Statistica Sinica

HYBRID RESAMPLING CONFIDENCE INTERVALS FOR CHANGE-POINT OR STATIONARY HIGH-DIMENSIONAL STOCHASTIC REGRESSION MODELS

Wei Dai and Ka Wai Tsang

The Chinese University of Hong Kong, Shenzhen

Abstract: Herein, we use hybrid resampling 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 a sequential Monte Carlo for the latent states (respresenting the change times and changed parameters) of a hidden Markov model. Asymptotic efficiency theory and simulation and empirical studies demonstrate the advantages of the proposed methods.

Key words and phrases: change-point ARX-GARCH models, coverage probability 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

$$y_t = \beta_0 + \sum_{j=1}^{p_n} \beta_j x_{tj} + \varepsilon_t, \quad t = 1, \ldots, n, \tag{1.1}$$

with p_n predictor variables $x_{t1}, \ldots, x_{t,p_n}$ that are uncorrelated with the mean-zero random disturbances ε_t. By centering y_t and x_{tj} at their respective means, we assume, without loss of generality, that $\beta_0 = 0$. In the case of big data, p_n is often larger than n. As a result, we need to rely on sparsity assumptions on the regression coefficients to carry out least squares or penalized least squares regressions, together with greedy variable selection algorithms, such as the Lasso, L_2-boosting, orthogonal matching pursuit (also called orthogonal greedy algorithm, or OGA), sure independence screening, and high-dimensional information criterion (HDIC), as reviewed by Ing and Lai (2011). The latter study also derives (a) the convergence rate of OGA under weak sparsity assumptions, (b)the variable selection consistency of OGA+HDIC under strong sparsity, and (c) the oracle property of OGA+HDIC+Trimming (i.e., the elimination of irrelevant variables) such that the resultant regression estimate is equivalent to a least squares regression on an asymptotically minimal set of relevant regressors.

1.1 Post-selection hypothesis testing and confidence regions

Consider independent $Y_t \sim N(\mu, 1)$, for $1 \leq t \leq n$, and a selection rule that considers the largest observation $Y_{(n)}$, where $(n) = \arg\max_{1 \leq t \leq n} Y_t$. Then, $Y_{(n)} - \mu$ is no longer $N(0, 1)$, but has the distribution 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 regression 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. Furthermore, Ing and Lai (2015) incorporate the selection effect to construct post-selection confidence regions for OGA selection, whereas Lee and Wu (2018) do so using a "bootstrap recipe".

Lee and Taylor (2014) and Lockhart et al. (2014) have shown that the conditional distributions for the components of $E\hat{\boldsymbol{\beta}}_{\hat{J}}$, given that $\hat{J}$ is selected using greedy-type algorithms, such as OGA and the Lasso, can be used to construct valid confidence intervals for these components, where $\boldsymbol{\beta}_J$ is the column vector consisting of β_j, for $j \in J \subset \{1, \ldots, p_n\}$, and $\hat{\boldsymbol{\beta}}_J$ is the corresponding least squares or penalized least squares estimate. Because $\mathbf{X} = (x_{tj})_{1 \le t \le n, 1 \le j \le p_n}$ is nonrandom,

$$E\hat{\boldsymbol{\beta}}_{\hat{J}} = \boldsymbol{\beta}_{\hat{J}} + \sum_{j \in \hat{J}^c} \beta_j (\mathbf{X}_{\hat{J}}^T \mathbf{X}_{\hat{J}})^{-1} \mathbf{X}_{\hat{J}}^T \mathbf{X}_j; \tag{1.2}$$

hence, $E\hat{\boldsymbol{\beta}}_{\hat{J}} \neq \boldsymbol{\beta}_{\hat{J}}$ unless $\beta_j = 0$, for $j \in \hat{J}^c$ (i.e., all relevant variables are selected) or $\mathbf{X}_{\hat{J}}^T \mathbf{X}_j = 0$ (i.e., $\mathbf{X}_j$ is orthogonal to the linear space spanned by the column vectors of $\mathbf{X}_{\hat{J}}$), for $j \in \hat{J}^c$. That the valid confidence intervals are actually about the components of $E\boldsymbol{\beta}_{\hat{J}}$ (rather than those of $\boldsymbol{\beta}_{\hat{J}}$) 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. With regard to 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 the classical example $Y_i \sim N(\mu, 1)$, $\sqrt{T}(\bar{Y}_T - \mu)$ is highly nonpivotal when the fixed sample size n is replaced by a group sequential stopping rule T, because the distribution of T depends on μ. Hence, we require the exact method introduced by Rosner and Tsiatis (1988), which they generalize as follows. If $\{\mathcal{F}_\theta : \theta \in \Theta\}$ is indexed by a real-valued parameter θ, 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 θ is equal to θ_0. Let $R(\mathbf{Y}, \theta_0)$ be some real-valued test statistic based on the data $\mathbf{Y}$. Let $u_\alpha(\theta_0)$ be the α-quantile of the distribution of $R(\mathbf{Y}, \theta_0)$ under the distribution F_{θ_0}. The null hypothesis is accepted if $u_\alpha(\theta_0) < R(\mathbf{Y}, \theta_0) < u_{1-\alpha}(\theta_0)$. An exact equal-tailed confidence region with coverage probability $1 - 2\alpha$ consists of all θ_0 not rejected by the test, and is therefore given by $\{\theta : u_\alpha(\theta) < R(\mathbf{Y}, \theta) < u_{1-\alpha}(\theta)\}$. The exact method, however, applies only when there are no nuisance parameters. The bootstrap method replaces the quantiles $u_\alpha(\theta)$ and $u_{1-\alpha}(\theta)$ with the approximate quantiles u^*_α and $u^*_{1-\alpha}$, respectively, obtained in the following manner. Based on $\mathbf{Y}$, construct an estimate $\hat{F}$ of $F \in \mathcal{F}$. The quantile u^*_α is defined as the α-quantile of the distribution of $R(\mathbf{Y}^*, \hat{\theta})$, with $\mathbf{Y}^*$ generated from $\hat{F}$ and $\hat{\theta} = \theta(\hat{F})$, yielding the confidence region $\{\theta : u^*_\alpha < R(\mathbf{Y}, \theta) < u^*_{1-\alpha}\}$ 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 θ is the unknown parameter of interest. Let $\hat{u}_\alpha(\theta)$ be the α-quantile of the sampling distribution of $R(\mathbf{Y}, \theta)$ under the assumption that $\mathbf{Y}$ 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 $\{\theta : \hat{u}_\alpha(\theta) < R(\mathbf{Y}, \theta) < \hat{u}_{1-\alpha}(\theta)\}$; the quantiles $\hat{u}_\alpha(\theta)$ and $\hat{u}_{1-\alpha}(\theta)$ can be computed using 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 $\mathbf{x}_t = (x_{t1}, \ldots, x_{t,p_n})^T$ and $\mathcal{F}_t$-measurable ε_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 $\{(\mathbf{x}_s, y_s), 1 \leq s \leq t\}$ (i.e., observations up to time t), Lai and Wei (1982) use martingale theory to analyze the convergence properties and the asymptotic normality of least squares estimates in time series models and dynamic input-output systems. For Lasso regularization, Loh and Wainwright (2012) consider the case of stochastic regressors $\mathbf{x}_t$ that are Gaussian VAR(1) processes $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\xi}_t$ with spectral norm $\|\mathbf{A}\| < 1$, where $\{\xi_s\}$ is independent of $\{\varepsilon_s\}$. Furthermore, Basu and Michailidis

(2015) consider $\mathcal{F}_{t-1}$-measurable regressors $\mathbf{x}_t$ and $\mathcal{F}_t$-measurable errors ε_t "generated according to independent, centered, Gaussian stationary processes." They derive the asymptotic properties of the Lasso estimate of $\boldsymbol{\beta}$ under usual regularity conditions, such as the restricted eigenvalue or restricted strong convexity assumption, for a theoretical analysis of the 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 the Lasso, we develop valid post-selection confidence regions for selected parameters in stochastic regression models with conditionally heteroscedastic martingale difference random disturbances ε_t, called "martingale regression models" by Guo et al. (2017, p.36). In Sections 2.2 and 2.6 of the latter work, the authors describe the historical background and development of models of "speculative prices" in equity markets. In particular, Section 2 focuses on GARCH(1,1) errors

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

and independent and identically distributed (i.i.d.) ξ_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, as well as further discusson 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 using OGA, we can write (1.1) with $\beta_0 = 0$ as $y_t = \sum_{j \in \hat{J}} \beta_j x_{tj} + \sum_{j \in \hat{J}^c} \beta_j x_{tj} + \varepsilon_t$, for $t = 1, \ldots, n$, or $\mathbf{Y} = \mathbf{X}_{\hat{J}} \boldsymbol{\beta}_{\hat{J}} + \mathbf{w}$ in vector form, as in (1.2), where $\mathbf{w} = \mathbf{X}_{\hat{J}^c} \boldsymbol{\beta}_{\hat{J}^c} + \boldsymbol{\varepsilon}$. We begin by describing the exact method in the preceding paragraph when it is applied with resampling to construct a confidence interval for β_j, with $j \in \hat{J}$, assuming $\boldsymbol{\beta}_{\hat{J} \setminus \{j\}}$ and that the distribution of $\mathbf{w}$ is known. We compare the test statistic $R_j(\mathbf{Y}, \mathbf{X}, \theta)$ with the quantiles $u_\alpha(\theta)$ and $u_{1-\alpha}(\theta)$

of $R_j(\mathbf{Y}_j^{(b)}, \mathbf{X}, \theta)$, for $b = 1, \ldots, B$, 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 enables the hybrid resampling method to construct confidence intervals for $\beta_j,\ j \in \hat{J}$. The consistency of $\hat{\boldsymbol{\beta}}_{\hat{J}}$, the least squares estimate of $\boldsymbol{\beta}_{\hat{J}}$, in a martingale regression has recently been established by Lai, Xu, and Yuan (2020), who 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 one as $n \to \infty$. This and (1.1) with $\beta_0 = 0$ then imply that with probability approaching one 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 $\hat{\boldsymbol{\beta}}_J$. 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 ε_t. Let $\mathcal{W} = (\hat{w}_1, \ldots, \hat{w}_n)$. Because $\hat{w}_t \to \varepsilon_t$ with probability approaching one as the sample size n approaches ∞, the block bootstrap is applicable when the ε_t are asymptotically stationary; see Bühlmann (2002). For a block length $\ell (\leq n)$, define overlapping blocks $B_{i,\ell} = (\hat{w}_i, \ldots, \hat{w}_{i+\ell-1})$, for $1 \leq i \leq n - \ell + 1$. 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, such that "the 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 fails to improve upon a normal approximation impeded by the widespread use of the block bootstraps. They propose the following double block bootstrap procedure to reduce the coverage error by an order of magnitude to achieve a marked improvement over the normal approximation.

The first level of the double bootstrap gives the above block bootstrap series $\mathcal{W}^*$. For $i = 1, \ldots, m$, divide block $B^*_{i,\ell} \in \mathcal{W}^*$ into $\ell - \tilde{\ell} + 1$ subseries $B^*_{i,j,\tilde{\ell}}$, each of length $\tilde{\ell}$. The second level of the 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 $\mathcal{W}^{**}$ 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 β_j, for $j \in \hat{J}$, in the martingale regression model. Here, we use the Studentized correction of the 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

$$R_j(\mathbf{Y}, \mathbf{X}, \theta) = (\hat{\beta}_j - \theta)/(s\sqrt{c_{jj}}) \tag{2.1}$$

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 $\mathcal{W}^{**}$

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

2.2 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: $\mathcal{W}^{**}$

to test $\beta_j = \theta$, where $\hat{\beta}_j$ is the jth component of the least squares estimate $(\mathbf{X}^T_{\hat{J}}\mathbf{X}_{\hat{J}})^{-1}\mathbf{X}_{\hat{J}}\mathbf{Y}_{\hat{J}}$ of $\boldsymbol{\beta}_{\hat{J}}$, c_{jj} is the jth diagonal element of $(\mathbf{X}^T_{\hat{J}}\mathbf{X}_{\hat{J}})^{-1}$, $s^2 = (\sum_{t=1}^n w_t^2)/(n-m_n)$, and $m_n = O(\sqrt{n/\log p_n})$ is the cardinality of the set $\hat{J}$ of regressors selected using OGA+HDIC. The quantiles $\hat{u}_\alpha(\theta)$ and $\hat{u}_{1-\alpha}(\theta)$ of $R_j(\mathbf{Y}, \mathbf{X}, \theta)$ are evaluated using Algorithm 1. Although the $(1-2\alpha)$-level confidence set $\{\theta : \hat{u}_\alpha(\theta) < R_j(\mathbf{Y}, \mathbf{X}, \theta) < \hat{u}_{1-\alpha}(\theta)\}$ may not be an interval, it often suffices to give the upper limit $\theta^U_j = \arg\min_\theta |R_j(\mathbf{Y}, \mathbf{X}, \theta) - \hat{u}_\alpha(\theta)|$ and the lower limit $\theta^L_j = \arg\min_\theta |R_j(\mathbf{Y}, \mathbf{X}, \theta) - \hat{u}_{1-\alpha}(\theta)|$ as the confidence interval $[\theta^L_j, \theta^U_j]$ of β_j; see Chuang and Lai (2000, p.4). We choose $m = m_n$ in the double block bootstrap (Algorithm 1), and hence the block length ℓ, using OGA+HDBIC, as in Section 4 of Ing and Lai (2011, p.1484):

$$m_n = \arg \min_{1 \le k \le 2\lfloor n/\log p_n \rfloor} \{n \log \hat{\sigma}^2_{\{\hat{j}_1, \ldots, \hat{j}_k\}} + k(\log n)(\log p_n)\}, \tag{2.2}$$

where $\hat{j}_1, \ldots, \hat{j}_k$ are selected along the OGA path until the kth 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 $\mathbf{Y}$ is projected into the linear space spanned by the column vectors of $\mathbf{X}_J$; see Ing and Lai (2011, p.1479). In the Supplementary Material S1, we prove the following theorem, showing that the post-selection confidence interval for β_j, for $j \in \hat{J}$, constructed using this hybrid resampling approach has coverage error $o(n^{-1/2})$, under certain regularity conditions, if ℓ is chosen suitably.

Theorem 1. *If $\ell \propto n^{1/3}$, then the preceding confidence interval for β_j, for $j \in \hat{J}$, 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 the 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 from ignoring structural changes in the GARCH model, respectively. Their Section 9.5.2 also describes change-point AR-GARCH models, definitively generalized by Lai and Xing (2013), who use a Bayesian change-point ARX-GARCH model of the form

$$y_t = \boldsymbol{\beta}_t^T \mathbf{x}_t + \nu_t \sqrt{h_t} \epsilon_t, \tag{2.3}$$

where ϵ_t are i.i.d. with a standardized Student t-distribution with $\kappa > 2$ degrees of freedom and are independent of $\mathbf{x}_t$, the parameter vector $\boldsymbol{\beta}_t$ and the unconditional variance ν_t^2 are piecewise constant with jumps at times of structural change, the vector $\mathbf{x}_t$ consists of exogenous variables and past observations $y_{t-1}, y_{t-2}, \ldots, y_{t-k}$, and h_t represents short-term proportional fluctuations in the variance generated by the GARCH model

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

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 $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t^T, \tau_t)^T$ is piecewise constant and satisfies the following conditions:

(A1) Let $t_0 = \max(k, k')$, $I_{t_0} = 1$, and there be no change point prior to t_0. For $t > t_0$, the change times of $\boldsymbol{\theta}_t$ form a renewal process with i.i.d. inter-arrival times that are geometrically distributed with parameter

p or, equivalently,

$$I_t := 1_{\{\boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t-1}\}} \text{ are i.i.d. Bernoulli}(p) \text{ with } P(I_t = 1) = p. \qquad (2.5)$$

(A2) $\boldsymbol{\theta}_t = (1 - I_t)\boldsymbol{\theta}_{t-1} + I_t(\mathbf{z}_t^T, \gamma_t)^T$, where $(\mathbf{z}_1^T, \gamma_1)^T, (\mathbf{z}_2^T, \gamma_2)^T, \ldots$ are i.i.d. random vectors such that $\mathbf{z}_t \mid \gamma_t \sim \text{Normal}(\mathbf{z}, \mathbf{V}/(2\gamma_t)), \gamma_t \sim \chi_d^2/\rho$, with χ_d^2 the chi-square distribution with d degrees of freedom.

(A3) The processes $\{I_t\}, \{(\mathbf{z}_t^T, \gamma_t)\}$, and $\{(\mathbf{x}_t, \epsilon_t)\}$ are independent.

We now focus on the case $k = k' = 1$ in (2.4), that is, 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 $\boldsymbol{\eta} = (a, b, \kappa), p, \mathbf{z}, \mathbf{V}, \rho$, 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 χ_d^2/ρ prior distribution for $1/(2\nu_t^2)$ given by (A2) is written as Gamma$(d/2, \rho/2)$ so that d does not need to be an integer. Second, $\mathbf{z}, \mathbf{V}, \rho$, and d are estimated by applying the following method of moments to the "stationary distribution of the Markov chain $(I_t, \boldsymbol{\theta}_t, \epsilon_t)$ that is partially observed via $(\mathbf{x}_t, y_t), 1 \leq t \leq n$." From (A2) and (A3), it follows that $E(\boldsymbol{\beta}_t) = \mathbf{z}, \text{Cov}(\boldsymbol{\beta}_t) = (E\nu_t^2)\mathbf{V}$, and $E(\mathbf{x}_t y_t) = E(\mathbf{x}_t \mathbf{x}_t^T)\mathbf{z}$. From $n - L$ moving windows $\{(\mathbf{x}_t, y_t) : s \leq t \leq s + L\}$

of these data, compute the least squares estimate, and denote it as $\widetilde{\boldsymbol{\beta}}^{(s)}$. Each $\widetilde{\boldsymbol{\beta}}^{(s)}$ is a method-of-moments estimate of $\mathbf{z}$, and thus so is $\overline{\boldsymbol{\beta}} = (n - L)^{-1} \sum_{s=1}^{n-L} \widetilde{\boldsymbol{\beta}}^{(s)}$. If an oracle revealed 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 $\widetilde{\boldsymbol{\beta}}^{(s)}$ based on the moving window $\{(\mathbf{x}_t, y_t) : s \le t \le s + L\}$. 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 $e_t = y_t - \mathbf{x}_t^T \widetilde{\boldsymbol{\beta}}^{(s)}$ 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 $\bar{r} := (n-L)^{-1} \sum_{s=1}^{n-L} \hat{r}_s$. In this connection, recall that $\boldsymbol{\theta}_t^T = (\boldsymbol{\beta}_t^T, \tau_t)$ and $2\nu_t^2 = 1/\tau_t$, and note that by (A2), $\boldsymbol{\beta}_t | \nu_t^2 \sim N(\mathbf{z}, 2\nu_t^2 \mathbf{V})$ and $2\nu_t^2 (= 1/\tau_t)$ 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, $2\rho^2/[(d-2)^2(d-4)]$ 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, $\rho^2/2[(d-2)^2(d-4)]$ can be estimated by the average $\bar{v} := (n-L)^{-1} \sum_{s=1}^{n-L} \hat{v}_s$ over the moving windows. Because $\mathbf{V} = \mathrm{Cov}(\boldsymbol{\beta}_t)/E\nu_t^2$, an obvious method-of-moments estimate of $\mathbf{V}$

is $\widetilde{\mathbf{V}} := (n-L)^{-1}\sum_{s=1}^{n-L}\widehat{\mathbf{Cov}}_s/\hat{r}_s$, where $\widehat{\mathbf{Cov}}_s$ is the sample covariance matrix of $\boldsymbol{\beta}^{(s)}$ for moving window s. Lastly, "with $\mathbf{z}, \mathbf{V}, \rho$, and d replaced by these estimates," $\boldsymbol{\eta}$ and p can be estimated by maximum likelihood using a recursive representation of the log-likelihood function based on $(\mathbf{x}_t, y_t)$, for $1 \leq t \leq n$; see Eq.(2.5) of Lai and Xing (2013, p.1578), where the consistency of these hyperparameter estimates (as $n \to \infty$ and with pL sufficiently small) is also established.

2.3 Sequential Monte Carlo to sample the latent variables I_t,$\boldsymbol{\theta}_t$

Chen and Lai (2007) and Chan and Lai (2013) have shown how particle filters that use a sequential Monte Carlo (SMC) with 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 ϵ_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 $\mathbf{z}, \mathbf{V}, \rho, d, p$, and $\boldsymbol{\eta}$ 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 assume ϵ_t is 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 $(\boldsymbol{\beta}_s^T, 2\nu_s^2)$ that does not change for $C_t \leq s \leq t$, the posterior distribution of $\boldsymbol{\theta}_t$ (with components $\boldsymbol{\beta}_t$ and $\tau_t = (2\nu_t^2)^{-1}$), given C_t and $\{(\mathbf{x}_s, y_s) : C_t \leq s \leq t\}$, also belongs to this conjugate family, with $\boldsymbol{\beta}_t | \nu_{C_t}^2 \sim N(\boldsymbol{\mu}_{C_t,t}, \nu_{C_t}^2 \mathbf{V}_{C_t,t})$ and $\tau_t \sim \text{Gamma}(d/2 + (t - C_t)/2, \rho_t/2)$, where

$$\begin{aligned} \boldsymbol{\mu}_{s,t} &= \mathbf{V}_{s,t}(\mathbf{V}^{-1}\mathbf{z} + \sum_{i=s}^{t} \mathbf{y}_i \mathbf{x}_i), \quad \mathbf{V}_{s,t} = (\mathbf{V}^{-1} + \sum_{i=s}^{t} \mathbf{x}_i \mathbf{x}_i^T)^{-1}, \\ \rho_t &= \rho + \mathbf{z}^T \mathbf{V}^{-1} \mathbf{z} + \sum_{s=C_t}^{t} y_s^2 - \boldsymbol{\mu}_{C_t,t}^T \mathbf{V}_{C_t,t}^{-1} \boldsymbol{\mu}_{C_t,t}, \end{aligned} \tag{2.6}$$

assuming the hyperparameters of the Bayesian model and C_t are specified; see Box and Tiao (1973, Chapter 8). Without prespecifying the hyperparameters $\mathbf{z}, \mathbf{V}, \rho$, and d, we can replace them in (2.6) with their method-of-moments estimates described in Section 2.2, which also gives the estimates $\tilde{p}$ and $\tilde{\boldsymbol{\eta}}$ of the hyperparameters p and $\boldsymbol{\eta}$, respectively in (A1) and (2.4). Replacing $\boldsymbol{\beta}_t, \nu_t$, and h_t in $\epsilon_t = (y_t - \boldsymbol{\beta}_t^T \mathbf{x}_t)/(\nu_t \sqrt{h_t})$ (from (2.3)) with their estimates involves C_t, which we now construct using an SMC with sequential importance sampling and resampling, as described by Chen and Lai (2007, Section II) and Chan and Lai (2013, Section 2.1).

To explain the SMC, we first assume that the hyperparameters $\mathbf{z}, \mathbf{V}, \rho, d, p,$ and $\boldsymbol{\eta}$ 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, respectively, of the random variables indicated. Sequential importance sampling samples $I_{t_0} = 1, I_{t_0+1}, \ldots, I_n$ sequentially from the proposed distribution, for which $I_t|(I_{t_0}, \ldots, I_{t-1})$ is Bernoulli, assuming the values one and zero with respective probabilities that are proportional to $a_t(p, \boldsymbol{\eta}) : b_t(p, \boldsymbol{\eta})$, where

$$\begin{aligned} a_t(p, \boldsymbol{\eta}) &= \frac{p}{(\mathbf{x}_t^T \mathbf{V} \mathbf{x}_t + \nu_t^2 h_t)^{1/2}} \pi\Big(\frac{y_t}{(\mathbf{x}_t^T \mathbf{V} \mathbf{x}_t + \nu_t^2 h_t)^{1/2}}\Big), \\ b_t(p, \boldsymbol{\eta}) &= \frac{1-p}{(\mathbf{x}_t^T \mathbf{V}_{C_t,t} \mathbf{x}_t + \nu_t^2 h_t)^{1/2}} \pi\Big(\frac{y_t - \boldsymbol{\mu}_{C_t,t}^T \mathbf{x}_t}{(\mathbf{x}_t^T \mathbf{V}_{C_t,t} \mathbf{x}_t + \nu_t^2 h_t)^{1/2}}\Big), \end{aligned} \tag{2.7}$$

in which π is the standardized Student t-density (with κ degrees of freedom) for ϵ_t in (2.3), and h_t is defined recursively by (2.4), with $k = k' = 1$ and the GARCH parameter vector $\boldsymbol{\eta} = (a, b, \kappa)$. 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 weights,

$$w_t = w_{t-1}\{a_t(p,\boldsymbol{\eta}) + b_t(p,\boldsymbol{\eta})\}, t > t_0; w_{t_0} = 1, \tag{2.8}$$

for the posterior distribution of $I_{t_0:t} := (I_{t_0}, \dots, I_t)$. For the Monte Carlo implementation, B sample paths are generated, with importance weights $w_t^{(b)}(b = 1, \dots, B)$. As such, the Monte Carlo estimate of the posterior mean $E(\boldsymbol{\theta}_t|(\mathbf{x}_s, y_s) : s \le t)$ is given by

$$\boldsymbol{\beta}_{t,\text{Bayes}} = (\sum_{b=1}^{B} w_t^{(b)} \boldsymbol{\mu}_{C_t^{(b)},t}) / \sum_{b=1}^{B} w_t^{(b)}, \quad 2\nu^2_{t,\text{Bayes}} = \sum_{b=1}^{B} \frac{\rho_t}{(d+t-C_t)} \frac{w_t^{(b)}}{\sum_{b=1}^{B} w_t^{(b)}}, \tag{2.9}$$

where $C_t^{(1)}, \dots, C_t^{(B)}$ 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 $\mathbf{z}, \mathbf{V}, \rho$, and d, followed by maximum likelihood estimates of $\boldsymbol{\eta}$ and p, as in the second paragraph of Section 2.2. These estimates are used in (2.7) and (2.8), which 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 $t_0, t_1, \dots (< n)$

under the posterior distribution, given the observations $(\mathbf{x}_t, y_t), 1 \leq t \leq n$. After the change times are generated, the moving windows of length $L+1$ in the second paragraph of Section 2.2 can be replaced with the times $t \in \{t_i, t_i+1, \ldots, t_{i+1}-1\}$ between consecutive change-points. Thus, the hyperparameters are re-estimated using these time segments during which $\boldsymbol{\theta}_t$ stays constant. These revised hyperparameter estimates are denoted by $\hat{\mathbf{z}}, \widehat{\mathbf{V}}, \hat{\rho}, \hat{d}, \hat{\boldsymbol{\eta}}$, 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). The algorithm describes Monte Carlo simulations of B independent sample paths $I_{t_0:n}^{(b)}$ from the estimated posterior distribution, given $\{(\mathbf{x}_t, y_t) : 1 \leq t \leq n\}$, with importance weights that use $\hat{\mathbf{z}}, \widehat{\mathbf{V}}, \hat{\rho}, \hat{d}, \hat{\boldsymbol{\eta}}$, and $\hat{p}$ for 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 zero while having to sum to one. To address this difficulty, bootstrap resampling is 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, the SMC consists of an importance sampling step followed by a resampling that transforms the weighted sample to a bootstrap sample with equal weights. Section 2.1 of Chan and Lai (2013) provides an asymptotic theory of the SMC with bootstrap resampling at every stage t. In their Section 2.2, they extend this to occasional resampling when the coefficient of variation of the normalized weights exceeds some threshold, as 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 that are used to segment the data so that the hyperparameters can be re-estimated using these time segments. Specifically, recall that we use $n - L$ moving windows of length $L + 1$ to replace the time segments between two consecutive change times "revealed by an oracle" to estimate the hyperparameters using the method of moments or maximum likelihood. In lieu of the oracle, Algorithm 2 simulates B samples of $I_{t0:n}$ from its posterior distribution, given the observations. From each simulated sample $I_{t0:n}^{(b)}$, 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 $\hat{\mathbf{z}}$, $\hat{\mathbf{V}}$, $\hat{\rho}$, and $\hat{d}$ (respectively, $\hat{\boldsymbol{\eta}}$ and $\hat{p}$).

Algorithm 2 plays a basic role in constructing hybrid resampling cred-

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

INPUT: Observations $(\mathbf{x}_t, y_t), 1 \le t \le n$

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

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

OUTPUT: $I^{(b)}_{t_0:n}, b = 1, \ldots, B$, and $\hat{\mathbf{z}}, \widehat{\mathbf{V}}, \hat{\rho}, \hat{d}, \hat{\boldsymbol{\eta}}, \hat{p}$

ible intervals for the regression parameters $\boldsymbol{\beta}_t$ and the GARCH parameters ν^2_t in the change-point ARX-GARCH(1,1) model (2.3), similar to that played by Algorithm 1 (double block bootstrap) for the hybrid confidence intervals for the selected regression parameters in Section 2.1. For ν^2_t and scalar functions of $\boldsymbol{\beta}_t$, we construct EB credible intervals for their values over a designated period, which may contain zero,one 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." Lai and Xing (2011, pp.539–540) provide a survey of the frequentist approach, including the works of Bai and Perron (1998, 2003), and Olshen et al. (2004) on segmentation, and those of Yao (1988) , Birgé and Massart (2001), Zhang and Siegmund (2006), and Davis et al. (2006) on model selection criteria. Their Section 4 describes 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 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, in addition to the hyperparameter vector $\boldsymbol{\eta}$ 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 the piecewise constant ν_t

and $\beta_{t,i}$ in the ARX-GARCH(1,1) model. The following theorem, which is proved in the 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

We begin with a simulation study in Section 3.1 of the coverage errors of the selected confidence intervals $[\theta_j^L, \theta_j^U]$, for $j \in \hat{J}$, in the regression model (1.1) with $\beta_0 = 0$ and GARCH(1,1) errors ε_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 $[\theta_j^L, \theta_j^U]$, for $j \in \hat{J}$, by simulating the martingale regression model (1.1) with $\beta_0 = 0$ and GARCH(1,1) errors ε_t; see (1.3). We consider $n = 400$, $p_n = 500$, the regressors x_{tj} generated by i.i.d. standard normal z_{tj} and $\tilde{z}_t$ from $x_{tj} = f_t + z_{tj}$ and $f_t = 0.9f_{t-1} + \tilde{z}_t(f_0 = 0)$, and $\alpha = 0.1$ for the nominal confidence level $1 - 2\alpha$ (hence, coverage error 2α). 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 $[\theta_j^L, \theta_j^U]$, for five parameter settings of the martingale regression model. In the last setting, the GARCH parameters $a = 0 = b$ and $\nu = \infty$; hence, $\varepsilon_t \sim 0.1N(0, 1)$ 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 using this conditional approach differ substantially from the nominal value of 20%, even though it is applicable in this setting. This is due to the spill-over effect. In contrast, the hybrid resampling confidence intervals $[\theta_j^L, \theta_j^U]$, for $j \in \hat{J}$, have coverage errors close to 20%. Note that the double block bootstrap in Algorithm 1 is nonparametric and assumes only that ε_t is asymptotically stationary. On the other hand, if it is assumed that ε_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 β_j, for $j \in \hat{J}$. This is similar to the 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, pp.13–15, 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

We 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 January 3, 2001, to December 31, 2008. To do so, we use weekly log returns of Wells Fargo & Company (WFC) and a 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 data set 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 the ARX-GARCH(1,1) model to the data with the following time-invariant parameters:

$$y_t = \beta_1 x_{t-1} + \beta_2 y_{t-1} + w_t, \quad w_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega + a w_{t-1}^2 + b\sigma_{t-1}^2. \tag{3.1}$$

The results, 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

$$y_t = \beta_{t,1} x_{t-1} + \beta_{t,2} y_{t-1} + \nu_t w_t, \quad w_t = \sqrt{h_t}\epsilon_t, \quad h_t = 1 - a - b + a w_{t-1}^2 + b h_{t-1} \tag{3.2}$$

to these data. The four identified change times correspond to December 26, 2002, April 26, 2007, December 22, 2007 and July 5, 2008. These dates correspond to the following significant economic events in the U.S. economy: the aftermath of the September 11, 2011, terrorist attacks and the preparation for the Iraq War, following congressional authorization by President Bush to launch a military attack against Iraq; concern for the sustainability of the upward earnings trend of the DJIA (Dow Jones Industrial Average), which actually switched to a negative regime on April 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 and broadened it to tackle two challenging problems related to inferences on martingale regression models. The effectiveness of our enhancement in terms of constructing valid confidence intervals for selected regression parameters in high-dimensional martingale regression models after model selection under sparsity assumptions, and in terms of empirical Bayes estimations of the change times with credible intervals for the piecewise constant regression coefficients and volatility parameters in change-point martingale regression models, opens up new possibilities and applications in time series analysis.

Supplementary Materials The online supplement contains the following sections:

S1 Proof of Theorem 1 and simulation study in Section 3.1

S2 Proof of Theorem 2 and empirical study in Section 3.2

Acknowledgements This research is supported by the National Natural Science Foundation of China under Contract No. 12001461.

References

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* **18**, 1–22.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Annals of Statistics* **43**, 1535–1567.

Belloni, A., Chernozhukov, V. and Kato, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102**, 77–94.

Birgé, L. and Massart, P. (2001). Gaussian model selection. *Journal of the European Mathematical Society* **3**, 203–268.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Wiley, Hoboken NJ.

Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science* **17**, 53–72.

Chan, H. P. and Lai, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Annals of Statistics* **41**, 2877–2904.

Chen, Y. and Lai, T. L. (2007). Identification and adaptive control of change-point ARX models via Rao-Blackwellized particle filters. *IEEE transactions on automatic control* **52**, 67–72.

Chuang, C.S. and Lai, T.L. (1998). Resampling methods for confidence intervals in group sequential trials. *Biometrika* **85**, 317–332.

Chuang, C.S. and Lai, T.L. (2000). Hybrid resampling methods for confidence intervals (with discussion and rejoinder). *Statistica Sinica* **10**, 1–50.

Davis, R., Lee, T. C. M. and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101**, 223–239.

Guo, X., Lai, T.L., Shek, H. and Wong, S.P.S. (2017). *Quantitative Trading: Algorithms, Analytics, Data, Models, Optimization.* Chapman and Hall/CRC.

Ing, C.K. and Lai, T.L. (2011). A stepwise regression method and consistent model selection for high-dimensional sparse linear models. *Statistica Sinica* **21**, 1473–1513.

Ing, C. K. and Lai, T. L. (2015). Fixed-size confidence regions in high-dimensional sparse linear regression models. *Sequential Analysis* **34**, 324–335.

Ing, C.K., Lai, T.L., Shen, M., Tsang, K.W. and Yu, S.H. (2017). Multiple testing in regression models with applications to fault diagnosis in the big data era. *Technometrics* **59**, 351–360.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* **10**, 154–166.

Lai, T.L. and Xing, H. (2008). *Statistical Models and Methods in Financial Markets*. Springer, New York.

Lai, T. L. and Xing, H. (2011). A simple Bayesian approach to multiple change-points. *Statistica Sinica* **21**, 539–569.

Lai, T. L. and Xing, H. (2013). Stochastic change-point ARX-GARCH models and their applications to econometric time series. *Statistica Sinica* **23**, 1573–1594.

Lai, T.L., Xu, H. and Yuan, H. (2020). Self-normalized martingales, recursive orthogonal matching pursuit and modified gradient boosting in high-dimensional stochastic regression models. *Technical Report* , Department of Statistics, Stanford Univ.

Lee, J.D. and Taylor, J.E. (2014). Exact post model selection inference for marginal screening. In *Advances in Neural Information Processing Systems* **27**, 136–144.

Lee, S. M. S. and Lai, P. Y. (2009). Double block bootstrap confidence intervals for dependent data. *Biometrika* **96**, 427–443.

Lee, S. M. S. and Wu, Y. (2018). A bootstrap recipe for post-model-selection inference under linear regression models. *Biometrika* **105**, 873–890.

Lockhart, R., Taylor, J., Tibshirani, R. and Tibshirani, R. (2014). A significance test for the Lasso. *Annals of Statistics* **42**, 413–468.

Loh, P. L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Annals of Statistics* **40**, 1637–1664.

Olshen, A. B., Venkatraman, E. S., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572.

Rosner, G.L. and Tsiatis, A.A. (1988). Exact confidence intervals following a group sequential trial: A comparison of methods. *Biometrika* **75**, 723–729.

Sorić, B. (1989). Statistical "discoveries" and effect-size estimation. *J. Amer. Statist. Assoc.* **84**, 608–610

Yao, Y. (1988). Estimating the number of change-points via Schwarz criterion. *Statistics and Probability Letters* **6**, 181–189.

Zhang, N. and Siegmund, D. (2006). A modified Bayes information criterion with applications to comparative genomic hybridization data. *Biometrics* **63**, 22-32.

Zhang, C. H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. Roy. Statist. Soc. Ser B* **76**, 217–242.

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