant regressors.

1.1 Post-selection hypothesis testing and confidence regions

Consider independent $Y_t \sim N(\mu, 1)$, $1 \leq t \leq n$, and the selection
rule that takes the largest observation $Y_{(n)}$ for consideration, where $(n) = \arg \max_{1 \leq t \leq n} Y_t$. Then $Y_{(n)} - \mu$ is no longer $N(0, 1)$ but has the distribu-
tion of the maximum of $n$ independent standard normal variables. Hence
valid post-selection confidence intervals have to incorporate the “selection
effect” on their coverage probability, as noted by Sorić (1989). For the re-
gression parameters in (1.1), Zhang and Zhang (2014) and Belloni et al.
(2015) have shown how to incorporate this selection effect to construct
post-selection confidence regions for the regression parameters selected by
Lasso-type methods; Ing and Lai (2015) (respectively, Lee and Wu 2018) in-
corporate the selection effect to construct post-selection confidence regions
for OGA selection (respectively, via a “bootstrap recipe”).

Lee and Taylor (2014) and Lockhart et al. (2014) have shown that
the conditional distributions for the components of $E\hat{\beta}_j$ given that $\hat{J}$ is
selected by greedy-type algorithms, including OGA and Lasso, can be used
1.2 Hybrid resampling

to construct valid confidence intervals for these components, where \( \beta_j \) is the column vector consisting of \( \beta_j, j \in J \subset \{1, \ldots, p_n\} \), and \( \hat{\beta}_j \) is the corresponding least squares or penalized least squares estimate. Since \( X = (x_{ij})_{1 \leq i \leq n, 1 \leq j \leq p_n} \) is nonrandom,

\[
E \hat{\beta}_j = \beta_j + \sum_{j \in J^c} \beta_j (X_j^T X_j)^{-1} X_j^T X_j, \tag{1.2}
\]

hence \( E \hat{\beta}_j \neq \beta_j \) unless \( \beta_j = 0 \) for \( j \in J^c \) (i.e., all relevant variables are selected) or \( X_j^T X_j = 0 \) (i.e., \( X_j \) is orthogonal to the linear space spanned by the column vectors of \( X_j \)) for \( j \in J^c \). The issue that the valid confidence intervals are actually about the components of \( E \beta_j \) (rather than those of \( \beta_j \)) has been pointed out by Ing et al. (2017) who called it the “spill-over effect” on these post-selection confidence intervals.

1.2 Hybrid resampling

Hybrid resampling, introduced by Chuang and Lai (2000), is a hybrid of exact and bootstrap methods, for constructing confidence regions when the data generating processes are too complex for standard approaches to be applicable. Concerning the exact method, Chuang and Lai (1998) note that even though the standardized sample mean \( \sqrt{n}(\bar{Y}_n - \mu) \) is a pivot in
1.2 Hybrid resampling

the classical example $Y_i \sim N(\mu, 1)$, $\sqrt{T}(\bar{Y}_T - \mu)$ is highly non-pivotal when the fixed sample size $n$ is replaced by a group sequential stopping rule $T$ because the distribution of $T$ depends on $\mu$, hence the need for the exact method introduced by Rosner and Tsiatis (1988), which they generalize as follows. If $\{F_\theta : \theta \in \Theta\}$ is indexed by a real-valued parameter $\theta$, an exact equal-tailed confidence region can always be found by using the well-known duality between hypothesis tests and confidence regions. Suppose one would like to test the null hypothesis that $\theta$ is equal to $\theta_0$. Let $R(Y, \theta_0)$ be some real-valued test statistic based on the data $Y$. Let $u_\alpha(\theta_0)$ be the $\alpha$-quantile of the distribution of $R(Y, \theta_0)$ under the distribution $F_{\theta_0}$. The null hypothesis is accepted if $u_\alpha(\theta_0) < R(Y, \theta_0) < u_{1-\alpha}(\theta_0)$. An exact equal-tailed confidence region with coverage probability $1 - 2\alpha$ consists of all $\theta_0$ not rejected by the test and is therefore given by $\{\theta : u_\alpha(\theta) < R(Y, \theta) < u_{1-\alpha}(\theta)\}$. The exact method, however, applies only when there are no nuisance parameters. The bootstrap method replaces the quantiles $u_\alpha(\theta)$ and $u_{1-\alpha}(\theta)$ by the approximate quantiles $u_\alpha^*$ and $u_{1-\alpha}^*$ obtained in the following manner. Based on $Y$, construct an estimate $\hat{F}$ of $F \in \mathcal{F}$. The quantile $u_\alpha^*$ is defined to be $\alpha$-quantile of the distribution of $R(Y^*, \hat{\theta})$ with $Y^*$ generated from $\hat{F}$ and $\hat{\theta} = \theta(\hat{F})$, yielding the confidence region $\{\theta : u_\alpha^* < R(Y, \theta) < u_{1-\alpha}^*\}$ with approximate coverage probability $1 - 2\alpha$. 
The hybrid resampling confidence region is based on reducing the family of distributions $\mathcal{F}$ to another family of distributions $\{\hat{F}_\theta : \theta \in \Theta\}$, which is used as the “resampling family” and in which $\theta$ is the unknown parameter of interest. Let $\hat{u}_\alpha(\theta)$ be the $\alpha$-quantile of the sampling distribution of $R(Y, \theta)$ under the assumption that $Y$ has distribution $\hat{F}_\theta$. The hybrid confidence region results from applying the exact method to $\{\hat{F}_\theta : \theta \in \Theta\}$ and is given by $\{\theta : \hat{u}_\alpha(\theta) < R(Y, \theta) < \hat{u}_{1-\alpha}(\theta)\}$, the quantiles $\hat{u}_\alpha(\theta)$ and $\hat{u}_{1-\alpha}(\theta)$ can be computed as in the bootstrap method.

### 1.3 Stochastic regressors and martingale regression models

For the case of constant $p_n$ in $[1,1]$ with $\mathcal{F}_{t-1}$-measurable regressor $x_t = (x_{t1}, \ldots, x_{tpn})^T$ and $\mathcal{F}_t$-measurable $\varepsilon_t$ such that $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ (i.e., $\{\varepsilon_t, \mathcal{F}_t, t \geq 1\}$ is a martingale difference sequence), where $\mathcal{F}_t$ is the filtration generated by $\{(x_s, y_s), 1 \leq s \leq t\}$ (i.e., observations up to time $t$), the seminal work of Lai and Wei (1982) uses martingale theory to analyze convergence properties and asymptotic normality of least squares estimates in time series models and dynamic input-output systems. For Lasso regularization, Loh and Wainwright (2012) considered the case of stochastic regressors $x_t$ that are Gaussian VAR(1) processes $x_t = Ax_{t-1} + \xi_t$ with spectral norm $\|A\| < 1$ and $\{\xi_s\}$ is independent of $\{\varepsilon_s\}$, and Basu and Michailidis (2015)
1.3 Stochastic regressors and martingale regression models

considered $\mathcal{F}_{t-1}$-measurable regressors $x_t$ and the $\mathcal{F}_t$-measurable errors $\varepsilon_t$ that are “generated according to independent, centered, Gaussian stationary processes.” They derived asymptotic properties of the Lasso estimate of $\beta$ under usual regularity conditions such as the restricted eigenvalue or restricted strong convexity assumption for theoretical analysis of Lasso.

In their discussion section, Basu and Michailidis (2015, p.1565) note that a serious limitation of their approach is that “popular models exhibiting nonlinear dependencies such as ARCH and GARCH are not covered.” Working with OGA instead of Lasso, we develop in the next two sections valid post-selection confidence regions of selected parameters in stochastic regression models with conditionally heteroscedastic martingale difference random disturbances $\varepsilon_t$, which are called “martingale regression models” by Guo et al. (2017, p.36) whose Sections 2.2 and 2.6 describe the historical background and development of models of “speculative prices” in equity markets. In particular, Section 2 focuses on GARCH(1,1) errors

$$\varepsilon_t = \sigma_t \xi_t, \quad \sigma_t^2 = \omega + a \varepsilon_{t-1}^2 + b \sigma_{t-1}^2$$ (1.3)

and i.i.d. $\xi_t$ having a common standardized Student $t$-distribution with $\nu > 2$ degrees of freedom in the martingale regression model (1.1) with
\( \beta_0 = 0; \) the case \( \nu = \infty \) corresponds to the standard normal distribution (Lai and Xing 2008, p.150).

In Section 2 we develop hybrid resampling methods for inference in martingale regression models, including (a) confidence intervals for selected regression parameters, and (b) credible intervals for the regression and GARCH parameters in an empirical Bayes model of the change-points of the time series. We also give implementation details and asymptotic efficiency results. Section 3 provides simulation and empirical studies to illustrate the advantages of the proposed methods, and gives further discussion and concluding remarks.

2. Hybrid Resampling in Martingale Regression Models

2.1 Valid confidence intervals for selected regression parameters

Letting \( \hat{J} \subset \{1, \ldots, p_n\} \) with cardinality \( m \) be selected by OGA, we can write (1.1) with \( \beta_0 = 0 \) as \( y_t = \sum_{j \in \hat{J}} \beta_j x_{tj} + \sum_{j \in \hat{J}^c} \beta_j x_{tj} + \varepsilon_t, \ t = 1, \ldots, n, \) or \( Y = X_{\hat{J}} \beta_{\hat{J}} + w \) in vector form as in (1.2), where \( w = X_{\hat{J}^c} \beta_{\hat{J}^c} + \varepsilon. \) We begin by describing the exact method in the preceding paragraph when it is applied to construct via resampling a confidence interval for \( \beta_j \) with \( j \in \hat{J}, \) assuming \( \beta_{\hat{J} \setminus \{j\}} \) and the distribution of \( w \) to be known. We compare the test statistic \( R_j(Y, X, \theta) \) with the quantiles \( u_{\alpha}(\theta) \) and
2.1 Valid confidence intervals for selected regression parameters

$u_{1-\alpha}(\theta)$ of $R_j(\mathbf{Y}_j^{(b)}, \mathbf{X}, \theta)$, $b = 1, \ldots, B$, where $\mathbf{Y}_j^{(b)} = \sum_{i \in \hat{J} \setminus \{j\}} \beta_i \mathbf{x}_i + \theta \mathbf{x}_j + \mathbf{w}^{(b)}$. This paves the way for the hybrid resampling method to construct confidence intervals for $\beta_j$, $j \in \hat{J}$. The consistency of $\hat{\beta}_j$, the least squares estimate of $\beta_j$, in martingale regression has recently been established by Lai, Xu and Yuan (2020) who have also extended the result of Ing and Lai (2011, Theorem 3) that the set $\hat{J}$ of regressors selected by OGA+HDIC contains all relevant variables with probability approaching 1 as $n \to \infty$.

This and (1.1) with $\beta_0 = 0$ then imply that with probability approaching 1 as $n \to \infty$, $\hat{w}_t \to \varepsilon_t$, where $\hat{w}_t = y_t - \sum_{j \in \hat{J}} \hat{\beta}_j x_{tj}$ are the residuals and $\hat{\beta}_j (j \in \hat{J})$ are the components of the least squares estimate $\hat{\mathbf{\beta}}_J$. Because $\{\varepsilon_t\}$ is a conditionally heteroscedastic martingale difference sequence (and not an i.i.d. sequence), we apply the double block bootstrap (instead of the usual bootstrap for i.i.d. data), introduced by Lee and Lai (2009), to the residuals $\hat{w}_t$ that approximate the unobservable $\varepsilon_t$. Let $\mathcal{W} = (\hat{w}_1, \ldots, \hat{w}_n)$. Because $\hat{w}_t \to \varepsilon_t$ with probability approaching 1 as the sample size $n$ approaches $\infty$, the block bootstrap is applicable when the $\varepsilon_t$ are asymptotically stationary; see Bühlmann (2002). For a block length $\ell(\leq n)$, define overlapping blocks $B_{i,\ell} = (\hat{w}_i, \ldots, \hat{w}_{i+\ell-1})$, $1 \leq i \leq n - \ell + 1$. A generic block bootstrap series $\mathcal{W}^*$ consists of $m = \lfloor n/\ell \rfloor$ blocks $B_{i,\ell}^*$ sampled with replacement from $\{B_{i,\ell} : i = 1, \ldots, n - \ell + 1\}$ and pasted end-to-end so that “the
2.1 Valid confidence intervals for selected regression parameters

Block-starting points $S_1, \ldots, S_m$ are i.i.d. uniform on the possible starting
locations” (i.e., $1, \ldots, n - \ell + 1$). Under certain regularity conditions, the
choice $\ell \propto n^{1/3}$ yields the smallest coverage error, of order $O(n^{-1/3})$ for the
block bootstrap confidence, As noted by Lee and Lai (2009), that it “fails
to improve upon normal approximation impeded the widespread use” of
the block bootstraps. They proposed the following double block bootstrap
procedure to reduce the coverage error by an order of magnitude to achieve
marked improvement over the normal approximation.

The first level of double bootstrap gives the above block bootstrap series
$W^*$. For $i = 1, \ldots, m$, divide block $B_{i,\ell}^* \in W^*$ into $\ell - \tilde{\ell} + 1$ subseries $B_{i,j,\tilde{\ell}}^*$
each of length $\tilde{\ell}$. The second level of double block bootstrap samples $\tilde{m} =
\lfloor n/\tilde{\ell} \rfloor$ blocks with replacement from $\{B_{i,j,\tilde{\ell}}^*: i = 1, \ldots, m; j = 1, \ldots, \ell - \tilde{\ell} + 1\}$, yielding the block bootstrap series $W^{**}$ that consists of $\tilde{m}$ blocks $B_{i,j,\tilde{\ell}}^{**}$
pasted end-to-end. The procedure is summarized in Algorithm 1, which we
now use to construct confidence intervals for $\beta_j, j \in \hat{J}$, in the martingale
regression model. Here we use the studentized correction of coverage error
as in Lee and Lai (2009, Sections 2.2 and 3) and Chuang and Lai (2000,
Section 3.2); specifically, we consider the studentized statistic

$$R_j(Y, X, \theta) = (\hat{\beta}_j - \theta)/(s\sqrt{c_{jj}})$$

(2.1)
2.1 Valid confidence intervals for selected regression parameters

**Algorithm 1 Double block bootstrap**

**INPUT:** \( \mathcal{W} = (\hat{w}_1, \ldots, \hat{w}_n) \)

**Step 1:** Resample first-level block bootstrap series \( \mathcal{W}^* \) consisting of \( m = \lfloor n/\ell \rfloor \) blocks \( B^*_{i,\ell} \) pasted end-to-end.

**Step 2:** Resample second-level block bootstrap series \( \mathcal{W}^{**} \)

1. Divide \( B^*_{i,\ell} \) into \( \ell - \tilde{\ell} + 1 \) sub-blocks, each of length \( \tilde{\ell} \).

2. Sample \( \tilde{m} = \lfloor n/\tilde{\ell} \rfloor \) blocks \( B^{**}_{i,j,\ell} \) with replacement from these sub-blocks and paste them end-to-end to form \( \mathcal{W}^{**} \).

**OUTPUT:** \( \mathcal{W}^{**} \)

for testing \( \beta_j = \theta \), where \( \hat{\beta}_j \) is the \( j \)th component of the least squares estimate \( (X_j^T X_j)^{-1} X_j Y \) of \( \beta_j \), \( c_{jj} \) is the \( j \)th diagonal element of \( (X_j^T X_j)^{-1} \), 

\[
 s^2 = \frac{\sum_{t=1}^n w_{t}^2}{(n - m_n)} \quad \text{and} \quad m_n = O(\sqrt{n/\log p_n}) \quad \text{is the cardinality of the set} \ J \text{ of regressors selected by OGA+HDIC. The quantiles} \ \hat{u}_\alpha(\theta) \text{ and} \ \hat{u}_{1-\alpha}(\theta) \text{ of} \ R_j(Y, X, \theta) \text{ are evaluated by Algorithm 1. Although the (1-2\alpha)-level confidence set} \ \{ \theta : \hat{u}_\alpha(\theta) < R_j(Y, X, \theta) < \hat{u}_{1-\alpha}(\theta) \} \text{ may not be an interval, it often suffices to give the upper limit} \ \theta^U_j = \arg \min_{\theta} |R_j(Y, X, \theta) - \hat{u}_\alpha(\theta)| \text{ and the lower limit} \ \theta^L_j = \arg \min_{\theta} |R_j(Y, X, \theta) - \hat{u}_{1-\alpha}(\theta)| \text{ as the confidence interval} \ [\theta^L_j, \theta^U_j] \text{ of} \ \beta_j \text{; see Chuang and Lai (2000, p.4). We choose} \ m = m_n \text{ in the double block bootstrap (Algorithm 1), and hence the block length} \ \ell, \text{ by OGA+HDBIC as in Section 4 of Ing and Lai (2011, p.1484):} \]

\[
 m_n = \arg \min_{1 \leq k \leq 2\lfloor n/\log p_n \rfloor} \left\{ n \log \hat{\sigma}^2_{\{j_1, \ldots, j_k\}} + k(\log n)(\log p_n) \right\}, \quad (2.2)
\]
where $\hat{j}_1, \ldots, \hat{j}_k$ are selected along the OGA path until the $k$th iteration and $\hat{\sigma}_J^2 = n^{-1} \sum_{t=1}^{n} (y_t - \hat{y}_{t,J})^2$ for $J \subset \{1, \ldots, p_n\}$, in which $\hat{y}_{t,J}$ denotes the fitted value of $y_t$ when $Y$ is projected into the linear space spanned by the column vectors of $X_J$; see Ing and Lai (2011, p.1479). In Supplementary Material S1 we prove the following theorem showing that the post-selection confidence interval for $\beta_{j,j} \in \hat{J}$, constructed by this hybrid resampling approach has coverage error $o(n^{-1/2})$ under certain regularity conditions if $\ell$ is chosen suitably.

**Theorem 1.** If $\ell \propto n^{1/3}$, then the preceding confidence interval for $\beta_{j,j} \in \hat{J}$, in the martingale regression model has coverage error $O(n^{-2/3})$.

### 2.2 Change-point ARX-GARCH models and empirical Bayes

GARCH models are widely used in econometric time series for dynamic modeling of volatilities; see Chapter 6 of Lai and Xing (2008) whose Sections 6.3.2 and 9.5 describe the commonly observed pattern of “volatility persistence” and its cause due to ignoring structural changes in the GARCH model. Their Section 9.5.2 also describes change-point AR-GARCH models, which have been definitively generalized by Lai and Xing (2013) who
2.2 Change-point ARX-GARCH models and empirical Bayes

use a Bayesian change-point ARX-GARCH model of the form

\[ y_t = \beta_t^T x_t + \nu_t \sqrt{h_t} \epsilon_t, \quad (2.3) \]

in which \( \epsilon_t \) are i.i.d. having the standardized Student \( t \)-distribution with \( \kappa > 2 \) degrees of freedom and are independent of \( x_t \), the parameter vector \( \beta_t \) and the unconditional variance \( \nu_t^2 \) are piecewise constant, with jumps at times of structural change, the vector \( x_t \) consists of exogenous variables and past observations \( y_{t-1}, y_{t-2}, \ldots, y_{t-k} \) and \( h_t \) is used to represent short-term proportional fluctuations in variance generated by the GARCH model

\[ h_t = (1 - \sum_{i=1}^{k} a_i - \sum_{l=1}^{k'} b_l) + \sum_{i=1}^{k} a_i w_{t-i}^2 + \sum_{l=1}^{k'} b_l h_{t-l}, \text{ with } w_s = \sqrt{h_s} \epsilon_s. \quad (2.4) \]

where the time-invariant GARCH parameters \( a_1, \ldots, a_k, b_1, \ldots, b_{k'} \) are assumed to satisfy \( a_i \geq 0, b_l \geq 0 \) and \( \sum_{i=1}^{k} a_i + \sum_{l=1}^{k'} b_l \leq 1 \). Letting \( \tau_t = 1/(2\nu_t^2) \), they assume \( \theta_t = (\beta_t^T, \tau_t)^T \) to be piecewise constant and satisfy the following conditions:

(A1) Letting \( t_0 = \max(k, k') \), \( I_{t_0} = 1 \) and there is no change-point prior to \( t_0 \). For \( t > t_0 \), the change-times of \( \theta_t \) form a renewal process with i.i.d. inter-arrival times that are geometrically distributed with parameter
2.2 Change-point ARX-GARCH models and empirical Bayes

\[ p \text{ or, equivalently,} \]
\[ I_t := 1_{\theta_t \neq \theta_{t-1}} \text{ are i.i.d. Bernoulli}(p) \text{ with } P(I_t = 1) = p. \] (2.5)

(A2) \( \theta_t = (1 - I_t)\theta_{t-1} + I_t(z_t^T, \gamma_t)^T, \) where \( (z_1^T, \gamma_1)^T, (z_2^T, \gamma_2)^T, \ldots \) are i.i.d. random vectors such that \( z_t \mid \gamma_t \sim \text{Normal}(z, V/(2\gamma_t)), \gamma_t \sim \chi^2_d/\rho, \) with \( \chi^2_d \) the chi-square distribution with \( d \) degrees of freedom.

(A3) The process \( \{I_t\}, \{(z_t^T, \gamma_t)\}, \) and \( \{(x_t, \epsilon_t)\} \) are independent.

We now focus on the case \( k = k' = 1 \) in (2.4), i.e., the GARCH(1,1) model with nonnegative parameters \( a \) and \( b \) such that \( a + b \leq 1, \) and summarize Lai and Xing’s empirical Bayes (EB) approach to determining the hyperparameters \( \eta = (a, b, \kappa, p, z, V, \rho) \) and \( d \) in (2.5) and (A2) that define the change-point ARX-GARCH(1,1) model. Section 2.2 of Lai and Xing(2013) provides details of the EB approach. First, the \( \chi^2_d/\rho \) prior distribution for \( 1/(2\nu^2_t) \) given by (A2) is written as Gamma\( (d/2, \rho/2) \) so that \( d \) does not need to be an integer. Second, \( z, V, \rho \) and \( d \) are estimated by applying the following method of moments to the “stationary distribution of the Markov chain \( (I_t, \theta_t, \epsilon_t) \) that is partially observed via \( (x_t, y_t), 1 \leq t \leq n. \)” From (A2) and (A3), it follows that \( E(\beta_t) = z, \text{Cov}(\beta_t) = (E\nu^2_t)V, \) and \( E(x_t y_t) = E(x_t x^T_t)z. \) From \( n - L \) moving windows \( \{(x_t, y_t) : s \leq t \leq \)}
2.2 Change-point ARX-GARCH models and empirical Bayes

of these data, compute the least squares estimate and denote it as \( \tilde{\beta}^{(s)} \). Each \( \tilde{\beta}^{(s)} \) is a method-of-moments estimate of \( z \), and so is \( \overline{\beta} = (n - L)^{-1}\sum_{s=1}^{n-L} \tilde{\beta}^{(s)} \). If an oracle would reveal the change-times up to time \( n \), then one would segment the time series accordingly and use the least squares estimate of each segment to estimate the regression parameter for that segment. This suggests using (a) moving windows of length \( L + 1 \) to replace the segments and (b) the least squares estimate \( \tilde{\beta}^{(s)} \) based on the moving window \( \{(x_t, y_t) : s \leq t \leq s + L\} \). The average \( \hat{r}_s := (L + 1)^{-1}\sum_{t=s}^{s+L} (e_t - \bar{e}_s)^2 \) of the squared (centered) residuals in the moving window, where \( e_t = y_t - x_t^T \tilde{\beta}^{(s)} \) and \( \bar{e}_s = (L + 1)^{-1}\sum_{t=s}^{s+L} e_t \), is a method-of-moments estimate of \( E(\nu_t^2) = \rho/[2(d - 2)] \) for this time segment, hence \( \rho/[2(d - 2)] \) can be estimated by \( \overline{r} := (n - L)^{-1}\sum_{s=1}^{n-L} \hat{r}_s \). In this connection, recall that \( \theta_t^{-1} = (\beta_t^T, \tau_t) \) and \( 2\nu_t^2 = 1/\tau_t \) and note that by (A2), \( \beta_t | \nu_t^2 \sim N(z, 2\nu_t^2 V) \) and \( 2\nu_t^2 \) has the inverse gamma distribution with shape parameter \( d/2 \) and scale parameter \( \rho/2 \), the mean (respectively, variance) of which is \( \rho/(d - 2) \) if \( d > 2 \) (respectively, \( 2\rho^2/[(d - 2)(d - 4)] \) if \( d > 4 \)). The variance \( E[(\nu_t^2 - E\nu_t^2)^2] \) can be estimated from the centered residuals in moving window \( s \) by \( \hat{v}_s := L^{-1}\sum_{t=s}^{s+L} (e_t^2 - \hat{r}_s)^2 \), hence \( \rho^2/2[(d - 2)(d - 4)] \) can be estimated by the average \( \overline{v} := (n - L)^{-1}\sum_{s=1}^{n-L} \hat{v}_s \) over the moving windows.

Since \( V = \text{Cov}(\beta_t) / E\nu_t^2 \), an obvious method-of-moments estimate of \( V \)
2.3 Sequential Monte Carlo to sample the latent variables $I_t, \theta_t$

Chen and Lai (2007) and Chan and Lai (2013) have shown how particle filters that use sequential Monte Carlo (SMC) involving importance sampling and resampling can be used to sample the latent variables $I_t$ and the changed regression parameters or means in the change-point ARX model and mean-shift model, respectively, without GARCH dynamics for $\epsilon_t$. We next extend the SMC procedure in Section II of Chen and Lai (2007, pp. 67-68) to the change-point ARX-GARCH(1,1) model, first assuming that the hyperparameters $z, V, \rho, d, p, \eta$ are known.

Let $C_t = \max\{j \leq t : I_j = 1\}$ be the most recent change-time up to time $t$ ($\max \emptyset = t_0$), which plays a key role in Chen and Lai’s SMC procedure for the change-point ARX model with $h_t = 1$ in (2.3), for which they also
2.3 Sequential Monte Carlo to sample the latent variables \( I_t, \Theta_t \)

assume the \( \epsilon_t \) to be standard normal. This distributional assumption is overly restrictive for GARCH models, for which the standardized Student \( t \)-distribution with unspecified degrees of freedom \( \kappa > 2 \) is often assumed in likelihood inference; see Lai and Xing (2008, p.150). In view of the normal-inverse gamma prior distribution in (A2) for \( (\beta_s^T, 2\nu_s^2) \) that does not change for \( C_t \leq s \leq t \), the posterior distribution of \( \theta_t \) (with components \( \beta_t \) and \( \tau_t = (2\nu_t^2)^{-1} \)) given \( C_t \) and \( \{(x_s, y_s) : C_t \leq s \leq t\} \) also belongs to this conjugate family with \( \beta_t | \nu_{C_t}^2 \sim N(\mu_{C_t,t}, \nu_{C_t}^2 V_{C_t,t}) \) and \( \tau_t \sim \text{Gamma}(d/2 + (t - C_t)/2, \rho_t/2) \), where

\[
\begin{align*}
\mu_{s,t} &= V_{s,t}(V^{-1}z + \sum_{i=s}^{t} y_i x_i), \\
V_{s,t} &= (V^{-1} + \sum_{i=s}^{t} x_i x_i^T)^{-1}, \\
\rho_t &= \rho + z^T V^{-1} z + \sum_{s=C_t}^{t} y_s^2 - \mu_{C_t,t}^T V_{C_t,t}^{-1} \mu_{C_t,t}, \\
\end{align*}
\]

(2.6)

assuming the hyperparameters of the Bayesian model and \( C_t \) to be specified; see Box and Tiao (1973, Chapter 8). Without prespecifying the hyperparameters \( z, V, \rho, d \), we can replace them in (2.6) by their method-of-moments estimates described in Section 2.2, which also gives the estimates \( \tilde{\rho}, \tilde{\eta} \) of the hyperparameters \( p \) and \( \eta \) in (A1) and (2.4). Replacing \( \beta_t, \nu_t \) and \( h_t \) in \( \epsilon_t = (y_t - \beta_t^T x_t)/(\nu_t \sqrt{h_t}) \) (from (2.3)) by their estimates involves \( C_t \) that we now proceed to construct by using SMC which comprises sequential
2.3 Sequential Monte Carlo to sample the latent variables $I_t, \theta$,

importance sampling and resampling, as described by Chen and Lai (2007, Section II) and Chan and Lai (2013, Section 2.1).

To explain SMC, we first assume that the hyperparameters $z, V, \rho, d, p, \eta$ are known. As in Chen and Lai (2007, p.68), let $\pi(\cdot|\cdot)$ and $\pi(\cdot)$ denote the conditional and the joint density functions of the random variables indicated. Sequential importance sampling samples $I_{t_0} = 1, I_{t_0+1}, \ldots, I_n$ sequentially from the proposal distribution for which $I_t | (I_{t_0}, \ldots, I_{t-1})$ is Bernoulli, assuming the values 1 and 0 with respective probabilities that are proportional to $a_t(p, \eta) : b_t(p, \eta)$, where

$$
a_t(p, \eta) = \frac{p}{(x_t^T V x_t + \nu_t^2 h_t)^{1/2} \pi \left( \frac{y_t}{(x_t^T V x_t + \nu_t^2 h_t)^{1/2}} \right)},
$$

$$
b_t(p, \eta) = \frac{1 - p}{(x_t^T V C_{t,t} x_t + \nu_t^2 h_t)^{1/2} \pi \left( \frac{y_t - \mu_{C_{t,t}}^T x_t}{(x_t^T V C_{t,t} x_t + \nu_t^2 h_t)^{1/2}} \right)},
$$

(2.7)

in which $\pi$ is the standardized Student $t$-density (with $\kappa$ degrees of freedom) for $\epsilon_t$ in (2.3) and $h_t$ is defined recursively by (2.4) with $k = k' = 1$ and the GARCH parameter vector $\eta = (a, b, \kappa)$. Hence the argument in Section IIA of Chen and Lai (2007) can be extended to the change-point ARX-GARCH(1,1) model, yielding the recursive formula for the importance
2.3 Sequential Monte Carlo to sample the latent variables $I_t, \theta_t$.

weights

$$w_t = w_{t-1}\{a_t(p, \eta) + b_t(p, \eta)\}, \quad t > t_0; \quad w_{t_0} = 1,$$

(2.8)

for the posterior distribution of $I_{t_0:t} := (I_{t_0}, \ldots, I_t)$. For Monte Carlo implementation, $B$ sample paths are generated, with importance weights $w_t^{(b)} (b = 1, \ldots, B)$, so that the Monte Carlo estimate of the posterior mean $E(\theta_t | (x_s, y_s) : s \leq t)$ is given by

$$\beta_{t, \text{Bayes}} = \frac{\sum_{b=1}^{B} w_t^{(b)} \mu_{C_t^{(b)}, t}}{\sum_{b=1}^{B} w_t^{(b)}}, \quad 2\nu_{t, \text{Bayes}}^2 = \frac{\sum_{b=1}^{B} \rho_t}{(d + t - C_t)} \frac{w_t^{(b)}}{\sum_{b=1}^{B} w_t^{(b)}}$$

(2.9)

in which $C_t^{(1)}, \ldots, C_t^{(B)}$ are $B$ independent replicates of $C_t$ and the Bayes model has hyperparameters, which we estimate sequentially as described in the next paragraph.

To initialize, we use moving windows of length $L + 1$ and the method-of-moment estimates of $z, V, \rho$ and $d$, followed by maximum likelihood estimates of $\eta$ and $p$, as in the second paragraph of Section 2.2. These estimates are used in (2.7) and (2.8) that define the importance weights in the SMC procedure to generate $I_{t_0:t}$. This represents the initialization $\tilde{I}_{t_0:n}$ of the SMC procedure to generate the successive change-times $t_0, t_1, \ldots (< n)$.
2.3 Sequential Monte Carlo to sample the latent variables $I_t, \theta_t$ under the posterior distribution given the observations $(x_t, y_t), 1 \leq t \leq n$. After the change-times are generated, the moving windows of length $L + 1$ in the second paragraph of Section 2.2 can be replaced by the times $t \in \{t_i, t_i + 1, \ldots, t_{i+1} - 1\}$ between consecutive change-points so that the hyperparameters are re-estimated using these time segments during which $\theta_t$ stays constant. These revised hyperparameter estimates are denoted by $\hat{z}, \hat{V}, \hat{\rho}, \hat{d}, \hat{\eta}$ and $\hat{p}$ in the SMC procedure summarized in Algorithm 2, which consists of sequential importance sampling and bootstrap resampling as in Chan and Lai (2013, Section 2.1), and which describes Monte Carlo simulations of $B$ independent sample paths $I_{t_0:n}^{(b)}$ from the estimated posterior distribution given $\{(x_t, y_t) : 1 \leq t \leq n\}$ with importance weights that use $\hat{z}, \hat{V}, \hat{\rho}, \hat{d}, \hat{\eta}$ and $\hat{p}$ to substitute the unspecified hyperparameters in the Bayesian model.

As explained in Section IIB of Chen and Lai (2007), the importance weights (2.8) have difficulties for large $t$ when they tend to have large variance, as reflected by the normalized weights $w_t^{(i)}/\sum_{b=1}^{B} w_t^{(b)}$ mostly converging to 0 while they have to sum to 1. To address this difficulty, bootstrap resampling has been used to resample $\{I_{t_0:t}^{(1)}, \ldots, I_{t_0:t}^{(B)}\}$ with probabilities proportional to the importance weights. Thus, at every $t$, SMC consists of an importance sampling step followed by a resampling that transforms
2.3 Sequential Monte Carlo to sample the latent variables $I_t, \theta$.

The weighted sample to a bootstrap sample with equal weights. Section 2.1 of Chan and Lai (2013) provides an asymptotic theory of SMC with bootstrap resampling at every stage $t$, which is extended in their Section 2.2 to occasional resampling when the coefficient of variation of the normalized weights exceeds some threshold, as has been used in Section IIB of Chen and Lai (2007). Algorithm 2 uses bootstrap resampling at every stage because the goal is to generate a sequence of consecutive change-times which are used to segment the data so that the hyperparameters can be re-estimated using these time segments as mentioned in the preceding paragraph. Specifically, recall that in the second paragraph of Section 2.2 we have used $n - L$ moving windows of length $L + 1$ to replace the time segments between two consecutive change-times “revealed by an oracle” to estimate the hyperparameters by the method of moments or maximum likelihood. In lieu of the oracle, Algorithm 2 simulates $B$ samples of $I_{t_0:n}$ from its posterior distribution given the observations. From each simulated sample $I_{t_0:n}^{(b)}$ we have the time segments $\{t_i^{(b)}, t_i^{(b)} + 1, \ldots, t_{i+1}^{(b)} - 1\}$ between consecutive change-times $t_i^{(b)}$ and $t_{i+1}^{(b)}$. Hence we can apply the method of moments (respectively, maximum likelihood) to each time segment and aggregate over these time segments to obtain the revised hyperparameter estimates $\hat{z}, \hat{V}, \hat{\rho}$ and $\hat{d}$ (respectively, $\hat{\eta}$ and $\hat{p}$).
2.3 Sequential Monte Carlo to sample the latent variables $I_t, \theta_{22}$

Algorithm 2: SMC to simulate $B$ samples $I_{t_0:n}^{(b)}$ from posterior distribution

**INPUT:** Observations $(x_t, y_t), 1 \leq t \leq n$

For $t = 2, \ldots, n$, initialize $I_{t_0} = 1, w_{t_0} = 1$ with $t_0 = 1$

**Importance Sampling:** Generate for $b = 1, \ldots, B$, Bernoulli $\tilde{I}_t^{(b)}|I_{t_0:t-1}$ taking the values 1 and 0 with respective probabilities that are proportional to $a_t(\tilde{p}, \tilde{\eta}) : b_t(\tilde{p}, \tilde{\eta})$ defined by (2.7) with the hyperparameters replaced by their estimates described in Section 2.2. Let $\tilde{I}_{t_0:t} = (\tilde{I}_t^{(b)}), \tilde{I}_t^{(b)}$, which is assigned the importance weight $\tilde{w}_t^{(b)} = w_{t-1}^{(b)}\{a_t(\tilde{p}, \tilde{\eta}) + b_t(\tilde{p}, \tilde{\eta})\}$.

**Bootstrap Resampling:** Resample $\tilde{I}_{t_0:t}^{(1)}, \ldots, \tilde{I}_{t_0:t}^{(B)}$ with probabilities proportional to the importance weights, yielding the bootstrap sample $I_{t_0:t}^{(1)}, \ldots, I_{t_0:t}^{(B)}$ with equal weights $w_t^{(1)} = \ldots = w_t^{(B)} = 1$.

**OUTPUT:** $I_{t_0:n}^{(b)}, b = 1, \ldots, B$, and $\hat{z}, \hat{V}, \hat{\rho}, \hat{d}, \hat{\eta}, \hat{p}$

Algorithm 2 plays a basic role in constructing hybrid resampling credible intervals for the regression parameters $\beta_t$ and the GARCH parameters $\nu_t^2$ in the change-point ARX-GARCH(1,1) model (2.3), similar to that played by Algorithm 1 (double block bootstrap) in hybrid confidence intervals for selected regression parameters in Section 2.1. For $\nu_t^2$ and scalar functions of $\beta_t$, we construct empirical Bayes(EB) credible intervals for their values over a designated period which may contain 0, 1 or more change-points in the Bayesian change-point ARX-GARCH(1,1) model. Details are given in the next subsection, which also establishes the asymptotic efficiency of the EB approach.
2.4  Frequentist segmentation, asymptotic efficiency of EB

Lai and Xing (2013, p.1575) point out that in contrast to the EB approach which assumes a “relatively simple stochastic model” for change-points, “the frequentist approach, often called segmentation, assumes the change-points in the pre- and post-change regression coefficients in regression models to be unknown parameters, and uses maximum likelihood to estimate them and a model selection criterion to determine the number of change-points.” Earlier Lai and Xing (2011, pp.539-540) have provided a survey of the frequentist approach, including in particular the works of Bai and Perron (1998, 2003) and Olshen et al. (2004) on segmentation and Yao (1988), Birgé and Massart (2001), Zhang and Siegmund (2006) and Davis et al. (2006) on model selection criteria. Their Section 4 have shown how “the relative simplicity of the posterior distribution (of the change-points in the EB approach) opens up new possibilities in resolving the long-standing difficulties in the frequentist problem of testing for change-points and determining the segmentation.” Lai and Xing (2013, pp.1584-1587) extend this work to the change-point ARX-GARCH model that we have described in Section 2.2, establishing the consistency of the EB estimates of the number of change-points and the post-change parameters in the frequentist segmentation model, besides the hyperparameter vector $\eta$ in the Bayesian model.
As noted in the last paragraph of Section 2.3, the hybrid resampling method is used to construct credible intervals of piecewise constant \( \nu_t \) and \( \beta_{t,i} \) in the ARX-GARCH(1,1) model. The following theorem, which is proved in Supplementary Material S2 together with additional references to the underlying developments of the EB approach to change-point stochastic regression models, establishes the asymptotic efficiency of the method.

**Theorem 2.** *Under a limiting Poisson assumption given in S2, the EB approach using the hybrid resampling method attains the asymptotic efficiency of the oracle procedure with knowledge of the successive change-times and the hyperparameters of the Bayesian change-point ARX-GARCH(1,1) model.*

3. Simulation and Empirical Studies, Concluding Remarks

We begin with a simulation study, in Section 3.1, of the coverage errors of the selected confidence intervals \([\theta_j^L, \theta_j^U] \), \( j \in \hat{J} \), in the regression model (1.1) with \( \beta_0 = 0 \) and GARCH(1,1) errors \( \varepsilon_t \). Section 3.2 presents an empirical study to illustrate the usefulness of the change-point model and the associated EB-SMC method, and gives some concluding remarks.
3.1 Coverage errors of post-selection confidence intervals

We now study the performance of \([\theta_{L_j}^j, \theta_{U_j}^j], j \in \hat{J}\), by simulating the martingale regression model (1.1) with \(\beta_0 = 0\) and GARCH(1,1) errors \(\varepsilon_t\); see (1.3). We consider \(n = 400, p_n = 500\), and regressors \(x_{tj}\) generated by i.i.d. standard normal \(z_{tj}\) and \(\tilde{z}_t\) via \(x_{tj} = f_t + z_{tj}\) and \(f_t = 0.9 f_{t-1} + \tilde{z}_t(f_0 = 0)\), and \(\alpha = 0.1\) for the nominal confidence level \(1 - 2\alpha\) (hence coverage error \(2\alpha\)). Table 1, given in Supplementary Material S1, presents the results, each of which is based on 500 simulations with \(B = 900\), of the actual coverage errors of the hybrid resampling confidence intervals \([\theta_{L_j}^j, \theta_{U_j}^j]\), for five parameter settings of the martingale regression model. In the last setting, the GARCH parameters \(a = 0 = b\) and \(\nu = \infty\), hence \(\varepsilon_t \sim 0.1 N(0, 1)\) and the conditional approach of Lockhart et al. (2014) is applicable. As shown in the last row of Table 1, the coverage errors of the confidence intervals attained by this conditional approach differ substantially from the nominal value of 20% even through it is applicable in this setting. This is due to the spill-over effect mentioned in the last paragraph of Section 1.1. In contrast, the hybrid resampling confidence intervals \([\theta_{L_j}^j, \theta_{U_j}^j], j \in \hat{J}\), have coverage errors close to 20%. Note that the double block bootstrap in Algorithm 1 is nonparametric and only assumes that \(\varepsilon_t\) is asymptotically stationary. On the other hand, if it is assumed that \(\varepsilon_t\) follows a GARCH(1,1) process as in
our martingale regression model, then hybrid resampling uses the parametric bootstrap to generate confidence intervals for $\beta_j, j \in \hat{J}$. This is similar to hybrid resampling confidence intervals for the correlation coefficient of a bivariate distribution in Example 4 (parametric case involving bivariate normal or regression model with double exponential errors) and Example 6 (nonparametric case) of Chuang and Lai (2000, pp13-15, 26-28). We discuss this in greater detail for the martingale regression model with GARCH(1,1) errors (1.3) in the next subsection.

3.2 Empirical study of EB change-point model and discussion

To illustrate the performance of the proposed EB change-point ARX-GARCH(1,1) modeling approach on real-world data in financial markets that have experienced changes during the period from Jan 3, 2001 to Dec 31, 2008, we use weekly log returns of Wells Fargo & Company (WFC), together with that of the market portfolio (represented by the S&P500 Index) in the Capital Asset Pricing Model (CAPM); see Guo et al. (2017, Sections 2.3 and 2.4.1). The dataset consists of $n = 416$ closing prices $P_t$ for the stock on the first day of the week from which the log returns $y_t = \log(P_t/P_{t-1})$ are computed. Let $x_t$ be the corresponding log returns of the market portfolio. Figure 1 in the Supplementary Material S2 plots
the time series of $x_t, y_t$. For comparison with the EB change-point ARX-GARCH(1,1) modeling approach, we use `garch` in MATLAB to fit to these data the ARX-GARCH(1,1) model with time-invariant parameters:

$$y_t = \beta_1 x_{t-1} + \beta_2 y_{t-1} + w_t, \quad w_t = \sigma_t \epsilon_t, \quad \sigma_t = \omega + aw_{t-1} + b\sigma_{t-1}. \quad (3.1)$$

The results, which are given in the first row of Table 2 in S2, show high volatility persistence. The other rows of Table 2 give corresponding results for the piecewise constant parameters (between the successive change-times identified) by applying the method in Section 2.3 to fit the change-point ARX-GARCH(1,1) model

$$y_t = \beta_{1,1} x_{t-1} + \beta_{2,1} y_{t-1} + \nu_t w_t, \quad w_t = \sqrt{h_t} \epsilon_t, \quad h_t = 1 - a - b + aw_{t-1}^2 + bh_{t-1}. \quad (3.2)$$

to these data. The four identified change-times correspond to Dec 26, 2002, Apr 26, 2007, Dec 22, 2007 and Jul 5, 2008, which are associated with tumultuous economic events in the U.S. economy: the aftermath of the 9/11/2001 terrorist attacks and the preparation for the Iraq War following congressional authorization of President Bush to launch a military attack against Iraq, concern for sustainability of upward earnings trend of DJIA
3.2 Empirical study of EB change-point model and discussion

(Dow Jones Industrial Average) which actually switched into the negative regime on 4/30/2007 after 2:30 pm, and the subprime mortgage meltdown and the collapse of Bear Stearns and Lehman-Brothers during the period June 2007 to July 2008. This empirical study demonstrates the effectiveness of the SMC procedure in identifying the unspecified change-times in the EB change-point ARX-GARCH(1,1) model (3.2) which is also shown to be amenable to valid inference on the piecewise constant regression coefficients and volatility parameters.

In conclusion, we have enhanced the hybrid resampling methodology introduced and developed by Professor Lai in collaboration with different generations of his Ph.D. students during the past twenty years and broadened it to tackle two challenging problems in inference on martingale regression models — another innovation by Professor Lai and his collaborators. The effectiveness of our enhancement in constructing valid confidence intervals for selected regression parameters in high-dimensional martingale regression models after model selection under sparsity assumptions, and empirical Bayes estimation of the change-times together with credible intervals for the piecewise constant regression coefficients and volatility parameters in change-point martingale regression models, opens up new possibilities and applications to time series analysis.
Acknowledgement This research is supported by the National Natural Science Foundation of China under Contract No. 12001461.

References


REFERENCES

_Sinica_ 21, 539-569.


Rosner, G.L. and Tsiatis, A.A. (1988). Exact confidence intervals following a group sequential


Wei Dai

E-mail: 216019004@link.cuhk.edu.cn

Ka Wai Tsang

E-mail: kwtsang@cuhk.edu.cn

*School of Data Science and Center for Statistical Science*

*The Chinese University of Hong Kong, Shenzhen*