View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Institutional Knowledge at Singapore Management University

Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection School Of Economics | School of Economics

8-2014

# Shrinkage Estimation of Regression Models with Multiple Structural Changes

Junhui QIAN
*Shanghai Jiaotong University*

Liangjun SU
*Singapore Management University*, ljsu@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/soe_research

 Part of the Econometrics Commons

Citation
QIAN, Junhui and SU, Liangjun. Shrinkage Estimation of Regression Models with Multiple Structural Changes. (2014). 1-51. Research Collection School Of Economics.
**Available at:** https://ink.library.smu.edu.sg/soe_research/1595

This Working Paper is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Shrinkage Estimation of Regression Models with Multiple Structural Changes

Junhui Qian and Liangjun Su

August 2014

Paper No. 06-2014

ANY OPINIONS EXPRESSED ARE THOSE OF THE AUTHOR(S) AND NOT NECESSARILY THOSE OF THE SCHOOL OF ECONOMICS, SMU

# Shrinkage Estimation of Regression Models with Multiple Structural Changes*

Junhui Qian
Antai College of Economics and Management, Shanghai Jiao Tong University

Liangjun Su
School of Economics, Singapore Management University

July 21, 2014

**Abstract**

In this paper we consider the problem of determining the number of structural changes in multiple linear regression models via group fused Lasso (*least absolute shrinkage and selection operator*). We show that with probability tending to one our method can correctly determine the unknown number of breaks and the estimated break dates are sufficiently close to the true break dates. We obtain estimates of the regression coefficients via post Lasso and establish the asymptotic distributions of the estimates of both break ratios and regression coefficients. We also propose and validate a data-driven method to determine the tuning parameter. Monte Carlo simulations demonstrate that the proposed method works well in finite samples. We illustrate the use of our method with a predictive regression of the equity premium on fundamental information.

**JEL Classification:** C13, C22

**Key Words:** Change point; Fused Lasso; Group Lasso; Penalized least squares; Structural change

## 1 Introduction

Since the 1950s a voluminous literature on issues related to structural changes has been developed. As Perron (2006) remarks, early works were mostly designed for the specific case of a single change. Andrews (1993) proposes supremum-type (*sup*-type) test for a one-time break in the GMM framework. Andrews

*The authors would like to thank the co-editor Victor Chernozhukov and an anonymous referee for constructive comments and suggestions. They also thank participants in the China Meeting of Econometric Society (CMES 2014) and SJTU-SMU Econometrics Conference (2014) at Shanghai for helpful comments. Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021. Address Correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; E-mail: ljsu@smu.edu.sg, Phone: +65 6828 0386.

and Ploberger (1994) consider the exponential-type (*exp*-type) and average-type (*avg*-type) tests for a one-time break in linear regression models and investigate their optimality properties under Pitman local alternatives. Bai (1995) and Bai (1998) consider the median estimation of a regression model with a single break and multiple breaks, respectively. Bai (1997a) and Bai (1997b) study the least squares estimation of a regression model with a single break and with multiple breaks, respectively. Bai and Perron (1998) extend the *sup*-type test to models with multiple changes and propose a double maximum test against the alternative under which only the maximum number of breaks is prescribed. They also consider a sequential test for the null hypothesis of $\ell$ breaks against the alternative of $\ell+1$ breaks. Bai et al. (1998) consider a sup Wald test for a single change in a multivariate system, Qu and Perron (2007) extend the analysis to the context of multiple structural changes in multivariate regressions, and Kurozumi and Arai (2006) study inferential problems for multivariate time series with change points, all allowing stationary or integrated regressors as well as trends. Su and White (2010) consider tests of structural changes in semiparametric models. As Bai and Perron (2006) show, the multiple structural change tests tend to be more powerful than the single structural change tests when multiple breaks are present. For a comprehensive survey on structural changes, see Perron (2006).

Despite the satisfactory power properties of multiple structural change tests, they are subject to some practical problems. First, one major practical difficulty is that one needs to consider all permissible partitions of the sample in order to construct the *avg*- and *exp*-type test statistics, the number of which is proportional to $T^m$ with $T$ and $m$ being the total number of observations in the sample and the number of breaks under the alternative. When $m \geq 3$, the computational burden can be prohibitively heavy. For this reason, Bai and Perron (2003a) propose an efficient dynamic-programming-based algorithm to compute the *sup*-type test statistic, which requires only $O\left(T^2\right)$ computations for any fixed number of breaks. Andrews (1993) and Bai and Perron (1998, 2003b) tabulate the critical values for the *sup*-type test for a one-time break and multiple breaks, respectively. Andrews and Ploberger (1994) tabulate critical values for the *exp*- and *ave*-type tests for a one-time break. The critical values for the last two types of tests in the case of multiple breaks have not been available until Kurozumi (2012) who tabulate the critical values for the *exp*-type test for at most three breaks and those for the *sup*- and *ave*-type tests for up to five breaks because the computation for the former test is prohibitively expensive in the case of $m \geq 3$ whereas the latter two tests only require $O\left(T^2\right)$ operations for any given number of breaks under the alternative. Second, for all tests for structural changes in the literature one has to apply some trimming parameter, say, by trimming $100\epsilon$ percentage of tail observations, and by requiring the minimum length of a segment be $\epsilon T$, where $\epsilon$ typically take values from 0.05 to 0.25. Not only the asymptotic distribution but also the finite sample performance of the test statistics heavily depend on the choice of $\epsilon$. One may draw different conclusions for different choices of $\epsilon$ and the desirable choice of $\epsilon$ heavily depends on the underlying data generating process (DGP). See Bai and Perron (2003a, 2006) for discussions on the importance of the choice of $\epsilon$ for the size and power of the test. Third, the asymptotic distributions of the test statistics depend on the number of regressors in the model. It remains unknown how the presence of irrelevant regressors affects the performance of the tests. Another undesirable feature of the test of no break versus a fixed number of breaks is that one has to pick a number of breaks under

the alternative, as practitioners often do not wish to pre-specify a particular number of breaks before making inferences.

In this paper we explore a different approach to the study of issues related to structural changes in regression models. For clarity, we focus on structural changes in a linear regression framework. But our methodology can be easily extended to the GMM framework, quantile regression, and system of equations. Unlike the early literature which tries to test the number of breaks first and then conduct estimation and inference subsequently, we focus on the simultaneous estimation of the number of breaks and model parameters via the method of group fused Lasso (*least absolute shrinkage and selection operator*). See Tibshirani (1996) for the introduction of Lasso and Knight and Fu (2000) for the first systematic study of the asymptotic properties of Lasso-type estimators. Tibshirani et al. (2005) propose a total-variation-based shrinkage technique, namely, the fused Lasso, a generalization of the Lasso designed for problems with features that can be ordered in some meaningful way. It penalizes the $L_1$-norm of both the coefficients and their successive differences and encourages sparsity of both the coefficients and their differences. Friedman et al. (2007) propose a pathwise coordinatewise optimization algorithm to solve the fused Lasso problem. Rinaldo (2009) considers three interrelated least squares procedures for the fused Lasso and study their asymptotic properties in the context of estimating an unknown blocky and sparse signal. Harchaoui and Lévy-Leduc (2010) apply the idea of fused Lasso to study the change point problem in one-dimensional piecewise constant signals. Bleakley and Vert (2011) propose fast algorithms to solve the *group fused Lasso* (hereafter GFL) problem to detect change points in a signal, and Angelosante and Giannakis (2012) develop an efficient block-coordinate descent algorithm to estimate piecewise-constants in time-varying autoregressive models. But they do not study the asymptotic properties of the resulting estimators of break points or regression coefficients.

We show that under suitable conditions on the tuning parameter, minimum regime length, minimum break size, and the underlying data generating process (DGP), the GFL procedure can not under-estimate the number of breaks in the DGP, and when the number of estimated breaks coincides with the true number of breaks, all break points can be "consistently" estimated as in Bai and Perron (1998). We further propose a BIC-type information criterion to determine a data-driven tuning parameter that can yield the correct number of breaks with probability approaching one (w.p.a.1). The limiting distributions of the break date estimates, the regression coefficients estimates and their post-Lasso versions are also derived. We emphasize that we derive all asymptotic results under a set of fairly general conditions. In particular, the number of observations within each regime may not be proportional to the sample size, the break magnitudes may differ across different break points, and the number of breaks may diverge to infinity as the sample size passes to infinity. Simulations demonstrate that our procedure works reasonably well in finite samples in comparison of the commonly used approach by Bai and Perron (1998, 2003a).

To proceed, it is worth mentioning that our paper contributes to the recent literature on the applications of Lasso-type shrinkage techniques in econometrics. These include Caner (2009) and Fan and Liao (2011) who consider covariate selection in GMM estimation; Belloni et al. (2012), Caner and Fan (2011), García (2011), and Liao (2013) who consider instruments or moment conditions selection in the GMM framework. In addition, Caner and Knight (2013) and Kock (2013) apply bridge estimators to differ-

entiate a unit root from a stationary alternative and to study oracle efficient estimation of linear panel data models with fixed or random effects, respectively; Liao and Phillips (2014) apply adaptive shrinkage techniques to cointegrated systems; Lu and Su (2013) apply adaptive group Lasso to select both relevant regressors and the number of unobserved factors in panel data models with interactive fixed effects.

The rest of the paper is organized as follows. Section 2 introduces our GFL procedure. Section 3 analyzes its asymptotic properties. Section 4 reports the Monte Carlo simulation results. Section 5 provides an empirical application and Section 6 concludes. All proofs are relegated to the appendix.

NOTATION. Throughout the paper we adopt the following notation. For an $m \times n$ real matrix $A$, we denote its transpose as $A'$, its Frobenius norm as $\|A\|$ $(\equiv [\mathrm{tr}\,(AA')]^{1/2})$, and its Moore-Penrose generalized inverse as $A^{+}$. When $A$ is symmetric, we use $\mu_{\max}(A)$ and $\mu_{\min}(A)$ to denote its largest and smallest eigenvalues, respectively. $\mathbb{I}_p$ denotes a $p \times p$ identity matrix and $\mathbf{0}_{a\times b}$ an $a \times b$ matrix of zeros. Let $\mathbf{1}\{\cdot\}$ denote the usual indicator function. The operator $\xrightarrow{P}$ denotes convergence in probability, $\xrightarrow{d}$ convergence in distribution, $\Rightarrow$ weak convergence, and plim probability limit.

# 2 Penalized Estimation of Linear Regression Models with Multiple Breaks

In this section we consider a linear regression model with an unknown number of breaks, which we estimate via the GFL.

## 2.1 The model

Consider the following linear regression model

$$y_t = \beta_t' x_t + u_t,\ t = 1, \ldots, T, \tag{2.1}$$

where $x_t$ is a $p \times 1$ vector of regressors, $u_t$ is the error term, and $\beta_t$ is a $p \times 1$ vector of unknown coefficients. We assume that the $\{\beta_1, ..., \beta_T\}$ exhibit certain *sparse* nature such that the total number of distinct vectors in the set is given by $m+1$, which is unknown but assumed to be much smaller than the sample size $T$. More specifically, we assume that

$$\beta_t = \alpha_j \text{ for } t = T_{j-1}, ..., T_j - 1 \text{ and } j = 1, ..., m+1$$

where we adopt the convention that $T_0 = 1$ and $T_{m+1} = T + 1$. The indices $T_1, ..., T_m$ indicate the unobserved $m$ break points/dates and the number $m+1$ denotes the total number of regimes. We are interested in estimating the *unknown* number $m$ of *unknown* break dates and the regression coefficients. Let $\boldsymbol{\alpha}_m = (\alpha_1', ..., \alpha_{m+1}')'$ and $\mathcal{T}_m = (T_1, ..., T_m)$.

Throughout, we denote the true value of a parameter with a superscript 0. In particular, we use $m^0$, $\boldsymbol{\alpha}^0_{m^0} = \left(\alpha_1^{0\prime}, ..., \alpha_{m^0+1}^{0\prime}\right)'$ and $\mathcal{T}^0_{m^0} = \left(T_1^0, ..., T_{m^0}^0\right)$ to denote the true number of breaks, the true vector of regression coefficients, and the true vector of break dates, respectively. Hence the data generating process

is assumed to be

$$y_t = \beta_t^{0\prime} x_t + u_t,\ t = 1, \ldots, T, \tag{2.2}$$

where $\beta_t^0 = \alpha_j^0$ for $t = T_{j-1}^0, ..., T_j^0 - 1$ and $j = 1, ..., m^0 + 1$; $T_0^0 = 1$ and $T_{m^0+1}^0 = T + 1$.

## 2.2 Penalized least squares estimation of $\{\beta_t\}$

Since neither $m$ nor the break dates are known and $m$ is typically much smaller than $T$, this motivates us to consider the estimation of $\beta_t$'s and $\mathcal{T}_m$ via a variant of fused Lasso *a la* Tibshirani et al. (2005). We propose to estimate $\{\beta_t\}$ by minimizing the following penalized least squares (PLS) objective function

$$V_{T\lambda}\left(\{\beta_t\}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - \beta_t' x_t\right)^2 + \lambda \sum_{t=2}^{T}\left\|\beta_t - \beta_{t-1}\right\| \tag{2.3}$$

where $\lambda = \lambda_T$ is a positive tuning parameter and $\|\cdot\|$ denotes the Frobenius norm. Harchaoui and Lévy-Leduc (2010) consider a special case where $p = 1$ and $x_t = 1$ so that the penalty term $\sum_{t=2}^{T}\left\|\beta_t - \beta_{t-1}\right\|$ becomes $\sum_{t=2}^{T}\left|\beta_t - \beta_{t-1}\right|$, the total variation of $\{\beta_t\}$. Note that the objective function in (2.3) is convex in $\{\beta_t\}$. The solution to the convex problem can be computed very fast. Let $\{\hat{\beta}_t = \hat{\beta}_t(\lambda)\}$ denote the solution to the above minimization problem. We frequently suppress the dependence of $\hat{\beta}_t$ on $\lambda$ as long as no confusion arises. Below we will propose a data-driven method to choose $\lambda$.

To see the connection of (2.3) with the group Lasso of Yuan and Lin (2006), we can rewrite (2.1) in an alternative format. Let $\theta_1 = \beta_1$ and $\theta_t = \beta_t - \beta_{t-1}$ for $t = 2, ..., T$. Let $\boldsymbol{\beta} = \left(\beta_1', ..., \beta_T'\right)'$, $\boldsymbol{\theta} = \left(\theta_1', ..., \theta_T'\right)'$, $Y = (y_1, ..., y_T)'$ and $U = (u_1, ..., u_T)'$. Define

$$\underset{T\times Tp}{X} = \begin{bmatrix} x_1' & & & \\ & x_2' & & \\ & & \ddots & \\ & & & x_T' \end{bmatrix}, \underset{Tp\times Tp}{A^*} = \begin{bmatrix} \mathbb{I}_p & & & \\ \mathbb{I}_p & \mathbb{I}_p & & \\ \cdots & \cdots & \ddots & \\ \mathbb{I}_p & \mathbb{I}_p & \mathbb{I}_p & \mathbb{I}_p \end{bmatrix}, \text{ and } \underset{T\times Tp}{X^*} = XA^*.$$

Then (2.1) can be rewritten as $Y = X\boldsymbol{\beta} + U = X^*\boldsymbol{\theta} + U$ and minimizing (2.3) is equivalent to minimizing the following group Lasso criterion function

$$\bar{V}_{T\lambda}\left(\{\theta_t\}\right) = \frac{1}{T}\left\|Y - X^*\boldsymbol{\theta}\right\|^2 + \lambda\sum_{t=2}^{T}\left\|\theta_t\right\| = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\sum_{s=1}^{t}\theta_s\right)^2 + \lambda\sum_{t=2}^{T}\left\|\theta_t\right\|. \tag{2.4}$$

For a given solution $\{\hat{\beta}_t\}$ to (2.3), there exists a block partition $\{\hat{B}_1, ..., \hat{B}_{\hat{m}+1}\}$ of $\{1, 2, ..., T\}$ such that

$$\hat{\beta}_t = \hat{\beta}_s \text{ for all } t, s \in \hat{B}_j = \left[\hat{T}_{j-1}, \hat{T}_j - 1\right] \text{ and } \hat{\beta}_{\hat{T}_j} \neq \hat{\beta}_{\hat{T}_j - 1},\ j = 1, ..., \hat{m} + 1$$

where $\hat{T}_0 = 1$ and $\hat{T}_{\hat{m}+1} = T + 1$. That is, $\hat{m}$ and $\hat{\mathcal{T}}_{\hat{m}} = (\hat{T}_1, ..., \hat{T}_{\hat{m}})$ denote the estimated number of breaks and estimated set of break points, respectively. Given the above block partition, we define $\hat{\alpha}_j = \hat{\alpha}_j(\hat{\mathcal{T}}_{\hat{m}}) = \hat{\beta}_{\hat{T}_j - 1}$ as the estimate of $\alpha_j$ for $j = 1, ..., \hat{m} + 1$. Frequently we suppress the dependence of $\hat{\alpha}_j$ on $\hat{\mathcal{T}}_{\hat{m}}$ (and

$\lambda$) unless necessary. Let $\hat{\boldsymbol{\alpha}}_{\hat{m}} = \hat{\boldsymbol{\alpha}}_{\hat{m}}(\hat{\mathcal{T}}_{\hat{m}}) = (\hat{\alpha}_1(\hat{\mathcal{T}}_{\hat{m}})', ..., \hat{\alpha}_{\hat{m}+1}(\hat{\mathcal{T}}_{\hat{m}})')'$. For any $\boldsymbol{\alpha}_m = \left(\alpha_1', ..., \alpha_{m+1}'\right)'$ and $\mathcal{T}_m = \{T_1, ..., T_m\}$ with $1 < T_1 < \cdots < T_m < T$, we can define

$$Q_{T\lambda}\left(\boldsymbol{\alpha}_m; \mathcal{T}_m\right) = \frac{1}{T}\sum_{j=1}^{m+1}\sum_{t=T_{j-1}}^{T_j-1}\left(y_t - \alpha_j' x_t\right)^2 + \lambda\sum_{j=1}^{m}\left\|\alpha_{j+1} - \alpha_j\right\|. \tag{2.5}$$

Then $Q_{T\lambda}(\hat{\boldsymbol{\alpha}}_{\hat{m}}; \hat{\mathcal{T}}_m) = V_{T\lambda}(\{\hat{\beta}_t\})$.

As we shall show in Theorem 3.3 below, under some weak conditions $P\left(\hat{m} \geq m^0\right) \to 1$ as $T \to \infty$. That is, the estimated number of breaks based on the GFL will be no less than the true number of breaks w.p.a.1. Without further conditions, it is not guaranteed that the GFL will produce the correct number of breaks w.p.a.1. For this reason, we also propose an information criterion that chooses the tuning parameter $\lambda$ in a set of candidate tuning parameters satisfying some basic requirements such that the true number of breaks can be estimated w.p.a.1.

# 3 Asymptotic Properties

In this section we address the statistical properties of the estimation procedure presented in the previous section.

## 3.1 Consistency of the GFL

Let $I_j^0 = T_j^0 - T_{j-1}^0$ for $j = 1, ..., m^0 + 1$. Define

$$I_{\min} = \min_{1\leq j\leq m^0+1}\left|I_j^0\right|,\ J_{\min} = \min_{1\leq j\leq m^0}\left\|\alpha_{j+1}^0 - \alpha_j^0\right\|, \text{ and } J_{\max} = \max_{1\leq j\leq m^0}\left\|\alpha_{j+1}^0 - \alpha_j^0\right\|.$$

Apparently, $I_{\min}$ denotes the minimum interval length among the $m^0 + 1$ regimes, and $J_{\min}$ and $J_{\max}$ denote the minimum and maximum jump sizes, respectively.

To study the consistency of the GFL, we make the following assumptions.

**Assumption A1**. (i)$\{(x_t, u_t),\ t = 1, 2, ...\}$ is a strong mixing process with mixing coefficients $\alpha(\cdot)$ satisfying $\alpha(\tau) \leq c_\alpha \rho^\tau$ for some $c_\alpha > 0$ and $\rho \in (0, 1)$. $E(x_t u_t) = 0$ for each $t$.

(ii) Either one of the following two conditions is satisfied: (a) $\sup_{t\geq 1} E\left\|x_t\right\|^{4q} < \infty$ and $\sup_{t\geq 1} E\left|u_t\right|^{4q} < \infty$ for some $q > 1$; (b) There exist some constants $c_{xx}$ and $c_{xu}$ such that $\sup_{t\geq 1} E[\exp(c_{xx}\left\|x_t\right\|^{2\gamma})] \leq C_{xx} < \infty$ and $\sup_{t\geq 1} E\left[\exp\left(c_{xu}\left\|x_t u_t\right\|^\gamma\right)\right] \leq C_{xu} < \infty$ for some $\gamma \in (0, \infty]$.

**Assumption A2**. (i) There exist two positive constants $\underline{c}_{xx}$ and $\bar{c}_{xx}$ and a positive sequence $\{\delta_T\}$ declining to zero as $T \to \infty$ such that

$$\underline{c}_{xx} \leq \inf_{1\leq s<r\leq T+1, r-s\geq T\delta_T}\mu_{\min}\left(\frac{1}{r-s}\sum_{t=s}^{r-1}E\left(x_t x_t'\right)\right) \leq \sup_{1\leq s<r\leq T+1, r-s\geq T\delta_T}\mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1}E\left(x_t x_t'\right)\right) \leq \bar{c}_{xx}.$$

(ii) $T\delta_T$ satisfies one of the following two conditions: (a) $T\delta_T \geq c_v T^{1/q}$ for some $c_v > 0$ if A1(ii.a) is satisfied; (b) $T\delta_T \geq c_v (\log T)^{(2+\gamma)/\gamma}$ for some $c_v > 0$ if A1(ii.b) is satisfied.

**Assumption A3**. (i) $m^0 = O(\log T)$ and $I_{\min}/(T\delta_T) \to \infty$ as $T \to \infty$.

(ii) $J_{\max} = O(1)$ and $T\delta_T J_{\min}^2/(\log T)^{c_\delta} \to \infty$ as $T \to \infty$ where $c_\delta = 6$ if A1(ii.a) is satisfied and $c_\delta = 1$ if A1(ii.b) is satisfied.

(iii) The tuning parameter $\lambda = \lambda_T$ satisfies $\lambda/(J_{\min}\delta_T) \to 0$ as $T \to \infty$.

(iv) $\frac{Tm^0\{(T^{-1/2}(\log m^0)^{c_\delta/2}+\delta_T)[(\lambda+\delta_T)TI_{\min}^{-1}+I_{\min}^{-1/2}]+T^{-1/2}+\lambda\}}{I_{\min}J_{\min}^2} \to 0$ as $T \to \infty$.

Assumption A1(i) requires that $\{(x_t, u_t)\}$ be a strong mixing process with geometric decay rate. It is satisfied by many well-known processes such as linear autoregressive moving average (ARMA) processes and a large class of processes implied by numerous nonlinear models, including bilinear, nonlinear autoregressive, and autoregressive conditional heteroskedastic (ARCH) type of models. Note that we do not require the error process $\{u_t\}$ to be a martingale difference sequence (m.d.s.) with respect to certain filtration. Let $\mathcal{F}_t = \sigma\text{-field}\{x_{t+1}, u_t, x_t, u_{t-1}, ...\}$. Bai and Perron (1998) specify two sets of conditions for the process $\{(x_t, u_t)\}$ : one requires that it be an $L^a$-mixingale sequence for some $a > 4$ but imposes independence between $x_t$ and $u_s$ for all $t$ and $s$ and thus rules out lagged dependent variables in $x_t$; the other requires that $\{u_t\}$ be an m.d.s. relative to $\mathcal{F}_t$, allowing the presence of lagged dependent variables in $x_t$, but ruling out serial correlation in $\{u_t\}$. In stark contrast, A1(i) allows both lagged dependent variables in $x_t$ and serial correlation and heteroskedasticity in $u_t$. This is important as the model can be dynamically misspecified.

The conditions stated in Assumption A1(ii) pertain to two specific cases related to the moments of $x_t$ and $u_t$. Part (a) in A1(ii) only requires finite moments for them whereas part (b) requires the existence of exponential moments. By Markov inequality, part (b) implies that

$$P\left(\|x_t\|^2 \geq v\right) \leq \exp\left(1 - \left(\frac{v}{K/c_{xx}}\right)^\gamma\right)$$

where $K = \max(1, \log C_{xx})$. That is, the distribution of $\|x_t\|^2$ has to decay exponentially fast. Similar remarks hold for $\|x_t u_t\|$. $\gamma = \infty$ in part (b) corresponds to the case where $\|x_t\|$ and $\|u_t\|$ are uniformly bounded. When combined with A1(i), the conditions in A2(ii) allow us to apply some exponential inequalities for strong mixing processes; see, e.g., Merlevède et al. (2009, 2011).

Assumption A2(i) requires that the sequence $\{E(x_t x_t')\}$ be well behaved. It is automatically satisfied if the process $\{x_t\}$ is covariance-stationary with positive definite covariance matrix. Nevertheless, we do not want to make such a strong assumption because the presence of lagged dependent variables in $x_t$ generally invalidates it when a structural change occurs. In sharp contrast, Assumption A3 in Bai and Perron (1998) requires that the matrix $B_{sr} \equiv \sum_{t=s}^{r-1} x_t x_t'$ be invertible for all $r - s \geq p$. A similar assumption is made in Bai (1998) and Kurozumi (2012), among others. It seems difficult to verify this condition if possible at all. Nevertheless, one can verify that $\frac{1}{r-s}B_{sr}$ is invertible w.p.a.1 under our Assumptions A1-A2 by assuming $r - s$ passes to infinity sufficiently fast. A2(ii) restricts the speed at which $\delta_T$ shrinks to zero. If $x_t$ and $u_t$ only exhibit finite $4q$-th moments for some $q > 1$, then the fastest speed at which $\delta_T \to 0$ is given by $\delta_T \propto T^{(1-q)/q}$. On the other hand, if A1(ii.b) is satisfied, the fastest speed at which $\delta_T \to 0$ is given by $\delta_T \propto (\log T)^{(2+\gamma)/\gamma}/T$, which is further simplified to $(\log T)/T$ if both $x_t$ and $u_t$ are uniformly bounded.

Assumption A3 mainly specifies conditions on $m^0$, $\delta_T$, $I_{\min}$, $J_{\min}$, and $\lambda$. Note that we allow the number of breaks to diverge to infinity slowly and the time intervals in different regimes to diverge to infinity at different rates as $T \to \infty$. This is in sharp contrast with Bai (1998), Bai and Perron (1998), and Kurozumi (2012), who assume that the *fixed* number of multiple break points are asymptotically distinct in the sense that $T_j^0 = \lfloor T\kappa_j^0 \rfloor$ where $0 < \kappa_1^0 < \cdots < \kappa_{m^0}^0 < 1$ and $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. As we shall see, $\delta_T$ will control the rate at which $\hat{T}_j/T$ converges to $T_j^0/T$ when the number of break points are correctly estimated. If one only cares the convergence rate of $\hat{T}_j/T$ to $T_j^0/T$ as in Theorem 3.1 below, A3(i) specifies the *slowest* rate at which $\delta_T$ is allowed to converge to zero: $\delta_T \ll I_{\min}/T$; A3(ii)-(iii) specifies the *fastest* rate at which $\delta_T$ is allowed to converge to zero: $\max\left(\frac{(\log T)^{c_\delta}}{TJ_{\min}^2}, \frac{\lambda}{J_{\min}}\right) \ll \delta_T$. Here $a \ll b$ indicates that $a = o(b)$ as $T \to \infty$. If one also wants to ensure the number of breaks is not underestimated, then the slowest rate for $\delta_T$ to converge to zero, as required by A3(iv), gets reduced: $\delta_T \ll \min(I_{\min}/T, \bar{\delta}_T, I_{\min}^{3/4} J_{\min} T^{-1} (m^0)^{-1/2})$, where $\bar{\delta}_T = I_{\min}^{3/2} J_{\min}^2 (T^2 m^0)^{-1} \min(\lambda^{-1}, T^{1/2} (\log m^0)^{-c_\delta/2})$. In addition, Assumptions A3(i)-(ii) imply that and $I_{\min} J_{\min}^2/(\log T)^{c_\delta} \to \infty$ as $T \to \infty$ and A3(i) and (iii) imply that $\lambda T/(I_{\min} J_{\min}) \to 0$ as $T \to \infty$, which will be frequently used in the proofs of the theorems below.

Admittedly, the conditions in A3 do not appear intuitive due to the generality of our model. In the special case where $I_{\min} \propto T$ (so that $m^0 = O(1)$), the conditions in A3 are reduced to

**Assumption A3**$^*$. As $T \to \infty$, $\delta_T \to 0$, $T\delta_T J_{\min}^2/(\log T)^{c_\delta} \to \infty$, $\lambda/(J_{\min}\delta_T) \to 0$, and $(T^{-1/2} + \lambda + \delta_T^2) J_{\min}^{-2} \to 0$.

If in addition $J_{\min}$ does not shrink to zero as $T \to \infty$ (so that $J_{\min}^{-1} = O(1)$), the conditions in A3(i)-(iii) are reduced to

**Assumption A3**$^{**}$. As $T \to \infty$, $\delta_T \to 0$, $T\delta_T/(\log T)^{c_\delta} \to \infty$, and $\lambda/\delta_T \to 0$.

In this case, A3(iv) becomes redundant given A3(i)-(iii) and we have $\max\left(\frac{(\log T)^{c_\delta}}{T}, \lambda\right) \ll \delta_T \ll 1$, i.e., $\delta_T$ has to converge to zero but at a rate not faster than either $\frac{(\log T)^{c_\delta}}{T}$ or $\lambda$.

The following theorem establishes the consistency of $\{\hat{T}_j\}$ and $\{\hat{\alpha}_j\}$ conditional on the event $\hat{m} = m^0$.

**Theorem 3.1** *Suppose that Assumptions A1-A2 and A3(i)-(iii) hold. If* $\hat{m} = m^0$, *then*
*(i)* $P\left(\max_{1 \le j \le m^0} \left|\hat{T}_j - T_j^0\right| \le T\delta_T\right) \to 1$ *as* $T \to \infty$,
*(ii)* $\hat{\alpha}_j - \alpha_j^0 = O_P\left((I_j^0)^{-1/2} + \lambda T/I_j^0 + \delta_T T/I_j^0\right)$ *for each* $j = 1, ..., m^0 + 1$.

The proof of Theorem 3.1(i) is quite involved. It builds upon some techniques that have been recently developed by Harchaoui and Lévy-Leduc (2010). The latter authors aim at estimating multiple location shifts by assuming independent and identically distributed (IID) errors that have exponential moments. Like Harchaoui and Lévy-Leduc (2010), our analysis is based on a careful inspection of the Karush-Kuhn-Tucker (KKT) optimality conditions for the solutions to the PLS problem in (2.4). Using these optimality conditions and some exponential inequalities for strong mixing processes, we prove Theorem 3.1(i) by contradiction. That is, if $\left|\hat{T}_j - T_j^0\right| \ge T\delta_T$ for some $j = 1, ..., m^0$, we show that w.p.a.1 the solutions will not satisfy all the KKT conditions and therefore cannot be optimal. Extra technicality appears here

because of the presence of regressors that may contain lagged dependent variables, the allowance of only finite $4q$-th moments for $x_t$ and $u_t$, and the allowance of serial dependence and heteroskedasticity in the error process. The proof of part (ii) in Theorem 3.1 simply relies on the result in part (i) and the inspection of the KKT optimality conditions.

Theorem 3.1 suggests that $\max_{1\leq j\leq m^0}\left|\hat{T}_j - T_j^0\right|/T = O\left(\delta_T\right)$, where $\max\left(\frac{(\log T)^{c_\delta}}{TJ_{\min}^2}, \frac{\lambda}{J_{\min}}\right) \ll \delta_T$ as explained above. On the one hand, because $\delta_T = o\left(1\right)$, we have $\left|\hat{T}_j - T_j^0\right|/T = o(1)$, implying that the break ratio $T_j^0/T$ can be consistently estimated. On the other hand, $\max\left(\frac{(\log T)^{c_\delta}}{TJ_{\min}^2}, \frac{\lambda}{J_{\min}}\right) \ll \delta_T$ implies that the fastest convergence rate for the break ratio estimator depends on $\frac{\lambda}{J_{\min}}$ and $\frac{(\log T)^{c_\delta}}{TJ_{\min}^2}$. Here, the first term signifies the effect of the penalty term in the GFL that interacts the minimal break size $J_{\min}$; the second term signifies the effect of moment conditions ($c_\delta = 6$ if the moment condition in Assumption A1(ii.a) is satisfied and 1 if that in Assumption A1(ii.b) is satisfied) and minimal break size. Generally speaking, the smaller the minimal break size is, the slower convergence rate we can achieve for the break ratio estimator; the stronger moment conditions we have, the faster convergence rate the break ratio estimator can have. The result in Theorem 3.1(ii) is intuitive. The first term ($(I_j^0)^{-1/2}$) results from the standard sample convergence as there are essentially $I_j^0$ observations in use for the estimation of $\alpha_j^0$; the second term ($\lambda T/I_j^0$) is derived from the penalty term in the GFL; the third term ($\delta_T T/I_j^0$) is derived from the estimation error of $T_j^0$. If one knows the break dates $\{T_j^0, j = 1, ..., m^0\}$ in advance, then the third term vanishes.

To compare with existing results in the literature, we restrict our attention to the case where $J_{\min}^{-1} = O\left(1\right)$ and $I_{\min} \propto T$ so that Assumption A3** is in effect. We further consider two specific cases that correspond to Assumption A1(ii.a) and A1(ii.b), respectively. In the case where Assumption A1(ii.a) is satisfied, both A2(ii) and A3** are satisfied if one chooses $\delta_T \propto T^{(1-q)/q}$ and $\lambda = \delta_T/\log T$. For small values of $q$, $\delta_T$ may converge to zero at a slower rate than the usual parametric rate $T^{-1/2}$. To ensure $\delta_T = o(T^{-1/2})$ so that the estimation of break dates has no effect on the first order asymptotic distribution of the regression coefficient estimators, we would require that $q > 2$, that is, both $x_t$ and $u_t$ exhibit finite eighth plus moments. In the case where Assumption A1(ii.b) is satisfied, by choosing $\delta_T = (\log T)^{(2+\gamma)/\gamma}/T$ and $\lambda = (\log T)/T$ we can ensure that both A2(ii) and A3** are satisfied. Then we can obtain an almost optimal rate for the estimation of $T_j^0/T$ for $j = 1, ..., m^0$ up to a logarithmic factor since the optimal rate obtained in the literature is of order $T^{-1}$; see, e.g., Bai and Perron (1998). The appearance of the logarithmic factor is due to the application of certain exponential inequality for strong mixing processes. Note that Bai and Perron (1998) make high level assumptions on $B_{sr}$ which are not directly verifiable and their proof does not rely on any exponential inequality. In the following we show that as long as $\hat{m} = m^0$ in large samples, the above convergence rates for the estimates of break dates can be improved. See the last paragraph in Section 3.3.

Unfortunately, the correct number $m^0$ of break points may be unknown. However, if we follow the literature (e.g., Bai and Perron (1998)) and assume that the true number of breaks is bounded by a number $m_{\max}$ with $m_{\max} \leq C\log T$ for a large number $C$, then we can show that for any single true break date $T_j^0 \in \mathcal{T}^0$, there exists an estimated break date in $\hat{\mathcal{T}}_{\hat{m}}$ that is sufficiently close to $T_j^0$ as long as $\hat{m} \geq m^0$.

In addition, under the extra conditions on $m^0$, $\lambda$, $I_{\min}$, $J_{\min}$, and $\delta_T$ detailed in Assumption A3(iv), we can ensure that the last condition is satisfied w.p.a.1. That is, the probability of under-estimating the number of true break points converges to zero as $T \to \infty$.

To proceed, let $\mathcal{D}(A,B) \equiv \sup_{b\in B} \inf_{a\in A} |a-b|$ for any two sets $A$ and $B$. Note that $\max\{\mathcal{D}(A,B), \mathcal{D}(B,A)\}$ denotes the Hausdorff distance between $A$ and $B$. The following theorem indicates all true break points in $\mathcal{T}^0$ can be "consistently" estimated by some points in $\hat{\mathcal{T}}_{\hat{m}}$ under the assumption that the estimated number of breaks is no less than the true number of breaks.

**Theorem 3.2** *Suppose that Assumptions A1-A2 and A3(i)-(iii) hold. If* $m^0 \leq \hat{m} \leq m_{\max}$, *then* $P(\mathcal{D}(\hat{\mathcal{T}}_{\hat{m}}, \mathcal{T}^0) \leq T\delta_T) \to 1$ *as* $T \to \infty$.

The proof of the above theorem is also done by contradiction and by the repeated utilization of the KKT optimality conditions under the same set of Assumptions required for Theorem 3.1. Theorem 3.2 assures us that even if the number of breaks is overestimated, there will be an estimated break date close to each unknown true break date.

The next theorem shows that $\hat{m}$ cannot be smaller than $m^0$ in large samples provided Assumption A3(iv) is also satisfied.

**Theorem 3.3** *Suppose that Assumptions A1-A3 hold. Then* $P\left(\hat{m} < m^0\right) \to 0$ *as* $T \to \infty$.

Theorem 3.3 implies that the probability of under-estimating the number of break points is asymptotically negligible.

## 3.2 Choosing the tuning parameter $\lambda$

Let $\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{\hat{m}_\lambda}}(\lambda) = (\hat{\alpha}_{1,\hat{\mathcal{T}}_{\hat{m}_\lambda}}(\lambda)', ..., \hat{\alpha}_{\hat{m}_\lambda+1,\hat{\mathcal{T}}_{\hat{m}_\lambda}}(\lambda)')'$ denote the set of post-Lasso OLS estimates of the regression coefficients based on the break dates in $\hat{\mathcal{T}}_{\hat{m}_\lambda} = \hat{\mathcal{T}}_{\hat{m}_\lambda}(\lambda)$, where we make the dependence of various estimates on $\lambda$ explicit. Let $\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}} \equiv Q_{T,1}(\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{\hat{m}_\lambda}}(\lambda), \hat{\mathcal{T}}_{\hat{m}_\lambda})$, where

$$Q_{T,1}(\boldsymbol{\alpha}_m; \mathcal{T}_m) \equiv \frac{1}{T}\sum_{j=1}^{m+1}\sum_{t=T_{j-1}}^{T_j-1}\left(y_t - \alpha_j' x_t\right)^2, \tag{3.1}$$

is the first term in the definition of $Q_{T\lambda}(\boldsymbol{\alpha}_m; \mathcal{T}_m)$ in (2.5). We propose to select the tuning parameter $\lambda$ by minimizing the following information criterion:

$$IC(\lambda) = \log(\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}}) + \rho_T p(\hat{m}_\lambda + 1). \tag{3.2}$$

Without any condition on $\lambda$, we are unable to study the asymptotic properties of $\hat{m}_\lambda$, $\hat{\mathcal{T}}_{\hat{m}_\lambda}$, and $\hat{\alpha}_j$ for $j = 1, ..., 1+\hat{m}_\lambda$. For this reason, we restrict our attention to the class of tuning parameters such that Assumptions A3(iii)-(iv) are satisfied.

To state the next result, we add the following assumption.

**Assumption A4**. (i) $m^0(I_{\min}^{-1/2} + \delta_T)\left[1 + T^{-1/2}(\log m^0)^{c_\delta/2}\delta_T^{-1}\right] = O(1)$ and $\delta_T I_{\min} \to \bar{c}_\delta \in (0, \infty]$ as $T \to \infty$.

(ii) $\left(1+\frac{T}{I_{\min}J_{\min}^2}\right)\rho_T m^0 \to 0$ and $\delta_T^{-1}\rho_T \to \infty$ as $T \to \infty$.

Assumption A4(i) imposes an extra condition on $\delta_T$ and it becomes redundant under Assumption A2 if $I_{\min} \propto T$. Assumption A4(ii) reflects the usual conditions for the consistency of model selection, that is, the penalty coefficient $\rho_T$ cannot shrink to zero either too fast or too slowly. If $I_{\min} \propto T$ and $J_{\min}^{-1} = O(1)$, the first part of A4(ii) requires that $\rho_T \to 0$, which is standard for an information-criterion function. $\delta_T$ indicates the probability order of the distance between the first term in the criterion function for an over-parametrized model and that for the true model.

**Theorem 3.4** *Suppose that Assumptions A1-A4 hold. Let* $\hat{\lambda} = \arg\min_\lambda IC(\lambda)$. *Then* $P\left(\hat{m}_{\hat{\lambda}} = m^0\right) \to 1$ *as* $T \to \infty$.

The proof of Theorem 3.4 in Appendix E suggests that the $\lambda$'s that yield the over-estimated or under-estimated number of breaks fail to minimize the information criterion w.p.a.1 provided that the minimization is restricted for a class of tuning parameters that satisfy some basic requirements stated in Assumptions A3(iii)-(iv). Consequently, the minimizer of $IC(\lambda)$ can only be the one that produces the correct number of estimated breaks in large samples. Conditional on $\hat{m}_\lambda = m^0$, we will study the asymptotic distributions of the Lasso estimates of regression coefficients and break dates below.

## 3.3 Limiting distributions of the Lasso estimates of regression coefficients and break dates

In this subsection we analyze the consistency of the regression coefficient estimates and break fraction estimates. We let $\hat{\kappa} = (\hat{\kappa}_1, ..., \hat{\kappa}_{m^0}) = (\hat{T}_1/T, ..., \hat{T}_{m^0}/T)$ with corresponding true values $\kappa^0 = (\kappa_1^0, ..., k_{m^0}^0) = (T_1^0/T, ..., T_{m^0}^0/T)$. Note that we allow $\kappa_j^0 - \kappa_{j-1}^0 = 0$ for some $j = 1, ..., m^0+1$, which occurs if $I_{\min} = o(T)$.

It is well known that the limiting distributions of the break date estimators obtained by specifying fixed magnitude of changes are dependent on the exact distribution of $\{x_t, u_t\}$. It is useful to consider asymptotic distributions under shrinking magnitude of changes. Now, $\alpha_j^0$'s is $T$-dependent and we frequently write $\alpha_{T,j}^0$ for $\alpha_j^0$ when we want to emphasize the dependence of $\alpha_j^0$'s on $T$. Let $d_{T,j}^0 = \alpha_{T,j+1}^0 - \alpha_{T,j}^0$ for $j = 1, ..., m^0$. The required conditions are stated in the next assumption.

**Assumption A5**. (i) For $j = 1, ..., m^0$, $d_{T,j}^0 = \bar{d}_{T,j}\Delta_j$ for some $\Delta_j$ independent of $T$ and $\bar{d}_{T,j} > 0$ is a scalar satisfying $\bar{d}_{T,j} \to 0$ and $T^{(1/2)-\vartheta_j}\bar{d}_{T,j} \to \infty$ for some $\vartheta_j \in (0, 1/2)$.

(ii) For $j = 1, ..., m^0+1$, as $I_j^0 \to \infty$, $(I_j^0)^{-1}\sum_{t=T_{j-1}^0}^{T_{j-1}^0+\lfloor sI_j^0 \rfloor} E(x_t x_t') \to s\Psi_j$ and $(I_j^0)^{-1}\sum_{t=T_{j-1}^0}^{T_{j-1}^0+\lfloor sI_j^0 \rfloor}\sum_{s=T_{j-1}^0}^{T_{j-1}^0+\lfloor sI_j^0 \rfloor} E(x_t x_s' u_t u_s) \to s\Phi_j$ uniformly in $s$.

**Assumption A6**. $m^0 T\lambda I_{\min}^{-1/2} \to 0$ as $T \to \infty$.

Assumption A5(i) specifies the magnitude of each break size: the smaller value of $\vartheta_j$, the smaller magnitude of the break size could be. Note that we allow different breaks to shrink to zero at different speeds. A5(ii) specifies the asymptotic average behavior of $E(x_t x_t')$ and $E(x_t x_s' u_t u_s)$ within each regime. In conjunction with Assumption A1, the first and second parts of A5(ii) ensure that $(I_j^0)^{-1}\sum_{t=T_{j-1}^0}^{T_{j-1}^0+\lfloor sI_j^0 \rfloor} x_t x_t' \xrightarrow{P} s\Psi_j$ and $(I_j^0)^{-1}\sum_{t=T_{j-1}^0}^{T_{j-1}^0+\lfloor sI_j^0 \rfloor} x_t u_t \Rightarrow B_j(s)$, respectively, by the uniform law

of large numbers and invariance principle for heterogeneous strong mixing processes, where $B_j(\cdot)$ is a multivariate Gaussian process on $[0,1]$ with mean 0 and covariance kernel $E[B_i(s)B_j(r)] = \min(s,r)\Phi_j$. See White (2001). A6 imposes some side condition on $\lambda$ to ensure the penalty term in the Lasso procedure does not affect the usual $(I_j^0)^{1/2}$-consistency of the Lasso estimator of $\alpha_j^0$.

Let $D_{m^0}$ =diag$((I_1^0)^{1/2}\mathbb{I}_p, ..., (I_{m^0+1}^0)^{1/2}\mathbb{I}_p)$, $Y = (y_1, ..., y_T)'$, $U = (u_1, ..., u_T)'$, and $\mathbb{X}$ =diag$(\mathbb{X}_1, ..., \mathbb{X}_{m^0+1})$, where $\mathbb{X}_j = (x_{T_{j-1}^0}, ..., x_{T_j^0-1})'$ for $j = 1, ..., m^0+1$. Let $\mathbf{\Psi} \equiv \text{plim}\, D_{m^0}^{-1}\mathbb{X}'\mathbb{X}D_{m^0}^{-1}$ and $\mathbf{\Phi} \equiv \text{plim}\, D_{m^0}^{-1}\mathbb{X}'UU'\mathbb{X}D_{m^0}^{-1}$. Note that both $\Psi$ and $\Phi$ are well behaved under Assumptions A1 and A5(ii). For $j = 1, ..., m^0$, define $\xi_j = \Delta_j'\Psi_{j+1}\Delta_j/\Delta_j'\Psi_j\Delta_j$, $\phi_{j,1} = \{\Delta_j'\Phi_j\Delta_j/\Delta_j'\Psi_j\Delta_j\}^{1/2}$, $\phi_{j,2} = \{\Delta_j'\Phi_{j+1}\Delta_j/\Delta_j'\Psi_{j+1}\Delta_j\}^{1/2}$, and let $W_{j,1}(s)$ and $W_{j,2}(s)$ be independent Wiener processes that are defined on $[0,\infty)$ with $W_{j,1}(0) = W_{j,2}(0) = 0$ and independent across $j$. Define

$$Z_j(s) = \begin{cases} \phi_{j,1}W_{j,1}(-s) - |s|/2 & \text{if } s < 0 \\ \sqrt{\xi_j}\phi_{j,2}W_{j,2}(s) - \xi_j|s|/2 & \text{if } s > 0 \end{cases} \quad \text{for } j = 1, ..., m^0.$$

The following theorem reports the asymptotic distributions of the Lasso estimators.

**Theorem 3.5** *Suppose that Assumptions A1-A6 hold. Let $S$ denote an $L \times p(m^0+1)$ selection matrix such that $\|S\|$ is finite, where $L \in [1, p(m^0+1)]$ is a fixed integer. Then*

*(i)* $SD_{m^0}(\hat{\boldsymbol{\alpha}}_{m^0} - \boldsymbol{\alpha}^0) \xrightarrow{d} N\left(0, S\mathbf{\Psi}^{-1}\mathbf{\Phi}\mathbf{\Psi}^{-1}S'\right)$;

*(ii)* $(\Delta_j'\Psi_j\Delta_j)\bar{d}_{T,j}^2 T\left(\hat{\kappa}_j - \kappa_j^0\right) \xrightarrow{d} \arg\max_s Z_j(s)$ *for $j = 1, ..., m^0$, and $\hat{\kappa}_j$'s are asymptotically mutually independent of each other.*

The above theorem lays down the foundation for inferences on the unknown regression coefficients and break fractions based on the GFL. Note that we specify a selection matrix $S$ in Theorem 3.5(i) that is not needed if $m^0$ is fixed. Intuitively, we allow the number of breaks, $m^0$, to diverge to infinity as the sample size $T$ passes to infinity. For this reason, the dimension of $\hat{\boldsymbol{\alpha}}_{m^0}$ is also divergent to infinity at the rate $m^0$ and we cannot derive the asymptotic normality of $\hat{\boldsymbol{\alpha}}_{m^0}$. Instead, we follow the literature on inferences with a diverging number of parameters (see, e.g., Fan and Peng (2004), Lam and Fan (2008), Lu and Su (2014)) and prove the asymptotic normality for any arbitrary linear combinations of elements of $\hat{\boldsymbol{\alpha}}_{m^0}$ after adapting to different convergence rates for different subvectors of $\hat{\boldsymbol{\alpha}}_{m^0}(\equiv (\hat{\alpha}_1', ..., \hat{\alpha}_{m^0+1}')')$. In the special case where $m^0$ is fixed, we can take $S = \mathbb{I}_{p(m^0+1)}$ and obtain the usual joint asymptotic normality of $\hat{\alpha}_j$'s. Alternatively, if we assume that $\{x_t u_t\}$ is an m.d.s., then like $\mathbf{\Psi}$, $\mathbf{\Phi}$ is also block diagonal and $\hat{\alpha}_j$'s are asymptotically mutually independent of each other. In this case, it suffices to report the asymptotic normality of $\hat{\alpha}_j$ for $j = 1, ..., m^0+1$. Interestingly, Theorem 3.5(ii) suggests that $\hat{\kappa}_j$'s are asymptotically mutually independent of each other even in the absence of any m.d.s. condition for $\{x_t u_t\}$.

A close examination of the proof of the above theorem suggests that the GFL estimators of the regression coefficients and break dates are closely tied with Bai and Perron's (1998) OLS estimators. If the number of breaks $m^0$ were known, one could obtain the GFL estimator by minimizing the following PLS objective function

$$S_{T\lambda}(\boldsymbol{\alpha}, \mathcal{T}_m) = \frac{1}{T}\sum_{j=1}^{m_0+1}\sum_{t=T_{j-1}}^{T_j-1}(y_t - x_t'\alpha_j)^2 + \lambda\sum_{j=1}^{m_0}\|\alpha_{j+1} - \alpha_j\|,$$

where the first term is the usual OLS objective function with $m^0$ unknown breaks and the second term is a penalty term. As expected, for sufficiently small $\lambda$, the solution to the above problem will share the same asymptotic distribution as that of Bai and Perron's estimator. When $m^0$ is unknown but can be estimated correctly by $\hat{m}$ w.p.a.1, we can treat $\hat{m}$ as $m^0$ to infer the above asymptotic result.

Given the result in Theorem 3.5(i), it is standard to estimate the asymptotic variance-covariance matrix and make inference on $\alpha^0$. In particular, one can obtain a HAC estimator for $\Phi$ to allow for both heteroskedasticity and serial correlation. Let $\hat{D}_{m^0}$ =diag$(\hat{I}_1^{1/2}\mathbb{I}_p, ..., \hat{I}_{m^0+1}^{1/2}\mathbb{I}_p)$ where $\hat{I}_j = \hat{T}_j - \hat{T}_{j-1}$ for $j = 1, ..., m^0+1$, $\hat{T}_0 = 1$, and $\hat{T}_{m^0+1} = T+1$. One can replace $D_{m^0}$ by $\hat{D}_{m^0}$ in the above theorem. Theorem 3.5(ii) indicates that the limiting distribution of the break fraction estimates is the same as that occurring in a single break model. As Bai and Perron (1998) remark, if $\Psi_j$ and $\Phi_j$ are the same for adjacent $j$'s and are given by $\Psi$ and $\Phi$, respectively, then we have the standard asymptotic pivotal limiting distribution for $\hat{\kappa}$ after normalization:

$$\frac{(\Delta_j'\Psi\Delta_j)^2}{\Delta_j'\Phi\Delta_j}\bar{d}_{T,j}^2 T\left(\hat{\kappa}-\kappa_j^0\right) \xrightarrow{d} \arg\max_s \{W_j(s) - |s|/2\}$$

where $W_j(s) = W_{j,1}(-s)$ for $s \leq 0$ and $W_j(s) = W_{j,2}(s)$ for $s > 0$. One can apply this result to construct confidence intervals for $\kappa_j^0$ or equivalently, $T_j^0$. See, e.g., Bai (1997a) and Su et al. (2013). We omit the details for brevity.

Theorem 3.5(ii), in conjunction with Assumption A3(ii), indicates that in the case of small breaks

$$\hat{T}_j - T_j^0 = O_P\left(\bar{d}_{T,j}^{-2}\right) = O_P(J_{\min}^{-2}) = o_P(T\delta_T),$$

which suggests an improved rate than that obtained in Theorem 3.1. For the fixed magnitude of breaks, although there is no way to obtain any asymptotic pivotal distribution for the break fraction estimates even after normalization, we can obtain $\hat{T}_j - T_j^0 = O_P(1) = o_P(T\delta_T)$, using similar arguments in the proof of Theorem 3.5. In either case, we can obtain the optimal rate of convergence for the estimation of the break dates provided that $\hat{m}_\lambda = m^0$ is ensured by a proper choice of the tuning parameter $\lambda$.

## 3.4 Limiting distribution of post-Lasso estimate of regression coefficients

In this subsection we study the asymptotic distribution of the post-Lasso estimate $\hat{\alpha}_{j,\hat{\mathcal{T}}_{m^0}}$. Let $\hat{\mathbb{X}}$ =diag$(\hat{\mathbb{X}}_1$, ..., $\hat{\mathbb{X}}_{m^0+1})$ where $\hat{\mathbb{X}}_j = (x_{\hat{T}_{j-1}}, ..., x_{\hat{T}_j-1})'$. We can write the DGP in matrix form

$$Y = \mathbb{X}\alpha^0 + U. \tag{3.3}$$

The model used for the post-Lasso estimation of $\alpha^0$ is given by

$$Y = \hat{\mathbb{X}}\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{m^0}} + \hat{U}, \tag{3.4}$$

where $\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{m^0}} = (\hat{\mathbb{X}}'\hat{\mathbb{X}})^{-1}\hat{\mathbb{X}}'Y$, and $\hat{U}$ is a $T\times 1$ vector of the post-Lasso residuals. The following assumption is needed for the analysis of the limiting distribution of $\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{m^0}}$.

**Assumption A7**. $m^0 T\delta_T I_{\min}^{-1/2} \to 0$ as $T\to\infty$.

Assumption A7 ensures that the estimation of the break dates has asymptotically negligible effect on the asymptotic distribution of the post-Lasso estimate of the regression coefficients. In the special case $I_{\min} \propto T$, Assumption A7 is satisfied as long as $\delta_T = o(T^{-1/2})$. In this case, Assumption A2(ii) indicates that we only need $x_t$ and $u_t$ to exhibit finite eighth plus moments. In the general case, the minimum interval length has crucial effect on the rate at which $\delta_T$ shrinks to zero. For example, if $I_{\min} \propto T^{1/2}$, $\delta_T$ has to converge to zero at a rate faster than $T^{-3/4}$, which, according to Assumption A2(ii), would in turn require that $x_t$ and $u_t$ exhibit finite sixteenth plus moments.

The following theorem reports the limiting distribution of $\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{I}}_{m^0}}$.

**Theorem 3.6** *Suppose that Assumptions A1-A4 and A7 hold. Let $S$ be defined as in Theorem 3.5. Then* $S\hat{D}_{m^0}(\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{I}}_{m^0}} - \boldsymbol{\alpha}^0) \xrightarrow{d} N\left(0, S\boldsymbol{\Psi}^{-1}\boldsymbol{\Phi}\boldsymbol{\Psi}^{-1}S'\right)$.

Note that Assumptions A5-A6 are not required for the above theorem. Define the infeasible estimator $\hat{\boldsymbol{\alpha}}_{\mathcal{I}^0_{m^0}} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y$. We can prove the theorem by showing that $S\hat{D}_{m^0}(\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{I}}_{m^0}} - \boldsymbol{\alpha}^0)$ shares the same asymptotic distribution as $SD_{m^0}(\hat{\boldsymbol{\alpha}}_{\mathcal{I}^0_{m^0}} - \boldsymbol{\alpha}^0)$. Similar idea was used by Bai (1997a) for the case of a single structural break. Extra care is needed as we allow the interval length to be different across different regimes and $m^0$ to be divergent. Given the above result, it is standard to make inference on $\boldsymbol{\alpha}^0$ based on the post Lasso estimate $\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{I}}_{m^0}}$.

As a referee kindly points out, the asymptotic distribution of the post-Lasso estimator is only valid pointwise and it does not provide uniformly valid inference for the regression coefficients; see Pötscher and Leeb (2009) and Pötscher and Schneider (2009). In particular, this limiting distribution ignores the randomness of the estimated number of breaks in finite samples. As a result, a robust inference procedure with correct asymptotic size is an important issue for the post-Lasso estimator; see, e.g., Belloni et al. (2014). This is closely related to the post model selection inference problem investigated by Leeb and Pötscher (2005, 2008), among others. Robust inference on the parameter of interest is beyond the scope of this paper.

# 4 Monte Carlo Simulations

In this section we conduct a set of Monte Carlo experiments to evaluate the finite sample performance of our GFL method. Throughout we use the block-coordinate descent algorithm (Angelosante and Giannakis, 2012) to solve the minimization problem in (2.3).[1] We select the penalty term $\lambda$ that minimizes the information criterion $IC(\lambda)$ by setting $\rho_T = 1/\sqrt{T}$ (c.f. Bai (1998)).[2] It is well known that there exists a $\lambda^{\max}$ such that any $\lambda \geq \lambda^{\max}$ will produce constant coefficients (i.e., no break) (Ohlsson et al., 2010). We thus search for a minimal IC on 20 evenly-distributed logarithmic grids on the interval $[0.01\lambda^{\max},\ \lambda^{\max}]$. Finally, to purge unwanted breaks, we employ a post-processing procedure similar to that used by Harchaoui and Lévy-Leduc (2010).

The main competitor of our approach is Bai and Perron (1998, 2003a, BP hereafter). We consider different trimming proportions (tr) for BP, namely, tr = 0.05, 0.1, 0.15, and 0.2. It should be noted that the comparison presented here is inevitably inconclusive. As shown in Bai and Perron (2006) and

in this section, the performance of BP is crucially dependent on the choice of trimming. For some of the data generate processes (DGPs) experimented here, which have either no break or only a small number of breaks in the middle range of the data, BP's tests with large trimmings generally give satisfactory performance. However, large trimming is an implicit assumption on the nature of the DGP. For example, a trimming of 0.2 implicitly assumes that the maximum number of breaks is 4 and that the break cannot happen in partitions at the beginning or in the end (each with a length of 20% of the sample). The assumption may be too restrictive for some applications. Small trimming can afford more breaks in the DGP but tend to overestimate the number of breaks. The size of trimming, indeed, plays a similar role as the penalty term in our approach.

## 4.1 The Case of No Break

We first evaluate the probability of falsely detecting breaks when no break exists. We consider the following DGPs

$$y_t = 1 + x_t + u_t,$$

with

- DGP-1: $x_t \sim i.i.d.\ N(0,1)$, $u_t \sim i.i.d.\ N(0,\sigma_u^2)$.
- DGP-2: $x_t \sim \text{AR}(1)$, $u_t \sim i.i.d.\ N(0,\sigma_u^2)$.
- GDP-3: $x_t \sim i.i.d.\ N(0,1)$, $u_t = \sigma_u v_t$, $v = 0.5v_{t-1} + \epsilon_t$, $\epsilon_t \sim i.i.d.\ N(0,0.75)$.
- DGP-4: $x_t \sim \text{AR}(1)$, $u_t = \sigma_u\sqrt{h_t}\epsilon_t$, $h_t = 0.05 + 0.05u_{t-1}^2 + 0.9h_{t-1}$, $\epsilon_t \sim i.i.d.\ N(0,1)$.
- DGP-5: $x_t \sim \text{AR}(1)$, $u_t \sim i.i.d.\ N(0,\sigma_1^2)$ for $t \in \{1,2,\ldots,T/2\}$ and $u_t \sim i.i.d.\ N(0,\sigma_2^2)$ for $t \in \{T/2, T/2+1,\ldots,T\}$.
- DGP-6: $y_t = ay_{t-1} + \epsilon_t$, $x_t = y_{t-1}$, $\epsilon_t \sim i.i.d.\ N(0,1-a^2)$.

DGP-1 is the basic benchmark. DGP-2 introduces serial correlation in $x_t$. Specifically, we generate $x_t$ by an AR(1) dynamics: $x_t = 0.5x_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim i.i.d.\ N(0,0.75)$, so that $x_t$ has unit variance. DGP-3 introduces serial correlation in $u_t$. DGP-4 introduces conditional heteroscedasticity (volatility clustering) in the error. DGP-5 considers heterogeneity in variance in the error. Finally, DGP-6 is an AR regression where $x_t$ is the lagged value of $y_t$. To evaluate the performance under different noise levels, we select the parameter $\sigma_u$ in DGP-1, DGP-2, DGP-3, and GDP-4 to be 0.5, 1, and 1.5. For the benchmark case, $\sigma_u = 1$ corresponds to a unit signal-to-noise ratio. In DGP-5, we set $\sigma_1 = 0.1$ and $\sigma_2 = 0.2$, 0.3, or 0.5. In essence, there is a regime shift in the variance of the residual. In DGP-6, the autoregressive coefficient $a$ is chosen from $\{0.2, 0.5, 0.9\}$. We compare our approach (GFL) with weighted double maximum tests (WDMax) and its robust version developed in BP with a theoretical size of 5%.[3] The robust version allows for heteroscedasticity and autocorrelation in the error. The results are summarized in Table 1, where we report proportions of false detections among 500 repetitions for each method.

In the benchmark case of DGP-1, our method (GFL) produces negligible percentages of false detection of breaks for all noise levels. The same is true for DGP-2 and DGP-4, where $x_t$ is endowed with serial correlation and severe conditional heteroscedasticity, respectively. However, when serial correlation is introduced in $u_t$ (DGP-3), there are sizable proportions of false detections when $T = 100$. As $T$ gets larger, the percentages of false detections quickly decline to nearly zero. When there is a moderate regime shift in the variance of the error process (DGP-5), the performance of our method is close to the benchmark case. So is true for the case of autoregression (DGP-6). Overall, we may conclude that GFL enjoys a low probability of falsely detecting breaks when none exists.

In comparison, the performance of WDMax and its robust version depends crucially on the choice of trimming. In most cases, the empirical sizes corresponding to tr=0.05 are substantially higher than 5%, the theoretical size we set. Except for DGP-3, DGP-4 in the case of $\sigma_u = 0.5$, and DGP-5 in the case of $\sigma_2 = 0.2$, empirical sizes of WDMax corresponding to tr=0.15 or 0.2 are reasonably close to 5%, especially when $T$ is large. In DGP-3, where the error is serially correlated, WDMax breaks down as expected, while the robust WDMax produces reasonable empirical sizes only when both trimming and sample size are large, as is true for other DGP's. The general under-performance of the robust WDMax may be understood by noting that the sample covariances for the robust correction need to be estimated from very small samples (say ten observations for the case where tr = 0.1 and $T = 100$).

Table 1: Proportions of False Detection When $m = 0$ (All figures are percentages (%) of false detection of breaks when there are none. )

| DGP | $\sigma_u$ | $T$ | GFL | WDMax | | | | robust WDMax | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | tr=.05 | .10 | .15 | .20 | .05 | .10 | .15 | .20 |
| 1 | 0.5 | 100 | 0.0 | 56.0 | 15.8 | 7.4 | 5.2 | 100.0 | 70.2 | 30.0 | 16.2 |
| | | 200 | 0.0 | 15.8 | 4.6 | 3.6 | 4.2 | 84.2 | 25.4 | 10.8 | 7.0 |
| | | 500 | 0.0 | 4.8 | 2.6 | 2.6 | 3.6 | 29.4 | 9.0 | 6.8 | 5.0 |
| | 1.0 | 100 | 0.0 | 57.4 | 12.4 | 6.4 | 5.2 | 100.0 | 65.6 | 32.0 | 16.8 |
| | | 200 | 0.0 | 23.2 | 5.8 | 4.4 | 3.4 | 88.0 | 29.4 | 17.0 | 10.4 |
| | | 500 | 0.0 | 6.4 | 4.2 | 5.0 | 4.8 | 29.2 | 9.8 | 8.2 | 6.0 |
| | 1.5 | 100 | 0.2 | 60.2 | 13.6 | 5.4 | 4.6 | 100.0 | 67.6 | 30.8 | 16.2 |
| | | 200 | 0.0 | 17.2 | 4.8 | 4.4 | 3.6 | 82.4 | 27.8 | 12.6 | 8.8 |
| | | 500 | 0.0 | 7.0 | 4.0 | 4.8 | 4.8 | 30.2 | 12.0 | 7.2 | 6.2 |
| 2 | 0.5 | 100 | 0.0 | 58.6 | 14.0 | 6.2 | 4.2 | 100.0 | 75.8 | 35.0 | 18.6 |
| | | 200 | 0.0 | 20.4 | 4.6 | 3.8 | 2.8 | 86.0 | 29.0 | 13.8 | 8.0 |
| | | 500 | 0.0 | 7.6 | 4.8 | 4.6 | 5.0 | 34.0 | 12.4 | 9.4 | 6.8 |
| | 1.0 | 100 | 0.0 | 59.4 | 12.6 | 6.8 | 5.4 | 99.6 | 74.6 | 38.6 | 18.0 |
| | | 200 | 0.0 | 20.6 | 6.2 | 4.0 | 3.0 | 86.4 | 29.2 | 15.4 | 9.0 |
| | | 500 | 0.0 | 6.2 | 3.0 | 3.2 | 3.4 | 35.8 | 10.8 | 8.4 | 7.2 |
| | 1.5 | 100 | 0.2 | 63.0 | 14.0 | 7.4 | 5.8 | 100.0 | 77.0 | 35.2 | 18.8 |
| | | 200 | 0.2 | 20.0 | 6.4 | 5.4 | 3.6 | 87.0 | 31.6 | 15.8 | 9.0 |
| | | 500 | 0.0 | 5.4 | 3.6 | 3.6 | 4.2 | 32.2 | 11.2 | 7.0 | 6.0 |
| | | 100 | 12.2 | 97.2 | 75.0 | 59.6 | 49.4 | 100.0 | 88.6 | 48.4 | 25.6 |

Continued on next page

Table 1: Continued

| DGP | $\sigma_u$ | $T$ | GFL | WDMax | | | | robust WDMax | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | tr=.05 | .10 | .15 | .20 | .05 | .10 | .15 | .20 |
| 3 | 0.5 | 200 | 3.2 | 90.8 | 67.8 | 54.4 | 45.2 | 94.4 | 39.6 | 16.6 | 10.2 |
| | | 500 | 0.0 | 85.6 | 64.2 | 54.4 | 47.8 | 38.4 | 11.2 | 7.8 | 6.2 |
| | 1.0 | 100 | 11.6 | 97.4 | 73.4 | 55.6 | 45.2 | 100.0 | 88.6 | 44.2 | 23.6 |
| | | 200 | 2.4 | 91.0 | 66.4 | 49.6 | 41.8 | 94.4 | 38.6 | 18.2 | 10.8 |
| | | 500 | 0.2 | 87.8 | 61.8 | 51.8 | 45.6 | 38.6 | 12.0 | 8.8 | 7.0 |
| | 1.5 | 100 | 11.8 | 98.0 | 78.2 | 61.8 | 49.4 | 100.0 | 87.4 | 46.6 | 27.6 |
| | | 200 | 2.8 | 92.8 | 68.2 | 53.8 | 45.0 | 95.0 | 38.2 | 17.4 | 10.8 |
| | | 500 | 0.2 | 84.6 | 60.8 | 52.6 | 42.8 | 38.8 | 10.6 | 9.0 | 5.8 |
| 4 | 0.5 | 100 | 0.2 | 86.2 | 43.0 | 27.0 | 19.6 | 100.0 | 99.4 | 75.2 | 43.2 |
| | | 200 | 0.0 | 60.2 | 27.8 | 21.4 | 17.0 | 99.8 | 81.6 | 46.2 | 28.2 |
| | | 500 | 0.0 | 29.6 | 16.0 | 12.2 | 11.0 | 86.2 | 41.0 | 21.2 | 14.0 |
| | 1.0 | 100 | 1.4 | 55.8 | 11.2 | 6.6 | 5.2 | 100.0 | 74.2 | 40.8 | 23.4 |
| | | 200 | 0.2 | 14.4 | 5.4 | 4.0 | 3.2 | 87.0 | 29.4 | 13.2 | 8.8 |
| | | 500 | 0.0 | 5.0 | 3.2 | 3.8 | 4.2 | 26.0 | 7.8 | 4.8 | 4.0 |
| | 1.5 | 100 | 0.4 | 53.0 | 10.6 | 4.4 | 3.4 | 100.0 | 79.4 | 39.0 | 18.6 |
| | | 200 | 0.0 | 18.0 | 5.2 | 4.4 | 3.2 | 89.4 | 29.4 | 13.0 | 9.0 |
| | | 500 | 0.0 | 5.6 | 2.6 | 3.4 | 4.0 | 31.6 | 8.6 | 4.6 | 4.0 |
| 5 | $\sigma_2 = .2$ | 100 | 0.2 | 86.2 | 43.0 | 27.0 | 19.6 | 100.0 | 99.4 | 75.2 | 43.2 |
| | | 200 | 0.0 | 60.2 | 27.8 | 21.4 | 17.0 | 99.8 | 81.6 | 46.2 | 28.2 |
| | | 500 | 0.0 | 29.6 | 16.0 | 12.2 | 11.0 | 86.2 | 41.0 | 21.2 | 14.0 |
| | $\sigma_2 = .3$ | 100 | 0.0 | 60.6 | 17.4 | 7.2 | 5.4 | 100.0 | 85.6 | 42.4 | 24.0 |
| | | 200 | 0.0 | 27.0 | 9.6 | 6.0 | 5.6 | 93.6 | 40.2 | 22.0 | 13.6 |
| | | 500 | 0.0 | 10.8 | 6.0 | 5.4 | 5.0 | 51.2 | 13.2 | 7.2 | 5.4 |
| | $\sigma_2 = .5$ | 100 | 0.6 | 44.0 | 10.6 | 6.0 | 3.2 | 100.0 | 70.6 | 32.0 | 18.2 |
| | | 200 | 0.0 | 23.8 | 7.0 | 5.6 | 5.2 | 89.2 | 32.0 | 17.4 | 13.6 |
| | | 500 | 0.0 | 6.4 | 3.0 | 2.6 | 2.8 | 37.6 | 11.4 | 6.2 | 4.6 |
| 6 | $a = 0.2$ | 100 | 0.0 | 60.0 | 15.2 | 7.6 | 6.8 | 100.0 | 75.0 | 32.8 | 18.0 |
| | | 200 | 0.0 | 15.2 | 4.4 | 3.0 | 3.0 | 86.4 | 28.4 | 12.2 | 8.0 |
| | | 500 | 0.0 | 6.8 | 3.2 | 3.6 | 3.4 | 32.4 | 10.6 | 7.2 | 5.0 |
| | $a = 0.5$ | 100 | 0.2 | 61.8 | 11.8 | 5.2 | 3.6 | 99.8 | 77.6 | 38.6 | 19.8 |
| | | 200 | 0.0 | 19.2 | 5.8 | 3.6 | 3.8 | 89.6 | 29.0 | 14.2 | 8.6 |
| | | 500 | 0.0 | 6.4 | 3.6 | 5.6 | 4.4 | 37.2 | 12.6 | 9.2 | 6.8 |
| | $a = 0.9$ | 100 | 0.2 | 56.4 | 12.6 | 5.8 | 5.0 | 100.0 | 75.6 | 33.6 | 17.8 |
| | | 200 | 0.0 | 19.0 | 7.2 | 5.4 | 5.2 | 88.4 | 31.4 | 17.6 | 12.6 |
| | | 500 | 0.0 | 7.0 | 3.0 | 4.0 | 3.8 | 33.2 | 10.6 | 7.6 | 7.0 |

## 4.2 The Case of One Break

In the following we evaluate the probability of correctly detecting the number of structural changes and the accuracy of change-point estimation when the true number of breaks is small. We generate data from

$$y_t = \beta_t x_t + u_t,$$

with

- DGP-1: $\beta_t = \mathbf{1}\{T/2 < t \leq T\}$, $x_t \sim i.i.d.\ N(0,1)$, $u_t \sim i.i.d.\ N(0, \sigma_u^2)$.
- DGP-2: $\beta_t = \mathbf{1}\{T/2 < t \leq T\}$, $x_t \sim i.i.d.\ N(0,1)$, $u_t = \sigma_u v_t$ with $v_t = 0.5v_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, 0.75)$.
- DGP-3: $\beta_t = \mathbf{1}\{T/2 < t \leq T\}$, $x_t \sim$ AR(1), $u_t \sim i.i.d.\ N(0, \sigma_u^2)$.
- DGP-4: $\beta_t = \mathbf{1}\{T/2 < t \leq T\}$, $x_t \sim$ AR(1), $u_t = \sigma_u \sqrt{h_t}\epsilon_t$, $h_t = 0.05 + 0.05u_{t-1}^2 + 0.9h_{t-1}$, $\epsilon_t \sim i.i.d.\ N(0,1)$.
- DGP-5: $\beta_t = \mathbf{1}\{T/2 < t \leq T\}$, $x_t \sim$ AR(1), $u_t = \sigma_u v_t$ with $v_t = \epsilon_t + 0.5\epsilon_{t-1}$, $\epsilon_t \sim i.i.d.\ N(0, 0.8)$.
- DGP-6: $\beta_t = 0.2\mathbf{1}\{1 \leq t \leq T/2\} + 0.8\mathbf{1}\{T/2 < t \leq T\}$, $x_t = y_{t-1}$, $u_t \sim i.i.d.\ N(0, \sigma_u^2)$.

In all the above DGP's, the coefficient on $x_t$ has a break at $T/2$ and the intercept is a constant zero.[4] DGP-1 is the benchmark case where both $x_t$ and $u_t$ are i.i.d. DGP-2 and DGP-3 introduce AR(1) structure to $u_t$ and $x_t$, respectively. As in the case of no breaks, we generate AR(1) processes with an AR coefficient of 0.5 and make sure that the processes have unit variances. DGP-4 considers GARCH(1,1) error along with an AR(1) regressor. DGP-5 considers MA(1) error along with an AR(1) regressor. And DGP-6 considers an auto-regression with a break in the AR coefficient. Again we set $\sigma_u = 0.5, 1$, and 1.5. We compare our approach with the sequential procedure in BP, which first looks at the UDMax or WDMax test to see if a break exists and then examines the $\sup F(\ell + 1|\ell)$ statistics sequentially. This procedure (BP) is the preferred strategy by Bai and Perron (2006). Here we only consider the nonrobust version of BP, since as shown above, the robust version gives poor size performance in general. Table 2 summarizes the proportions of correct estimation (pce) of $m$ (number of breaks) for each method and, conditional on correct estimation of $m$ ($\hat{m} = 1$), the accuracy of break date estimation, which we measure by average Hausdorff distance divided by $T$ (hd/$T$). All figures in the table are in percentages (%).

In the benchmark case of DGP-1, GFL gives satisfactory results in terms of both pce and hd/$T$ at low and medium noise levels. At the high noise level ($\sigma_u = 1.5$), pce's also rise quickly as $T$ increases. In comparison, BP outperforms GFL in terms of pce at the high noise level but underperforms at the low noise level ($\sigma_u = 0.5$). When $\sigma_u = 1$, the comparison is mixed. Comparing the accuracy of break-date estimation, GFL almost uniformly outperforms BP, especially when the latter takes small trimming sizes. Similar patterns emerge in the results from DGP-2 to DGP-5. In DGP-6, the pce of GFL appears to converge faster than BP to 100% as $T$ increases. BP, however, slightly outperforms GFL in terms of the accuracy of the break-date estimation, especially if BP takes a bigger trimming parameter.

Table 2: Detecting One Break (Under pce are proportions of correctly estimating the number of breaks. Under hd/$T$ are average Hausdorff distance between estimated and true sets of break dates in percentages of $T$, given that the estimated number of breaks is correct.)

| DGP | $\sigma_u$ | $T$ | GFL | | BP | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | tr=.05 | | .10 | | .15 | | tr=.20 | |
| | | | pce | $\frac{\text{hd}}{T}$ | pce | hd/$T$ | pce | hd/$T$ | pce | hd/$T$ | pce | hd/$T$ |
| 1 | 0.5 | 100 | 100.0 | 1.1 | 58.6 | 1.1 | 87.6 | 1.1 | 93.2 | 1.1 | 95.0 | 1.2 |
| | | 200 | 100.0 | 0.6 | 80.0 | 0.6 | 89.8 | 0.6 | 91.6 | 0.6 | 94.2 | 0.6 |
| | | 500 | 100.0 | 0.2 | 87.4 | 0.2 | 92.8 | 0.2 | 93.0 | 0.2 | 95.6 | 0.2 |
| | 1.0 | 100 | 71.8 | 3.0 | 54.4 | 4.1 | 80.8 | 3.8 | 88.0 | 3.8 | 90.6 | 3.7 |
| | | 200 | 92.0 | 1.7 | 79.0 | 2.0 | 89.8 | 2.0 | 91.0 | 2.0 | 94.0 | 1.9 |
| | | 500 | 99.8 | 0.7 | 89.4 | 0.7 | 92.0 | 0.7 | 93.4 | 0.7 | 95.6 | 0.7 |
| | 1.5 | 100 | 19.4 | 5.0 | 48.8 | 11.9 | 52.8 | 9.1 | 55.6 | 7.8 | 56.8 | 6.9 |
| | | 200 | 31.4 | 2.8 | 73.6 | 4.9 | 82.2 | 4.6 | 85.6 | 4.4 | 87.8 | 4.4 |
| | | 500 | 73.6 | 1.2 | 91.0 | 1.5 | 94.6 | 1.5 | 93.4 | 1.5 | 95.0 | 1.5 |
| 2 | 0.5 | 100 | 98.2 | 1.1 | 10.6 | 0.9 | 35.2 | 1.0 | 52.2 | 1.0 | 67.8 | 1.1 |
| | | 200 | 99.6 | 0.6 | 14.2 | 0.6 | 32.0 | 0.7 | 46.0 | 0.6 | 59.6 | 0.6 |
| | | 500 | 100.0 | 0.2 | 19.4 | 0.3 | 37.4 | 0.2 | 51.8 | 0.2 | 65.6 | 0.2 |
| | 1.0 | 100 | 74.8 | 3.9 | 13.4 | 6.1 | 35.0 | 5.1 | 49.4 | 5.2 | 62.2 | 4.5 |
| | | 200 | 90.2 | 1.9 | 20.0 | 2.3 | 38.8 | 2.6 | 55.2 | 2.4 | 69.4 | 2.4 |
| | | 500 | 98.0 | 0.7 | 18.0 | 0.8 | 35.8 | 0.7 | 50.0 | 0.8 | 61.6 | 0.7 |
| | 1.5 | 100 | 30.2 | 6.5 | 12.2 | 11.8 | 29.0 | 10.0 | 45.8 | 9.7 | 57.8 | 8.8 |
| | | 200 | 42.8 | 3.7 | 15.2 | 3.8 | 37.2 | 4.6 | 52.2 | 4.9 | 65.8 | 5.0 |
| | | 500 | 77.8 | 1.6 | 21.0 | 1.7 | 37.6 | 1.8 | 49.0 | 1.8 | 62.2 | 1.8 |
| 3 | 0.5 | 100 | 99.8 | 1.5 | 57.2 | 1.3 | 86.6 | 1.4 | 92.2 | 1.4 | 94.0 | 1.4 |
| | | 200 | 100.0 | 0.6 | 80.2 | 0.6 | 91.4 | 0.6 | 94.6 | 0.6 | 95.2 | 0.6 |
| | | 500 | 100.0 | 0.2 | 90.8 | 0.2 | 91.2 | 0.2 | 93.4 | 0.2 | 95.2 | 0.2 |
| | 1.0 | 100 | 64.4 | 3.3 | 58.2 | 4.6 | 82.0 | 4.3 | 87.0 | 4.1 | 90.8 | 4.1 |
| | | 200 | 91.2 | 1.8 | 83.2 | 2.0 | 92.0 | 2.1 | 94.8 | 2.1 | 96.4 | 2.1 |
| | | 500 | 98.6 | 0.8 | 88.6 | 0.8 | 91.0 | 0.8 | 91.8 | 0.8 | 95.6 | 0.8 |
| | 1.5 | 100 | 17.4 | 5.1 | 44.6 | 12.3 | 52.0 | 8.8 | 57.0 | 7.9 | 61.6 | 7.4 |
| | | 200 | 25.2 | 3.1 | 72.2 | 5.3 | 83.4 | 5.1 | 87.2 | 5.2 | 90.6 | 4.9 |
| | | 500 | 76.6 | 1.5 | 88.6 | 1.8 | 92.0 | 1.7 | 93.4 | 1.8 | 95.0 | 1.7 |
| 4 | 0.5 | 100 | 99.8 | 1.5 | 51.0 | 1.4 | 82.6 | 1.5 | 90.8 | 1.5 | 93.2 | 1.5 |
| | | 200 | 100.0 | 0.7 | 81.6 | 0.6 | 92.6 | 0.6 | 94.0 | 0.6 | 94.6 | 0.6 |
| | | 500 | 100.0 | 0.2 | 88.8 | 0.2 | 91.2 | 0.2 | 91.4 | 0.2 | 93.6 | 0.2 |
| | 1.0 | 100 | 62.0 | 3.4 | 57.6 | 4.5 | 79.6 | 4.5 | 87.6 | 4.5 | 90.0 | 4.4 |
| | | 200 | 89.6 | 1.9 | 79.6 | 2.2 | 90.8 | 2.2 | 94.6 | 2.2 | 96.4 | 2.2 |
| | | 500 | 98.8 | 0.8 | 89.0 | 0.8 | 93.0 | 0.8 | 94.2 | 0.9 | 95.0 | 0.8 |
| | 1.5 | 100 | 22.8 | 5.5 | 43.6 | 9.9 | 52.0 | 7.6 | 55.4 | 7.3 | 59.4 | 6.5 |

Continued on next page

Table 2: Continued

| DGP | $\sigma_u$ | $T$ | GFL | | BP | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | tr=.05 | | .10 | | .15 | | tr=.20 | |
| | | | pce | $\frac{\text{hd}}{T}$ | pce | hd/$T$ | pce | hd/$T$ | pce | hd/$T$ | pce | hd/$T$ |
| | 1.5 | 200 | 31.4 | 3.3 | 72.6 | 4.5 | 81.6 | 4.5 | 87.2 | 4.3 | 90.2 | 4.2 |
| | | 500 | 73.4 | 1.3 | 90.4 | 1.6 | 94.4 | 1.6 | 94.8 | 1.6 | 97.0 | 1.6 |
| | | 100 | 98.8 | 1.6 | 22.0 | 1.5 | 55.2 | 1.6 | 70.2 | 1.7 | 80.8 | 1.6 |
| | 0.5 | 200 | 100.0 | 0.8 | 39.8 | 0.7 | 59.0 | 0.7 | 69.2 | 0.7 | 79.2 | 0.7 |
| | | 500 | 100.0 | 0.3 | 45.8 | 0.3 | 59.2 | 0.3 | 68.0 | 0.3 | 78.2 | 0.3 |
| | | 100 | 66.2 | 4.4 | 22.0 | 7.7 | 52.0 | 6.6 | 65.6 | 5.9 | 76.2 | 5.5 |
| 5 | 1.0 | 200 | 92.4 | 2.2 | 37.4 | 2.7 | 61.6 | 2.5 | 71.2 | 2.5 | 78.6 | 2.5 |
| | | 500 | 98.8 | 1.0 | 46.4 | 0.9 | 60.4 | 1.0 | 70.6 | 1.1 | 79.4 | 1.1 |
| | | 100 | 28.0 | 6.6 | 24.4 | 16.3 | 46.4 | 13.3 | 51.4 | 11.2 | 57.2 | 9.5 |
| | 1.5 | 200 | 42.6 | 3.5 | 36.8 | 8.0 | 54.2 | 5.9 | 65.6 | 5.3 | 76.2 | 5.3 |
| | | 500 | 72.6 | 1.7 | 49.2 | 2.3 | 68.2 | 2.3 | 73.8 | 2.3 | 80.8 | 2.3 |
| | | 100 | 65.0 | 8.6 | 63.0 | 8.4 | 68.2 | 8.2 | 72.6 | 7.7 | 76.2 | 7.5 |
| | 0.5 | 200 | 87.8 | 5.4 | 88.6 | 5.0 | 92.6 | 4.7 | 92.8 | 4.6 | 94.8 | 4.5 |
| | | 500 | 98.0 | 2.5 | 93.8 | 1.7 | 93.8 | 1.7 | 95.8 | 1.7 | 97.8 | 1.7 |
| | | 100 | 65.8 | 7.6 | 62.2 | 8.2 | 67.8 | 7.8 | 71.4 | 7.1 | 76.6 | 6.8 |
| 6 | 1.0 | 200 | 88.2 | 5.0 | 90.2 | 4.6 | 94.2 | 4.6 | 95.6 | 4.5 | 97.0 | 4.3 |
| | | 500 | 98.0 | 2.6 | 91.8 | 1.6 | 93.2 | 1.6 | 95.0 | 1.6 | 96.4 | 1.6 |
| | | 100 | 66.8 | 8.4 | 63.6 | 7.4 | 69.6 | 7.3 | 72.8 | 7.0 | 77.8 | 6.5 |
| | 1.5 | 200 | 87.0 | 5.0 | 90.2 | 4.4 | 94.2 | 4.4 | 95.0 | 4.3 | 96.6 | 4.2 |
| | | 500 | 98.4 | 2.5 | 91.0 | 1.6 | 94.2 | 1.6 | 95.2 | 1.6 | 96.4 | 1.6 |

## 4.3 The Case of Many Breaks

To evaluate the finite-sample performance for the case of many breaks, we consider two setups. First we set constant regime length and let the number of regimes increase. In the second setup the number of breaks is fixed and regime lengths increases proportional to sample size. Specifically, let $\delta = T/R$, where $R$ is an even number of regimes $(m+1)$ that divides $T$ with no remainder. We generate data from the following equation,

$$y_t = x_t\beta_t + u_t,$$

where $x_t \sim i.i.d.\ N(0,1)$, $u_t \sim i.i.d.\ N(0,\sigma_u^2)$, and

$$\beta_t = \begin{cases} 0 & \delta(2i)+1 \le t < \delta(2i+1) \\ 1 & \delta(2i+1)+1 \le t < \delta(2i+2) \end{cases}, \ i = 0,1,\ldots,R/2.$$

We specify

- DGP-1: Fix $\delta = 30$ and vary $R = 6, 10, 20$.

Table 3: **Detecting Many Breaks**

| | | DGP 1 | | | | DGP 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | GFL | | BP | | | GFL | | BP | |
| $\sigma_u$ | $R$ | pce | hd/T | pce | hd/T | T | pce | hd/T | pce | hd/T |
| | 6 | 99 | 1 | 95 | 0.7 | 150 | 98 | 1.3 | 52.2 | 1 |
| 0.2 | 10 | 99.8 | 0.7 | 99.4 | 0.5 | 300 | 100 | 0.7 | 98.8 | 0.5 |
| | 20 | 99.4 | 0.4 | 0 | NaN | 600 | 100 | 0.4 | 99.6 | 0.3 |
| | 6 | 92.4 | 2 | 71 | 1.9 | 150 | 53.8 | 2.4 | 0.8 | 2.8 |
| 0.5 | 10 | 84.8 | 1.5 | 38.8 | 1.4 | 300 | 83.2 | 1.5 | 38.8 | 1.4 |
| | 20 | 36.4 | 1.2 | 0 | NaN | 600 | 93 | 0.7 | 99.4 | 0.7 |

Note: In DGP-1 (fixed regime length), $\delta = 30$. In DGP-2 (fixed number of regimes), $R = 10$. Under pce are proportions of correctly estimating the number of breaks. Under hd/$T$ are average Hausdorff distance between estimated and true sets of break dates in percentages of $T$, given that the estimated number of breaks is correct.

- DGP-2: Fix $R = 10$ and vary $T = 150, 300, 600$.

For the BP approach, we set trimming size of 0.05, allowing the maximum number of breaks to be 18. The results are summarized in Table 3.

In the case of DGP-1, GFL correctly estimates the number of breaks in most repetitions (close to 100%) at the low noise level. At the high noise level, pce drops significantly, especially when at the same time the true number of breaks is high. However, the performance of BP seems even more sensitive to noise. Notice that when $R = 20$, pce for BP is zero at both noise levels, since the number of breaks exceeds the maximum allowed by the trimming size. For both approaches, we witness a declining performance as the sample size increases along with the true number of breaks. In the case of DGP-2, $R$ is fixed at 10 and $\delta$ increases proportionally with the sample size $T$. We do see improving performance as $T$ increases. GFL dominates BP in terms of pce. In terms of the accuracy of break-date estimation, both approaches give satisfactory performances.

# 5 An Empirical Illustration

In this section we present an empirical illustration of our method. We consider the problem of predicting equity premium using fundamental information. We use a subset of the quarterly data of Welch and Goyal (2008), which has been updated to 2011. The equity premium ($y$) is the return on the stock market minus the prevailing risk-free rate. We use the return on S&P 500 index as the proxy of the stock market return and take the short-term T-bill rate as the risk-free rate. The fundamental information we consider includes earning price ratio ($ep$) and dividend price ratio ($dp$). We refer to Welch and Goyal

Table 4: **Summary Statistics**

| | Num | Mean | S.D. | Min | Max | Median | Skew. | Kurt. |
|---|---|---|---|---|---|---|---|---|
| $y$ | 363 | 0.0053 | 0.1050 | -0.5023 | 0.6226 | 0.0206 | 0.0683 | 10.7594 |
| $dp$ | 363 | 0.0407 | 0.0178 | 0.0112 | 0.1490 | 0.0377 | 1.0800 | 6.4649 |
| $ep$ | 363 | 0.0730 | 0.0291 | 0.0082 | 0.1695 | 0.0637 | 0.7299 | 3.1455 |

Note: $y$ is equity premium, $dp$ is dividend to price ratio, and $ep$ is earning price ratio.

(2008) for detailed description of the data and sources. Table 4 summarizes the data we use. We estimate the following predictive regression with structural breaks,

$$y_{t+1} = \beta_{0,t} + \beta_{1,t} dp_t + \beta_{2,t} ep_t + u_{t+1}.$$

The parameter $\beta_t = (\beta_{0,t}, \beta_{1,t}, \beta_{2,t})'$ may contain multiple breaks in the calendar range from 1921Q2 to 2011Q4, reflecting discrete changes in the way how equities are priced overall.

The main results are summarized in Table 5. The estimation contains two steps. First we estimate break dates, then we perform the usual OLS estimation in each regime. For each OLS regression, coefficient estimates and standard errors are tabulated along with $R^2$ and F statistics for model significance tests. Our approach (GFL) detects two breaks at 1932Q3 and 1942Q3. Possible reasons for the first break include the bottoming out of the stock market, election of FDR into presidency, and the passage of the Securities Act of 1933, which comprehensively regulated the securities industry. The second break may be attributed to the deepening US involvement in the World War II. Linear regressions in all three regimes are statistically significant at the 5% level. Before the first break, the slope on $dp$ is significantly negative and that on $ep$ significantly positive. This is reversed in the second regime, although the negative slope of $ep$ fails to be statistically significant at the 5% level. In the third regime, the effect of $dp$ remains significantly positive but weakens substantially and the effect of $ep$ remains insignificant.

For the purpose of comparison, we also estimate the model using Bai and Perron's approach (WDMax coupled with $\sup F(\ell + 1|\ell)$) with different trimming sizes. If trimming equals 15%, BP fails to detect any break. Under 10% trimming, one break is detected at 1932Q3, which coincides with the first break detected by our method. If trimming equals 5%, two breaks are detected at 1928Q2 and 1933Q2. These results once again show the importance of choosing a correct trimming size for Bai and Perron's approach. A large trimming implicitly imposes restrictive assumptions that may preclude detection of true breaks, but a small trimming like 5% tends to produce false structural breaks, as shown in simulations. Using our approach, in contrast, practitioners do not have to face such choices. The continuous nature of the tuning parameter $\lambda$ offers an even richer trade-offs between goodness of fit and model simplicity. And as shown in Theorem 3.4, our IC-based procedure to choose $\lambda$ naturally rules out the possibility of over- and under-fitting, at least asymptotically.

Table 5: **Empirical Results**

| Regime Range | $\hat{\beta}_{0,t}$ | $\hat{\beta}_{1,t}$ $(dp)$ | $\hat{\beta}_{2,t}$ $(ep)$ | $R^2$ | F |
|---|---|---|---|---|---|
| GFL: | | | | | |
| 1921Q2-1932Q2 | 0.027 (0.776) | -3.6747 (0.008) | 2.0546 (0.040) | 0.205 | 4.998 (0.007) |
| 1932Q3-1942Q2 | -0.1501 (0.080) | 5.8416 (0.002) | -2.259 (0.109) | 0.263 | 6.411 (0.002) |
| 1942Q3-2011Q4 | -0.0194 (0.126) | 1.5207 (0.021) | -0.3831 (0.221) | 0.0369 | 3.982 (0.020) |
| BP, trim=0.05: | | | | | |
| 1921Q2-1928Q1 | 0.1072 (0.285) | -0.6536 (0.000) | -0.4801 (0.416) | 0.214 | 0.453 (0.636) |
| 1928Q2-1933Q1 | -1.0497 (0.666) | -7.0342 (0.008) | 21.8756 (0.000) | 0.585 | 10.528 (0.000) |
| 1933Q2-2011Q4 | -0.0198 (0.416) | 1.6299 (0.000) | -0.4759 (0.106) | 0.0383 | 5.079 (0.007) |
| BP, trim=0.10: | | | | | |
| 1921Q2-1932Q2 | 0.027 (0.776) | -3.6747 (0.008) | 2.0546 (0.040) | 0.205 | 4.998 (0.007) |
| 1932Q3-2011Q4 | -0.0301 (0.037) | 2.1835 (0.000) | -0.6301 (0.024) | 0.0758 | 11.655 (0.000) |
| BP, trim=0.15: | | | | | |
| 1921Q2-2011Q4 | -0.0261 (0.090) | 0.4584 (0.293) | 0.174 (0.513) | 0.0161 | 2.492 (0.084) |

Note: p-value's for significance tests (t and F) are given in parentheses.

# 6 Conclusion

We propose a shrinkage procedure for the determination of the number of structural changes in a multiple linear regression model via GFL. We show that our method consistently determines the number of breaks and the estimated break dates are sufficiently close to the true break dates. Simulation results suggest that our new method performs well in finite samples in comparison with Bai and Perron (1998).

There are several interesting topics for further research. First, we consider the estimation and inference in OLS regression models with an unknown number of breaks in this paper. It is straightforward to extend to the GMM framework without essential changes. Second, following the lead of Andews (2003) who consider end-of-sample stability test, it is also possible to allow a break to occur at the end of a random sample. Third, it is also possible to extend our method to the panel data framework. The last decade has seen a growing literature on estimation and testing of common breaks in panel data models; see, De Watcher and Tzavalis (2005, 2012), Chan et al. (2008), Bai (2010), Kim (2011, 2014), Hsu and Lin (2012), Liao and Wang (2012), Baltagi et al. (2014), among others. We are exploring some of these topics in ongoing work.

# Notes

[1]Since the minimization in (2.6) is a convex problem, we may use a general-purpose convex solver system such as CVX (Grant et al., 2009). However, the general solver does not exploit the special structure of our problem, hence computationally inefficient.

[2]We also conduct a robustness check by considering $\rho_T = c_1 T^{-c_2}$ for $c_1 = 0.9$, 1 and 1.1 and $c_2 = 0.4$, 0.5, and 0.6. The results are available upon request.

[3]We also experimented with UDMax in BP and found differences between UDMax and WDMax negligible.

[4]The experiments for DGPs with two breaks yield similar results.

*REFERENCE*


Andrews, D. W. K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821-856.

Andrews, D. W. K. (2003) End-of-sample instability tests. *Econometrica* 71, 1661-1694.

Andrews, D. W. K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383-1414.

Angelosante, D. & G. B. Giannakis (2012) Group Lassoing change-points in piecewise-constant AR processes. *EURASIP Journal on Advances in Signal Processing* 1(70), 1-16.

Bai, J. (1995) Least absolute deviation estimation of a shift. *Econometric Theory* 11, 403-436.

Bai, J. (1997a) Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79, 551-563.

Bai, J. (1997b) Estimating multiple breaks one at a time. *Econometric Theory* 13, 315-352.

Bai, J. (1998) Estimation of multiple-regime regressions with least absolute deviation. *Journal of Statistical Planning and Inference* 74, 103-134.

Bai, J. (2010) Common breaks in means and variances for panel data. *Journal of Econometrics* 157, 78-92.

Bai, J. R. L. Lumsdaine & J. Stock (1998) Testing and dating common breaks in multivariate time series. *Review of Economic Studies* 65, 395-432.

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

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

Bai, J. & P. Perron (2003b) Critical values for multiple structural change tests. *Econometrics Journal* 6, 72-78.

Bai, J. & P. Perron (2006) Multiple structural change models: a simulation analysis. In D. Corbae, S. N. Durlauf, and B. E. Hansen (eds.), *Econometric Theory and Practice*. Cambridge University Press, Cambridge.

Baltagi, B. H., Q. Feng & C. Kao (2014) Estimation of heterogeneous panels with structural breaks. *Working Paper*, Syracuse University.

Belloni, A., V. Chernozhukov & C. Hansen (2012) Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80, 2369-2429.

Belloni, A., V. Chernozhukov & C. Hansen (2014) Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies,* forthcoming.

Bertsekas, D. (1995) *Nonlinear Programming.* Athena Scientific, Belmont, MA.

Bleakley, K. & J-P. Vert (2012) The group fused Lasso for multiple change point detection. *Working paper*, INRIA Saclay, Orsay, France.

Caner, M. (2009) Lasso-type GMM Estimator. *Econometric Theory* 25, 270-290.

Caner, M. & M. Fan (2011) A near minimax risk bound: adaptive Lasso with heteroskedastic data in instrumental variable selection. *Working Paper*, North Carolina State University.

Caner, M. & K. Knight (2013) An alternative to unit root tests: Bridge estimators differentiate between nonstationary versus stationary models and select optimal lag. *Journal of Statistical Planning and Inference* 143, 691-715.

Chan, F., T. Mancini-Griffoli & L. L. Pauwels (2008) Stability tests for heterogenous panel. *Working paper*, Curtin University of Technology.

De Watcher, S. & E. Tzavalis (2005) Monte Carlo comparison of model and moment selection and classical inference approaches to break detection in panel data models. *Economics Letters* 99, 91-96.

De Watcher, S. & E. Tzavalis (2012) Detection of structural breaks in linear dynamic panel data models. *Computational Statistics and Data Analysis* 56, 3020-3034.

Fan, J., & H. Peng (2004) Nonconcave penalized likelihood with a diverging number of parameters. *Annals of Statistics* 32, 928-961.

Friedman, J., T. Hastie, H. Höfling & R. Tibshirani (2007) Pathwise coordinate optimization. *Annals of Applied Statistics* 1, 302-332.

Grant, M., S. Boyd & Y. Ye (2009) CVX: Matlab Software for Disciplined Convex Programming. *Mimeo.*

Hall, P. & C. C. Heyde (1980) *Martingale Limit Theory and Its Applications.* Academic Press, New York.

Harchaoui, Z. & C. Lévy-Leduc (2010) Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association* 105, 1481-1493.

Hsu, C-C. & C-C. Lin (2012) Change-point estimation for nonstationary panel. *Working paper*, National Central University.

Kim, D. (2011) Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics* 164, 310-330.

Kim, D. (2014) Common breaks in time trends for large panel data with a factor structure. *The Econometrics Journal*, forthcoming.

Kock, A. B. (2013) Oracle efficient variable selection in random and fixed effects panel data models. *Econometric Theory* 29, 115-152.

Knight, K. & W. Fu (2000) Asymptotics for Lasso-type estimators. *Annals of Statistics* 28, 1356-1378.

Kurozumi, E. & Y. Arai (2006) Efficient estimation and inference in cointegrating regressions with structural breaks. *Journal of Time Series* Analysis 28, 545-575.

Kurozumi, E. (2012) Testing for multiple structural changes with non-homogeneous regressors. *Working paper*, Hitotsubashi University.

Lam, C. & J. Fan (2008) Profile-kernel likelihood inference with diverging number of parameters. *Annals of Statistics* 36, 2232-2260.

Leeb, H. & B. M. Pötscher (2005) Model selection and inference: facts and fiction. *Econometric Theory* 21, 21-59.

Leeb, H. & B. M. Pötscher (2008) Sparse estimators and the oracle property, or the return of the Hodges estimator. *Journal of Econometrics* 142, 201-211.

Liao, W. & P. Wang (2012) Structural breaks in panel data models: a common distribution approach. *Working paper*, HKUST.

Liao, Z. (2013) Adaptive GMM shrinkage estimation with consistent moment selection. *Econometric Theory* 29, 857-904.

Liao, Z. & P. C. B. Phillips (2014) Automated estimation of vector error correction models. *Econometric Theory*, forthcoming.

Liu, Q. & F. Watbled (2009) Exponential inequalities for martingales and asymptotic properties of the free energy of directed polymers in a random experiment. *Stochastic Processes and Their Applications* 119, 3101-3132.

Lu, X. & L. Su (2013) Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Working paper*, Singapore Management University.

Lu, X. & L. Su (2014) Jackknife model averaging for quantile regressions. *Working paper*, Singapore Management University.

Merlevède F., M. Peligrad & E. Rio (2009) Bernstein inequality and moderate deviations under strong mixing conditions. IMS collections. *High Dimensional Probability* V., 273-292.

Merlevède F., Peligrad, M., Rio, E., 2011. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields* 151, 435-474.

Ohlsson H., L. Ljung & S. Boyd (2010) Segmentation of ARX-models using sum-of-norms regularization. *Automatica* 46, 1107-1111.

Perron, P. (2006) Dealing with structural breaks. In T. C. Mills and K. Patterson (eds.), *Palgrave Handbook of Econometrics, Vol 1: Econometric Theory*, pp. 278-352. Palgrave Macmillan, New York.

Pötscher, B. M. & H. Leeb (2009) On the distribution of penalized maximum likelihood estimators: the LASSO, SCAD, and thresholding. *Journal of Multivariate Analysis* 100, 2065-2082.

Pötscher, B. M. & U. Schneider (2009) On the distribution of the adaptive LASSO estimator. *Journal of Statistical Planning and Inference* 139, 2775-2790.

Qu, Z. & P. Perron (2007) Estimating and testing structural changes in multiple regressions. *Econometrica* 75, 459-502.

Rinaldo, A. (2009) Properties and refinement of the fused Lasso. *Annals of Statistics* 37, 2922-2952.

Su, L. & H. White (2010) Testing structural change in partially linear models. *Econometric Theory* 26, 1761-1806.

Su, L., P. Xu & H. Ju (2013) Pricing for goodwill: a threshold quantile regression approach. *Working paper*, Singapore Management University.

Tibshirani, R. J. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 58, 267-288.

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu & K. Knight (2005) Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society, Series B* 67, 91-108.

Welch, I. & A. Goyal (2008) A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21, 1455-1508.

White, H. (2001) *Asymptotic Theory for Econometricians*. 2nd edition, Emerald, UK.

Yuan, M. & Y. Lin (2006) Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B* 68, 49-67.

Zou, H. (2006) The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* 101, 1418-1429.

# APPENDIX

## A Some Technical Lemmas

In this section we prove some technical lemmas that are used in the proof of the main results in the paper.

**Lemma A.1** *Consider the PLS problem in (2.3) or equivalently (2.4). Let* $\{\hat{\beta}_t,\ t=1,2...,T\}$ *and* $\{\hat{\theta}_t,\ t=2,...,T\}$ *denote the respective solutions. Then*

*(i)* $\frac{1}{T}\sum_{r=\hat{T}_j}^{T} x_r\left(y_r - x_r'\hat{\beta}_r\right) = \frac{\lambda}{2}\hat{\theta}_{\hat{T}_j}/\left\|\hat{\theta}_{\hat{T}_j}\right\|$ *for* $j=1,...,\hat{m}$;

*(ii)* $\frac{1}{T}\left\|\sum_{r=t}^{T} x_r\left(y_r - x_r'\hat{\beta}_r\right)\right\| \leq \frac{\lambda}{2}$ *for* $t=1,...,T$.

**Proof.** To prove the above lemma, we invoke subdifferential calculus (e.g., Bertsekas (1995, Appendix B.5)). We first rewrite the PLS criterion function as

$$\bar{V}_{T\lambda}\left(\{\theta_t\}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - x_t'\sum_{s=1}^{t}\theta_s\right)^2 + \lambda\sum_{t=2}^{T}\|\theta_t\|. \tag{A.1}$$

A necessary and sufficient condition for $\{\hat{\theta}_t\}$ to minimize (A.1) is that for each $t=1,...,T$, $\mathbf{0}_{p\times 1}$ belongs to the subdifferential of (A.1) with respect to $\theta_t$ evaluated at $\{\hat{\theta}_t\}$. That is,

$$-\frac{2}{T}\sum_{r=t}^{T} x_r\left(y_r - x_r'\sum_{s=1}^{r}\hat{\theta}_s\right) + \lambda e_t = \mathbf{0}_{p\times 1} \tag{A.2}$$

where for $t=2,...,T$

$$e_t = \frac{\hat{\theta}_t}{\left\|\hat{\theta}_t\right\|} \text{ if } \left\|\hat{\theta}_t\right\| \neq 0 \text{ and } \|e_t\| \leq 1 \text{ if } \left\|\hat{\theta}_t\right\| = 0, \tag{A.3}$$

and $e_1 = \mathbf{0}_{p\times 1}$. If $t=\hat{T}_j$ for some $j\in\{1,...,\hat{m}\}$, i.e., $t$ is one of the estimated break dates, then $\hat{\theta}_t = \hat{\beta}_t - \hat{\beta}_{t-1} \neq \mathbf{0}_{p\times 1}$ and we obtain (i) as the breaks cannot occur at $t=1$ and $\sum_{s=1}^{r}\hat{\theta}_s = \hat{\beta}_r$. In general, (A.2) and (A.3) imply that (ii) holds for all $t\geq 2$. When $t=1$, the first order condition with respect to $\theta_1$ yields $\sum_{r=1}^{T} x_r'\left(y_r - x_r'\sum_{s=1}^{r}\hat{\theta}_s\right) = \mathbf{0}_{p\times 1}$ so that (ii) is also satisfied for $t=1$. ■

**Lemma A.2** *Let* $\{\xi_t, t=1,2,...\}$ *be a zero-mean strong mixing process, not necessarily stationary, with the mixing coefficients satisfying* $\alpha(\tau)\leq c_\alpha\rho^\tau$ *for some* $c_\alpha>0$ *and* $\rho\in(0,1)$.

*(i) If* $\sup_{1\leq t\leq T}|\xi_t|\leq M_T$, *then there exists a constant* $C_0$ *depending on* $c_\alpha$ *and* $\rho$ *such that for any* $T\geq 2$ *and* $\epsilon>0$,

$$P\left(\left|\sum_{t=1}^{T}\xi_t\right|>\epsilon\right)\leq \exp\left(-\frac{C_0\epsilon^2}{v_0^2 T + M_T^2 + \epsilon M_T\left(\log T\right)^2}\right),$$

*where* $v_0^2 = \sup_{t\geq 1}\left[Var(\xi_t) + 2\sum_{s=t+1}^{\infty}|Cov(\xi_t,\xi_s)|\right]$.

*(ii) If* $\sup_{t\geq 1}P\left(|\xi_t|>v\right)\leq \exp\left(1-(v/b)^\gamma\right)$ *for some* $b\in(0,\infty)$ *and* $\gamma\in(0,\infty]$, *then there exist constants* $C_1$ *and* $C_2$ *depending only on* $b$, $c_\alpha$, $\rho$, *and* $\gamma$ *such that for any* $T\geq 4$ *and* $\epsilon\geq C_0(\log T)^{\eta_0}$ *with* $\eta_0$, $C_0>0$,

$$P\left(\left|\sum_{t=1}^{T}\xi_t\right|>\epsilon\right)\leq (T+1)\exp\left(-\frac{\epsilon^{\frac{\gamma}{1+\gamma}}}{C_1}\right)+\exp\left(-\frac{\epsilon^2}{TC_2}\right).$$

**Proof.** (i) Merlevède et al. (2009, Theorem 2) prove (i) under the condition $\alpha(\tau) \leq \exp(-2c\tau)$ for some $c > 0$. If $c_\alpha = 1$, we can take $\rho = \exp(-2c)$ and apply the theorem to obtain the claim in (i). Other values of $c_\alpha$ do not alter the conclusion.

(ii) Merlevède et al. (2011, Theorem 1) prove a result that is more general than that in (ii) under the condition $\alpha(\tau) \leq \exp(-c_1\tau^{\gamma_1})$ for some $c_1$, $\gamma_1 > 0$. If $c_\alpha = 1$ and $\gamma_1 = 1$, we can take $\rho = \exp(-2c_1)$ and apply the theorem to obtain the claim in (ii). Other values of $c_\alpha$ do not alter the conclusion. ■

**Lemma A.3** *Suppose Assumptions A1 and A2 hold. Let $v_T = T\delta_T$. Then*

*(i)* $\sup_{\substack{1\leq s<r\leq T+1 \\ r-s\geq v_T}} \mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} x_t x_t'\right) \leq \bar{c}_{xx} + o_P(1)$;

*(ii)* $\inf_{\substack{1\leq s<r\leq T+1 \\ r-s\geq v_T}} \mu_{\min}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} x_t x_t'\right) \geq \underline{c}_{xx} + o_P(1)$.

**Proof.** (i) By Weyl inequality, the fact that $|\mu_{\max}(A)| \leq \|A\|$ for any symmetric matrix $A$, and Assumption A2,

$$\begin{aligned}\mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} x_t x_t'\right) &\leq \mu_{\max}\left(\frac{1}{r-s}\sum_{t=s}^{r-1} E(x_t x_t')\right) + \left\|\frac{1}{r-s}\sum_{t=s}^{r-1}[x_t x_t' - E(x_t x_t')]\right\| \\ &\leq \bar{c}_{xx} + \left\|\frac{1}{r-s}\sum_{t=s}^{r-1}[x_t x_t' - E(x_t x_t')]\right\|.\end{aligned}$$

It suffices to prove the theorem by showing that $\max_{\substack{1\leq s<r\leq T+1 \\ r-s\geq v_T}} \left\|\frac{1}{r-s}\sum_{t=s}^{r-1}[x_t x_t' - E(x_t x_t')]\right\| = o_P(1)$.

We first consider the case where Assumption A1(ii.a) holds so that $v_T \geq c_v T^{1/q}$. Let $\eta_T = T^{1/(2q)}$. Let $\iota_{sp}$ be an arbitrary $p\times 1$ unit vector such that $\|\iota_{sp}\| = 1$ for $s = 1, 2$. Let $\zeta_t \equiv \iota_{1p}'[x_t x_t' - E(x_t x_t')]\iota_{2p}$, $\zeta_{1t} \equiv \iota_{1p}'[x_t x_t'\mathbf{1}_t - E(x_t x_t'\mathbf{1}_t)]\iota_{2p}$ and $\zeta_{2t} \equiv \iota_{1p}'[x_t x_t'\bar{\mathbf{1}}_t - E(x_t x_t'\bar{\mathbf{1}}_t)]\iota_{2p}$, where $\mathbf{1}_t \equiv \mathbf{1}\{\|x_t\|^2 \leq \eta_T\}$ and $\bar{\mathbf{1}}_t = 1 - \mathbf{1}_t$. Note that $\zeta_t = \zeta_{1t} + \zeta_{2t}$. By Boole inequality and Lemma A.2(i),

$$\begin{aligned}&P\left(\sup_{\substack{1\leq s<r\leq T+1 \\ r-s\geq v_T}}\left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1}\zeta_{1t}\right| \geq C(\log T)^3\right) \leq T^2 \sup_{\substack{1\leq s<r\leq T+1 \\ r-s\geq v_T}} P\left(\left|\sum_{t=s}^{r-1}\zeta_{1t}\right| \geq C\sqrt{r-s}(\log T)^3\right) \\ \leq\ & T^2 \sup_{\substack{1\leq s<r\leq T+1 \\ r-s\geq v_T}} \exp\left(-\frac{C_0 C^2 (r-s)(\log T)^6}{v_0^2(r-s) + 4\eta_T^2 + 2C\sqrt{r-s}(\log T)^3\eta_T\left[\log\left(\sqrt{r-s}\right)\right]^2}\right) \\ \leq\ & \exp\left(-\frac{C_0 C^2 v_T(\log T)^6}{v_0^2 C v_T + 4\eta_T^2 + \frac{1}{2}C\sqrt{v_T}(\log T)^3\eta_T[\log v_T]^2} + 2\log T\right) \\ \to\ & 0 \text{ as } T\to\infty.\end{aligned}$$

By Assumption A1(ii.a), Boole and Markov inequalities, and the dominated convergence theorem,

$$\begin{aligned}&P\left(\sup_{\substack{1\leq s<r\leq T+1 \\ r-s\geq v_T}}\left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1}\zeta_{2t}\right| \geq C(\log T)^3\right) \leq P\left(\max_{1\leq t\leq T}\|x_t\|^2 \geq \eta_T\right) \\ \leq\ & T\max_{1\leq t\leq T} P\left(\|x_t\|^2 \geq \eta_T\right) \leq \frac{T}{\eta_T^{2q}}\max_{1\leq t\leq T} E\left[\|x_t\|^{4q}\mathbf{1}\left\{\|x_t\|^2 \geq \eta_T\right\}\right] \to 0 \text{ as } T\to\infty.\end{aligned}$$

Noting that $\iota_{1p}$ and $\iota_{2p}$ are arbitrary unit vectors, we infer that $\max_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}} \left\| \frac{1}{r-s}\sum_{t=s}^{r-1}\left[x_t x_t' - E\left(x_t x_t'\right)\right]\right\|$ $= O_P\left(v_T^{-1/2}(\log T)^3\right) = o_P(1)$. Then (i) follows.

Now we consider the case where Assumption A1(ii.b) holds where $v_T \ge c_v (\log T)^{(2+\gamma)/\gamma}$. By Boole inequality and Lemma A.2(ii), for any sufficiently large $C$

$$\begin{aligned}
& P\left(\sup_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}}\left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1}\zeta_t\right| \ge C\sqrt{\log T}\right) \le T^2 \sup_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}} P\left(\left|\sum_{t=s}^{r-1}\zeta_t\right| \ge C\sqrt{(r-s)\log T}\right) \\
\le\ & T^2 \sup_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}}\left[(T+1)\exp\left(-\frac{[C(r-s)\log T]^{\gamma/[2(1+\gamma)]}}{C_1}\right) + \exp\left(-\frac{(r-s)\log T}{C_2}\right)\right] \\
\le\ & \exp\left(-\frac{(Cv_T\log T)^{\gamma/[2(1+\gamma)]}}{C_1} + 4\log T\right) + \exp\left(-\frac{Cv_T\log T}{C_2} + 2\log T\right) \\
\to\ & 0 \text{ as } T\to\infty,
\end{aligned}$$

as $(v_T \log T)^{\gamma/[2(1+\gamma)]} \propto \log T$ by construction. It follows that $\max_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}} \left\| \frac{1}{r-s}\sum_{t=s}^{r-1}\left[x_t x_t' - E\left(x_t x_t'\right)\right]\right\|$ $= O_P\left(v_T^{-1/2}(\log T)^{1/2}\right) = o_P(1)$.

(ii) The proof of (ii) is analogous and thus omitted. ■

**Lemma A.4** *Suppose Assumptions A1(i) and A2 hold. Let $v_T = T\delta_T$.*

*(i) If Assumption A1(ii.a) holds, then $\sup_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}} \left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1} x_t u_t\right| = O_P\left((\log T)^3\right)$;*

*(ii) If Assumption A1(ii.b) holds, then $\sup_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}} \left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1} x_t u_t\right| = O_P\left((\log T)^{1/2}\right)$.*

**Proof.** (i) In this case, $v_T \ge c_v T^{1/q}$. Let $\eta_T = T^{1/(2q)}$ and $\iota_{1p}$ be as defined in the proof of Lemma A.3(i). Let $\varsigma_t \equiv \iota_{1p}'\left[x_t u_t - E\left(x_t u_t\right)\right]$, $\varsigma_{1t} \equiv \iota_{1p}'\left[x_t u_t \mathbf{1}_t - E\left(x_t u_t \mathbf{1}_t\right)\right]$ and $\varsigma_{2t} \equiv \iota_{1p}'\left[x_t u_t \bar{\mathbf{1}}_t - E\left(x_t u_t \bar{\mathbf{1}}_t\right)\right]$, where now $\mathbf{1}_t \equiv \mathbf{1}\left\{\|x_t u_t\| \le \eta_T\right\}$ and $\bar{\mathbf{1}}_t = 1 - \mathbf{1}_t$. Note that $\varsigma_t = \varsigma_{1t} + \varsigma_{2t}$. Arguments like those used the proof of Lemma A.3(i) show that for any sufficiently large $C$, $P\left(\max_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}} \left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1}\varsigma_{lt}\right| \ge C(\log T)^3\right)$ $\to 0$ as $T\to\infty$ for $l = 1, 2$. Then (i) follows.

(ii) In this case, $v_T \ge c_v (\log T)^{(2+\gamma)/\gamma}$ and arguments like those used the proof of Lemma A.3(i) show that for any sufficiently large $C$, $P\left(\max_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}} \left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1}\varsigma_t\right| \ge C(\log T)^{1/2}\right) \to 0$ as $T\to\infty$. Then (ii) holds. ■

**Remark.** If in addition, $\{x_t u_t\}$ is an m.d.s. with respect to $\mathcal{F}_t$ in Lemma A.4(ii), then for any $v_T \to \infty$ and $C > 0$ we can apply Theorem 1.1 in Liu and Watbled (2009) to obtain

$$\begin{aligned}
P\left(\max_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}}\left|\frac{1}{\sqrt{r-s}}\sum_{t=s}^{r-1}\varsigma_t\right| \ge C\sqrt{\log T}\right) &\le T^2 \max_{\substack{1\le s<r\le T+1\\ r-s\ge v_T}} \exp\left(-C^2 C_3\sqrt{r-s}\log T\right) \\
&\le \exp\left(-C^2 C_3\sqrt{v_T}\log T + 2\log T\right) \to 0 \text{ as } T\to\infty,
\end{aligned}$$

where $C_3$ is a constant that does not depend on $T$.

# B Proof of Theorem 3.1

(i) Our proof strategy follows closely from that of Proposition 3 in Harchaoui and Lévy-Leduc (2010). Define

$$A_{T,j} = \left\{\left|\hat{T}_j - T_j^0\right| \geq T\delta_T\right\} \text{ and } C_T = \left\{\max_{1\leq l\leq m^0}\left|\hat{T}_l - T_l^0\right| < I_{\min}/2\right\}. \tag{B.1}$$

Since $P\left(\max_{1\leq j\leq m^0}\left|\hat{T}_j - T_j^0\right| \geq T\delta_T\right) \leq \sum_{j=1}^{m^0} P\left(A_{T,j}\right)$ and $m^0 < \infty$, it suffices to show that (i1) $\sum_{j=1}^{m^0} P(A_{T,j}\cap\, C_T) \to 0$ and (i2) $\sum_{j=1}^{m^0} P\left(A_{T,j}\cap C_T^c\right) \to 0$, where $C_T^c$ denotes the complement of $C_T$.

**We first prove** (i1) by showing that $\sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap C_T\right) \to 0$ and $\sum_{j=1}^{m^0} P\left(A_{T,j}^- \cap C_T\right) \to 0$, where $A_{T,j}^+ = \{T_j^0 - \hat{T}_j \geq T\delta_T\}$ and $A_{T,j}^- = \{\hat{T}_j - T_j^0 \geq T\delta_T\}$. Without loss of generality (Wlog) we prove that $\sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap C_T\right) \to 0$ as the other case follows analogously. By the definition of $C_T$, we have

$$T_{j-1}^0 < \hat{T}_j < T_{j+1}^0 \text{ for all } j \in \left\{1, ..., m^0\right\}. \tag{B.2}$$

By (2.2) and Lemma A.1, we have $\frac{-1}{T}\sum_{r=\hat{T}_j}^{T} x_r x_r'(\hat{\beta}_r - \beta_r^0) + \frac{1}{T}\sum_{r=\hat{T}_j}^{T} x_r u_r = \frac{\lambda}{2} e_{\hat{T}_j}$ and $||\frac{-1}{T}\sum_{r=T_j^0}^{T} x_r x_r'$ $(\hat{\beta}_r - \beta_r^0) + \frac{1}{T}\sum_{r=T_j^0}^{T} x_r u_r|| \leq \frac{\lambda}{2}$, where $e_{\hat{T}_j} = \hat{\theta}_{\hat{T}_j}/||\hat{\theta}_{\hat{T}_j}||$. By the triangle inequality and the fact that $||e_{\hat{T}_j}|| = 1$, we have

$$\begin{aligned}
\lambda &\geq \left\|\frac{-1}{T}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\hat{\beta}_r - \beta_r^0\right) + \frac{1}{T}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| = \left\|\frac{-1}{T}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\hat{\alpha}_{j+1} - \alpha_j^0\right) + \frac{1}{T}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| \\
&\geq \left\|\frac{1}{T}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\alpha_{j+1}^0 - \alpha_j^0\right)\right\| - \left\|\frac{1}{T}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\hat{\alpha}_{j+1} - \alpha_{j+1}^0\right)\right\| - \left\|\frac{1}{T}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| \\
&\equiv R_{T,j1} - R_{T,j2} - R_{T,j3}, \text{ say,}
\end{aligned} \tag{B.3}$$

where the equality follows from the fact that $\hat{\beta}_r = \hat{\alpha}_{j+1}$ and $\beta_r^0 = \alpha_j^0$ for $r \in [\hat{T}_j, T_j^0 - 1]$ by (B.2). Define the event $\bar{R}_{T,j}(\lambda) = \left\{\lambda \geq \frac{1}{3}R_{T,j1}\right\} \cup \left\{R_{T,j2} \geq \frac{1}{3}R_{T,j1}\right\} \cup \left\{R_{T,j3} \geq \frac{1}{3}R_{T,j1}\right\}$. It is easy to show that $P\left(\bar{R}_{T,j}(\lambda)\right) = 1$. It follows that

$$\begin{aligned}
P\left(A_{T,j}^+ \cap C_T\right) &\leq P\left(A_{T,j}^+ \cap C_T \cap \left\{\lambda \geq \frac{1}{3}R_{T,j1}\right\}\right) + P\left(A_{T,j}^+ \cap C_T \cap \left\{R_{T,j2} \geq \frac{1}{3}R_{T,j1}\right\}\right) \\
&\quad + P\left(A_{T,j}^+ \cap C_T \cap \left\{R_{T,j3} \geq \frac{1}{3}R_{T,j1}\right\}\right) \\
&\equiv AC_{j1} + AC_{j2} + AC_{j3}, \text{ say.}
\end{aligned}$$

We first bound $\sum_{j=1}^{m^0} AC_{j1}$. Noting that $\|AB\| = [\mathrm{tr}(BB'A'A)]^{1/2} \geq \mu_{\min}\left(A'A\right)^{1/2}\|B\|$, we have

$$\begin{aligned}
\sum_{j=1}^{m^0} AC_{j1} &\leq \sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap \left\{\lambda \geq \frac{1}{3}R_{T,j1}\right\}\right) \\
&= \sum_{j=1}^{m^0} P\left(\left\|\frac{1}{T_j^0 - \hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\alpha_{j+1}^0 - \alpha_j^0\right)\right\| \leq \frac{3T\lambda}{T_j^0 - \hat{T}_j};\ T_j^0 - \hat{T}_j \geq T\delta_T\right) \\
&\leq \sum_{j=1}^{m^0} P\left(c_{1T,j} \leq 3\lambda/(J_{\min}\delta_T);\ T_j^0 - \hat{T}_j \geq T\delta_T\right) \to 0,
\end{aligned}$$

where $c_{1T,j} \equiv \mu_{\min}\left(\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\right) \geq \underline{c}_{xx}/2 > 0$ w.p.a.1 by Lemma A.3(ii) and $\lambda/(J_{\min}\delta_T) \to 0$ by Assumption A3(iii). Next, we bound $\sum_{j=1}^{m^0} AC_{j2}$. Observe that

$$\begin{aligned} AC_{j2} &= P\left(A_{T,j}^+ \cap C_T \cap \left\{\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\hat{\alpha}_{j+1}-\alpha_{j+1}^0\right)\right\| \geq \frac{1}{3}\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\alpha_{j+1}^0-\alpha_j^0\right)\right\|\right\}\right) \\ &\leq P\left(A_{T,j}^+ \cap C_T \cap \left\{\bar{c}_{1T,j}\left\|\hat{\alpha}_{j+1}-\alpha_{j+1}^0\right\| \geq \frac{1}{3}c_{1T,j}\left\|\alpha_{j+1}^0-\alpha_j^0\right\|\right\}\right), \end{aligned}$$

where $\bar{c}_{1T,j} \equiv \mu_{\max}\left(\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\right) \leq 2\bar{c}_{xx}$ w.p.a.1 by Lemma A.3(i). Note that $\hat{\beta}_t = \hat{\alpha}_{j+1}$ for $t \in [T_j^0, \left(T_j^0+T_{j+1}^0\right)/2-1]$ as $\hat{T}_j < T_j^0$ given $A_{T,j}^+$ and $\hat{T}_{j+1} > \left(T_j^0+T_{j+1}^0\right)/2$ conditional on the event $C_T$. Using Lemma A.1(ii) with $t = \left(T_j^0+T_{j+1}^0\right)/2$ and $t = T_j^0$ and following the steps to obtain (B.3), we have $\lambda \geq \left\|\frac{1}{T}\sum_{r=T_j^0}^{\left(T_j^0+T_{j+1}^0\right)/2-1} x_r x_r'\left(\hat{\alpha}_{j+1}-\alpha_{j+1}^0\right)\right\| - \left\|\frac{1}{T}\sum_{r=T_j^0}^{\left(T_j^0+T_{j+1}^0\right)/2-1} x_r u_r\right\|$. It follows that conditional on $C_T$, $\left\|\hat{\alpha}_{j+1}-\alpha_{j+1}^0\right\| \leq \left(c_{2T,j}\right)^{-1}\left[\frac{2\lambda T}{I_{\min}} + \left\|\frac{2}{T_{j+1}^0-T_j^0}\sum_{r=T_j^0}^{\left(T_j^0+T_{j+1}^0\right)/2-1} x_r u_r\right\|\right]$, where $c_{2T,j} \equiv \mu_{\min}\left(\frac{2}{T_{j+1}^0-T_j^0}\sum_{r=T_j^0}^{\left(T_j^0+T_{j+1}^0\right)/2-1} x_r x_r'\right) \geq \underline{c}_{xx}/2$ w.p.a.1 by Lemma A.3(ii). Consequently, we have

$$\begin{aligned} &\sum_{j=1}^{m^0} P\left(\left\{\left\|\hat{\alpha}_{j+1}-\alpha_{j+1}^0\right\| \geq \bar{c}_{1T,j}^{-1}c_{1T,j}\left\|\alpha_j^0-\alpha_{j+1}^0\right\|/3\right\} \cap C_T\right) \\ \leq\ &\sum_{j=1}^{m^0} P\left(\frac{2\lambda T}{I_{\min}} \geq \bar{c}_{1T,j}^{-1}c_{1T,j}c_{2T,j}\left\|\alpha_j^0-\alpha_{j+1}^0\right\|/6\right) \\ &+\sum_{j=1}^{m^0} P\left(\left\|\frac{2}{T_{j+1}^0-T_j^0}\sum_{r=T_j^0}^{\left(T_j^0+T_{j+1}^0\right)/2-1} x_r u_r\right\| \geq \bar{c}_{1T,j}^{-1}c_{1T,j}c_{2T,j}\left\|\alpha_j^0-\alpha_{j+1}^0\right\|/6\right). \end{aligned}$$

The first term converges to zero because $\lambda T/(I_{\min}J_{\min}) \to 0$ under Assumptions A3(i) and (iii). The second term is bounded from above by $\sum_{j=1}^{m^0} P\left(\left\|\frac{1}{T_{j+1}^0-T_j^0}\sum_{r=T_j^0}^{\left(T_j^0+T_{j+1}^0\right)/2-1} x_r u_r\right\| \geq \bar{c}_{xx}\underline{c}_{xx}^2 J_{\min}/96\right) \to 0$ by analogous arguments as used in the proof of Lemma A.4 and the fact that $I_{\min}^{-1/2}\left(\log m^0\right)^{c_\delta/2} = o\left(J_{\min}\right)$ under Assumptions A3(i)-(ii). It follows that $\sum_{j=1}^{m^0} AC_{j2} \to 0$. Noting that $c_{1T,j} \geq \underline{c}_{xx}/2 > 0$ w.p.a.1 and $\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| = O_P\{[T\delta_T/(\log T)^{c_\delta}]^{-1/2}\} = o_P\left(J_{\min}\right)$ when $T_j^0 - \hat{T}_j \geq T\delta_T$ by Lemma A.4 and Assumption A2(ii), we have

$$\begin{aligned} \sum_{j=1}^{m^0} AC_{j3} &\leq \sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap \left\{R_{T,3} \geq \frac{1}{3}R_{T,1}\right\}\right) \\ &= \sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap \left\{\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| \geq \frac{1}{3}\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\alpha_{j+1}^0-\alpha_j^0\right)\right\|\right\}\right) \\ &\leq \sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap \left\{\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| \geq \frac{1}{3}c_{1T,j}J_{\min}\right\}\right) \to 0. \end{aligned}$$

Here, the last convergence is obtained by strengthening the results in Lemma A.4 through the squeezing of $\log m^0 (< \log T)$ into the exponent when applying the exponential inequality in Lemma A.3. So we have shown that $\sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap C_T\right) \to 0$.

**Now we prove** (i2). We prove this by showing that $\sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap C_T^c\right) \to 0$ and $\sum_{j=1}^{m^0} P\left(A_{T,j}^- \cap C_T^c\right) \to 0$. Wlog we prove that $\sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap C_T^c\right) \to 0$. Define

$$\begin{aligned} D_T^{(l)} &\equiv \left\{\exists j \in \{1,...,m^0\}, \hat{T}_j \leq T_{j-1}^0\right\} \cap C_T^c, \\ D_T^{(m)} &\equiv \left\{\forall j \in \{1,...,m^0\}, T_{j-1}^0 < \hat{T}_j < T_{j+1}^0\right\} \cap C_T^c, \text{ and} \\ D_T^{(r)} &\equiv \left\{\exists j \in \{1,...,m^0\}, \hat{T}_j \geq T_{j+1}^0\right\} \cap C_T^c. \end{aligned}$$

Then $\sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap C_T^c\right) = \sum_{j=1}^{m^0} P(A_{T,j}^+ \cap D_T^{(l)}) + \sum_{j=1}^{m^0} P(A_{T,j}^+ \cap D_T^{(m)}) + \sum_{j=1}^{m^0} P(A_{T,j}^+ \cap D_T^{(r)})$.

We first consider $\sum_{j=1}^{m^0} P(A_{T,j}^+ \cap D_T^{(m)})$. Observe that

$$\begin{aligned} & P\left(A_{T,j}^+ \cap D_T^{(m)}\right) \\ = \; & P\left(A_{T,j}^+ \cap \{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2} I_{\min}\} \cap D_T^{(m)}\right) + P\left(A_{T,j}^+ \cap \{\hat{T}_{j+1} - T_j^0 < \frac{1}{2} I_{\min}\} \cap D_T^{(m)}\right) \\ \leq \; & P\left(A_{T,j}^+ \cap \{\hat{T}_{j+1} - T_j^0 \geq \frac{1}{2} I_{\min}\} \cap D_T^{(m)}\right) + P\left(A_{T,j}^+ \cap \{T_{j+1}^0 - \hat{T}_{j+1} \geq \frac{1}{2} I_{\min}\} \cap D_T^{(m)}\right), \end{aligned}$$

where the inequality follows as $0 \leq \hat{T}_{j+1} - T_j^0 \leq I_{\min}/2$ implies that $T_{j+1}^0 - \hat{T}_{j+1} = (T_{j+1}^0 - T_j^0) - (\hat{T}_{j+1} - T_j^0) \geq I_{\min} - I_{\min}/2 = I_{\min}/2$. Further noticing that $\left\{A_{T,j}^+ \cap \{T_{j+1}^0 - \hat{T}_{j+1} \geq I_{\min}/2\} \cap D_T^{(m)}\right\} \subset \cup_{k=j+1}^{m^0-1} \left(\{T_k^0 - \hat{T}_k \geq I_{\min}/2\} \cap \{\hat{T}_{k+1} - T_k^0 \geq I_{\min}/2\} \cap D_T^{(m)}\right)$, we have

$$\begin{aligned} \sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap D_T^{(m)}\right) \leq & \sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap \{\hat{T}_{j+1} - T_j^0 \geq I_{\min}/2\} \cap D_T^{(m)}\right) \\ & + \sum_{j=1}^{m^0} \sum_{k=j+1}^{m^0-1} P\left(\{T_k^0 - \hat{T}_k \geq I_{\min}/2\} \cap \{\hat{T}_{k+1} - T_k^0 \geq I_{\min}/2\} \cap D_T^{(m)}\right). \quad \text{(B.4)} \end{aligned}$$

To bound the first term on the right-hand side of (B.4), we apply Lemma A.1 with $t = \hat{T}_j$ and $t = T_j^0$ to obtain $\frac{1}{T}\sum_{r=\hat{T}_j}^{T} x_r(y_r - x_r'\hat{\beta}_r) = \frac{\lambda}{2}\hat{\theta}_{\hat{T}_j}/||\hat{\theta}_{\hat{T}_j}||$ and $\frac{1}{T}\left\|\sum_{r=T_j^0}^{T} x_r'(y_r - x_r'\hat{\beta}_r)\right\| \leq \frac{\lambda}{2}$. This, in conjunction with (2.2), implies that $\frac{\lambda T}{T_j^0 - \hat{T}_j} \geq \frac{1}{T_j^0 - \hat{T}_j}\left\|-\sum_{r=\hat{T}_j}^{T_j^0-1} x_r x_r'\left(\hat{\alpha}_{j+1} - \alpha_j^0\right) + \sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| \geq c_{1T,j}\left\|\hat{\alpha}_{j+1} - \alpha_j^0\right\| - \left\|\frac{1}{T_j^0 - \hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\|$. It follows that

$$\left\|\hat{\alpha}_{j+1} - \alpha_j^0\right\| \leq c_{1T,j}^{-1}\left[\frac{\lambda T}{T_j^0 - \hat{T}_j} + \left\|\frac{1}{T_j^0 - \hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r'\right\|\right]. \quad \text{(B.5)}$$

Similarly, applying Lemma A.1 with $t = \hat{T}_{j+1}$ and $t = T_j^0$ yields $\frac{1}{T}\sum_{r=\hat{T}_{j+1}}^{T} x_r(y_r - x_r'\hat{\beta}_r) = \frac{\lambda}{2}\hat{\theta}_{\hat{T}_{j+1}}/||\hat{\theta}_{\hat{T}_{j+1}}||$ and $\frac{1}{T}\left\|\sum_{r=T_j^0}^{T} x_r'(y_r - x_r'\hat{\beta}_r)\right\| \leq \frac{\lambda}{2}$, which, in conjunction with and (2.2), implies that $\frac{\lambda T}{\hat{T}_{j+1} - T_j^0} \geq$

$\frac{1}{\hat{T}_{j+1}-T_j^0}\left\|\sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r\left(y_r - x_r'\hat{\beta}_r\right)\right\| \geq c_{3T,j}\left\|\hat{\alpha}_{j+1}-\alpha_{j+1}^0\right\| - \left\|\frac{1}{\hat{T}_{j+1}-T_j^0}\sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r u_r\right\|$, where $c_{3T,j} \equiv \mu_{\min}\left(\frac{1}{\hat{T}_{j+1}-T_j^0}\sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r x_r'\right) \geq \underline{c}_{xx}/2$ w.p.a.1 by Lemma A.3(ii). So

$$\left\|\hat{\alpha}_{j+1}-\alpha_{j+1}^0\right\| \leq c_{3T,j}^{-1}\left[\frac{\lambda T}{\hat{T}_{j+1}-T_j^0} + \left\|\frac{1}{\hat{T}_{j+1}-T_j^0}\sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r u_r\right\|\right]. \tag{B.6}$$

Define the event

$$\begin{aligned} E_{T,j} \equiv \Bigg\{ &\left\|\alpha_{j+1}^0-\alpha_j^0\right\| \leq \lambda\left(\frac{T}{T_j^0-\hat{T}_j}c_{1T,j}^{-1} + \frac{T}{\hat{T}_{j+1}-T_j^0}c_{3T,j}^{-1}\right) \\ &+ c_{1T,j}^{-1}\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| + c_{3T,j}^{-1}\left\|\frac{1}{\hat{T}_{j+1}-T_j^0}\sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r u_r\right\|\Bigg\}. \end{aligned} \tag{B.7}$$

By the triangle inequality, (B.5) and (B.6) imply that $E_{T,j}$ occurs with probability one. It follows that

$$\begin{aligned} &\sum_{j=1}^{m^0} P\left(A_{T,j}^+ \cap \left\{\hat{T}_{j+1}-T_j^0 \geq I_{\min}/2\right\} \cap D_T^{(m)}\right) \\ = &\sum_{j=1}^{m^0} P\left(E_{T,j} \cap A_{T,j}^+ \cap \left\{\hat{T}_{j+1}-T_j^0 \geq I_{\min}/2\right\} \cap D_T^{(m)}\right) \\ \leq &\sum_{j=1}^{m^0} P\left(E_{T,j} \cap \left\{T_j^0-\hat{T}_j > T\delta_T\right\} \cap \left\{\hat{T}_{j+1}-T_j^0 \geq I_{\min}/2\right\}\right) \\ \leq &\sum_{j=1}^{m^0} P\left(\lambda\delta_T^{-1}c_{1T,j}^{-1} + \frac{2\lambda T}{I_{\min}}c_{3T,j}^{-1} \geq \left\|\alpha_{j+1}^0-\alpha_j^0\right\|/3\right) \\ &+\sum_{j=1}^{m^0} P\left(\left\{c_{1T,j}^{-1}\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\| \geq \left\|\alpha_{j+1}^0-\alpha_j^0\right\|/3\right\} \cap \left\{T_j^0-\hat{T}_j > T\delta_T\right\}\right) \\ &+\sum_{j=1}^{m^0} P\left(\left\{c_{3T,j}^{-1}\left\|\frac{1}{\hat{T}_{j+1}-T_j^0}\sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r u_r\right\| \geq \left\|\alpha_{j+1}^0-\alpha_j^0\right\|/3\right\} \cap \left\{\hat{T}_{j+1}-T_j^0 \geq I_{\min}/2\right\}\right). \end{aligned} \tag{B.8}$$

The first term in (B.8) converges to zero because $\lambda/(J_{\min}\delta_T) = o(1)$ and $\lambda T/(I_{\min}J_{\min}) = o(1)$ by Assumptions A3(i) and (iii). The second and third terms in (B.8) converge to zero because $\left\|\frac{1}{T_j^0-\hat{T}_j}\sum_{r=\hat{T}_j}^{T_j^0-1} x_r u_r\right\|$ $= O_P\{[T\delta_T/(\log T)^{c_\delta}]^{-1/2}\} = o_P(J_{\min})$ by Lemma A.4 and Assumption A3(ii), $\left\|\frac{1}{\hat{T}_{j+1}-T_j^0}\sum_{r=T_j^0}^{\hat{T}_{j+1}-1} x_r u_r\right\|$ $= O_P\{[I_{\min}/(\log T)^{c_\delta}]^{-1/2}\} = o_P(J_{\min})$ by Lemma A.4 and Assumptions A3(i)-(ii), and by strengthening the results in Lemma A.4 through the squeezing of $\log m^0 (< \log T)$ into the exponent. Similarly, we can show that the second term in (B.4) converges to zero.

Now we consider $\sum_{j=1}^{m^0} P(A_{T,j}^+ \cap D_T^{(l)})$. Observe that

$$P\left(A_{T,j}^+ \cap D_T^{(l)}\right) \leq P\left(D_T^{(l)}\right) \leq \sum_{j=1}^{m^0} 2^{j-1} P\left(\max\left\{l \in \{1,...,m^0\} : \hat{T}_l \leq T_{l-1}^0\right\} = j\right),$$

and the event $\max\left\{l \in \{1,...,m^0\} : \hat{T}_l \leq T^0_{l-1}\right\} = j$ implies that $\hat{T}_j \leq T^0_{j-1}$ and $\hat{T}_{l+1} > T^0_l$ for all $l = j$, ..., $m^0$, and $\left\{\max\{l \in \{1,...,m^0\} : \hat{T}_l \leq T^0_{l-1}\} = j\right\} \subset \cup_{k=j}^{m^0-1}\left(\{T^0_k - \hat{T}_k \geq I_{\min}/2\} \cap \{\hat{T}_{k+1} - T^0_k \geq I_{\min}/2\}\right)$. It follows that

$$\begin{aligned}\sum_{j=1}^{m^0} P\left(A^+_{T,j} \cap D^{(l)}_T\right) &\leq m^0 \sum_{j=1}^{m^0-1} 2^{j-1} \sum_{k=j}^{m^0-1} P\left(\left\{T^0_k - \hat{T}_k \geq I_{\min}/2\right\} \cap \left\{\hat{T}_{k+1} - T^0_k \geq I_{\min}/2\right\}\right) \\ &\quad + m^0 2^{m^0-1} P\left(T^0_{m^0} - \hat{T}_{m^0} \geq I_{\min}/2\right). \end{aligned} \tag{B.9}$$

Consider the last term on the right-hand side of (B.9). Applying $j = m^0$ in (B.7) suggests that the event $E_{T,m^0}$ occurs with probability one. It follows that

$$\begin{aligned} & m^0 2^{m^0-1} P\left(T^0_{m^0} - \hat{T}_{m^0} \geq I_{\min}/2\right) \\ = \; & m^0 2^{m^0-1} P\left(E_{T,m^0} \cap \left\{T^0_{m^0} - \hat{T}_{m^0} \geq I_{\min}/2\right\}\right) \\ \leq \; & m^0 2^{m^0-1} P\left(\lambda \delta_T^{-1} c^{-1}_{1T,m^0} + \frac{2\lambda T}{I_{\min}} c^{-1}_{3T,m^0} \geq \left\|\alpha^0_{m^0+1} - \alpha^0_{m^0}\right\|/3\right) \\ & + m^0 2^{m^0-1} P\left(c^{-1}_{1T,m^0} \left\|\frac{1}{T^0_{m^0} - \hat{T}_{m^0}} \sum_{r=\hat{T}_{m^0}}^{T^0_{m^0}-1} x_r u_r\right\| \geq \left\|\alpha^0_{m^0+1} - \alpha^0_{m^0}\right\|/3,\; T^0_{m^0} - \hat{T}_{m^0} \geq I_{\min}/2\right) \\ & + m^0 2^{m^0-1} P\left(c^{-1}_{3T,m^0} \left\|\frac{1}{T - T^0_{m^0}} \sum_{r=T^0_{m^0}}^{T} x_r u_r\right\| \geq \left\|\alpha^0_{m^0+1} - \alpha^0_{m^0}\right\|/3\right) \\ \to \; & 0, \end{aligned}$$

by similar arguments to those used in the study of (B.8) and the fact that $m^0 2^{m^0-1} = O(T \log T)$ and $\log(T \log T) \leq \log\left(T^{1+\epsilon/2}\right)$ can be squeezed into the the exponent when applying the exponential inequality in Lemma A.3. Now, we consider the first term on the right-hand side of (B.9). Using (B.7) with $j = k$, similar arguments like those used in the study of (B.8), and the fact that $\log(\left(m^0\right)^2 2^{m^0-1}) = O(\log\left(T^{1+\epsilon/2}\right))$ yields

$$\begin{aligned} & m^0 \sum_{j=1}^{m^0-1} 2^{j-1} \sum_{k=j}^{m^0-1} P\left(\left\{T^0_k - \hat{T}_k \geq I_{\min}/2\right\} \cap \left\{\hat{T}_{k+1} - T^0_k \geq I_{\min}/2\right\}\right) \\ \leq \; & m^0 2^{m^0-1} \sum_{k=1}^{m^0-1} P\left(E_{T,k} \cap \left\{T^0_k - \hat{T}_k \geq I_{\min}/2\right\} \cap \left\{\hat{T}_{k+1} - T^0_k \geq I_{\min}/2\right\}\right) \\ \leq \; & m^0 2^{m^0-1} \sum_{k=1}^{m^0-1} P\left(\lambda \delta_T^{-1} c^{-1}_{1T,k} + \frac{2\lambda T}{I_{\min}} c^{-1}_{3T,k} \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/3\right) \\ & + m^0 2^{m^0-1} \sum_{k=1}^{m^0-1} P\left(\left\{c^{-1}_{1T,k} \left\|\frac{1}{\hat{T}_k - T^0_k} \sum_{r=\hat{T}_k}^{T^0_k-1} x_r u'_r\right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/3\right\} \cap \left\{T^0_k - \hat{T}_k \geq I_{\min}/2\right\}\right) \\ & + m^0 2^{m^0-1} \sum_{k=1}^{m^0-1} P\left(\left\{c^{-1}_{3T,k} \left\|\frac{1}{\hat{T}_{k+1} - T^0_k} \sum_{r=T^0_k}^{\hat{T}_{k+1}-1} x_r u_r\right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/3\right\} \cap \left\{\hat{T}_{k+1} - T^0_k \geq I_{\min}/2\right\}\right) \\ \to \; & 0. \end{aligned}$$

It follows that $\sum_{j=1}^{m^0} P(A_{T,j}^{+} \cap D_T^{(l)}) \to 0$. Analogously, we can show that $\sum_{j=1}^{m^0} P(A_{T,j}^{+} \cap D_T^{(r)}) \to 0$.

**We now prove (ii).** By the result in part (i) and Assumption A3(i), $\left|\hat{T}_j - T_j^0\right| = O_P(T\delta_T) = o_P(I_{\min})$ uniformly in $j = 1, ..., m^0$. It follows that either $(T_{j-1}^0 + T_j^0)/2 < \hat{T}_j < T_j^0$ or $T_j^0 \leq \hat{T}_j < (T_j^0 + T_{j+1}^0)/2$ holds for each $j$. Fix $l \in \{1, ..., m^0\}$, wlog we assume that $(T_{l-1}^0 + T_l^0)/2 < \hat{T}_l < T_l^0$ and consider two subcases: (ii1) $(T_l^0 + T_{l+1}^0)/2 < \hat{T}_{l+1} < T_{l+1}^0$ and (ii2) $T_{l+1}^0 \leq \hat{T}_{l+1}$. In subcase (ii1), using Lemma A.1(i) with $t = \hat{T}_l$ and $\hat{T}_{l+1}$ and (2.2) yields

$$\begin{aligned}
\lambda &\geq \left\| \frac{1}{T} \sum_{r=\hat{T}_l}^{T} x_r \left(y_r - x_r' \hat{\beta}_r\right) - \frac{1}{T} \sum_{r=\hat{T}_{l+1}}^{T} x_r \left(y_r - x_r' \hat{\beta}_r\right) \right\| = \left\| \frac{1}{T} \sum_{r=\hat{T}_l}^{\hat{T}_{l+1}-1} x_r \left(y_r - x_r' \hat{\beta}_r\right) \right\| \\
&= \left\| \frac{1}{T} \sum_{r=\hat{T}_l}^{T_l^0-1} \left[x_r x_r' \left(\alpha_l^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] + \frac{1}{T} \sum_{r=T_l^0}^{\hat{T}_{l+1}-1} \left[x_r x_r' \left(\alpha_{l+1}^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] \right\| \\
&\geq \left\| \frac{1}{T} \sum_{r=T_l^0}^{\hat{T}_{l+1}-1} \left[x_r x_r' \left(\alpha_{l+1}^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] \right\| - \left\| \frac{1}{T} \sum_{r=\hat{T}_l}^{T_l^0-1} \left[x_r x_r' \left(\alpha_l^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] \right\| \\
&\geq \frac{\hat{T}_{l+1} - T_l^0}{T} \left\{ c_{T,l} \left\|\hat{\alpha}_{l+1} - \alpha_{l+1}^0\right\| - O_P((I_{l+1}^0)^{-1/2}) \right\} - O_P\left((T_l^0 - \hat{T}_l)/T\right)
\end{aligned}$$

where $c_{T,l} \equiv \mu_{\min}\left(\frac{1}{\hat{T}_{l+1} - T_l^0} \sum_{r=T_l^0}^{\hat{T}_{l+1}-1} x_r x_r'\right) \geq \underline{c}_{xx}/2$ w.p.a.1 by Lemma A.3(ii) and the result in part (i). It follows that $\left\|\hat{\alpha}_{l+1} - \alpha_{l+1}^0\right\| = O_P[(\lambda + \delta_T) T / I_{l+1}^0 + (I_{l+1}^0)^{-1/2}]$. In subcase (ii2), using Lemma A.1(i) with $t = \hat{T}_l$ and $\hat{T}_{l+1}$, (2.2), and the triangle inequality yields

$$\begin{aligned}
\lambda &\geq \left\| \frac{1}{T} \sum_{r=\hat{T}_l}^{\hat{T}_{l+1}-1} x_r \left(y_r - x_r' \hat{\beta}_r\right) \right\| \\
&= \left\| \frac{1}{T} \sum_{r=\hat{T}_l}^{T_l^0-1} \left[x_r x_r' \left(\alpha_l^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] + \frac{1}{T} \sum_{r=T_l^0}^{T_{l+1}^0-1} \left[x_r x_r' \left(\alpha_{l+1}^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] \right. \\
&\quad \left. + \frac{1}{T} \sum_{r=T_{l+1}^0}^{\hat{T}_{l+1}-1} \left[x_r x_r' \left(\alpha_{l+2}^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] \right\| \\
&\geq \left\| \frac{1}{T} \sum_{r=T_l^0}^{T_{l+1}^0-1} \left[x_r x_r' \left(\alpha_{l+1}^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] \right\| - \left\| \frac{1}{T} \sum_{r=\hat{T}_l}^{T_l^0-1} \left[x_r x_r' \left(\alpha_l^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] \right\| \\
&\quad - \left\| \frac{1}{T} \sum_{r=T_{l+1}^0}^{\hat{T}_{l+1}-1} \left[x_r x_r' \left(\alpha_{l+2}^0 - \hat{\alpha}_{l+1}\right) + x_r u_r\right] \right\| \\
&\geq \frac{I_{l+1}^0}{T} \left\{ \underline{c}_{T,l} \left\|\hat{\alpha}_{l+1} - \alpha_{l+1}^0\right\| - O_P((I_{l+1}^0)^{-1/2}) \right\} - O_P\left((T_l^0 - \hat{T}_l)/T\right) - O_P\left((\hat{T}_{l+1} - T_{l+1}^0)/T\right)
\end{aligned}$$

where $\underline{c}_{T,l} \equiv \mu_{\min}\left(\frac{1}{I_{l+1}^0} \sum_{r=T_l^0}^{T_{l+1}^0-1} x_r x_r'\right) \geq \underline{c}_{xx}/2$ w.p.a.1. It follows that $\left\|\hat{\alpha}_{l+1} - \alpha_{l+1}^0\right\| = O_P[(\lambda + \delta_T) T / I_{l+1}^0 + (I_{l+1}^0)^{-1/2}]$. The same conclusion holds when $T_l^0 \leq \hat{T}_l < (T_l^0 + T_{l+1}^0)/2$. This implies that the result in part (ii) holds for all $j = 2, ..., m^0 + 1$.

To show (ii) holds for $j=1$, we apply Lemma A.1 with $t=\hat{T}_1$ and $t=1$ and the triangle inequality to obtain $\lambda \geq \left\|\frac{1}{T}\sum_{r=1}^{\hat{T}_1-1} x_r\left(y_r - x_r'\hat{\beta}_r\right)\right\| \geq \frac{\hat{I}_1}{T}\left[\bar{c}_{T,1a}\left\|\hat{\alpha}_1-\alpha_1^0\right\| - \left\|\frac{1}{\hat{I}_1}\sum_{r=1}^{\hat{T}_1-1} x_r u_r\right\|\right]$ if $\hat{T}_1 \leq T_1^0$, and $\lambda \geq \frac{I_1^0}{T}\left[\bar{c}_{T,1b}\left\|\hat{\alpha}_1-\alpha_1^0\right\| - \frac{1}{I_1^0}\sum_{r=T_1^0}^{\hat{T}_1-1}\|x_r u_r\|\left\|\hat{\alpha}_1-\alpha_2^0\right\| - \left\|\frac{1}{\hat{T}_1}\sum_{r=1}^{\hat{T}_1-1} x_r u_r\right\|\right]$ if $\hat{T}_1 > T_1^0$, where $\bar{c}_{T,1a} \equiv \mu_{\min}\left(\frac{1}{\hat{I}_1}\sum_{r=1}^{\hat{T}_1-1} x_r x_r'\right)$ and $\bar{c}_{T,1b} \equiv \mu_{\min}\left(\frac{1}{I_1^0}\sum_{r=1}^{T_1^0-1} x_r x_r'\right)$. One can readily show that $\frac{1}{\hat{T}_1}\sum_{r=1}^{\hat{T}_1-1} x_r u_r = \frac{1}{\hat{T}_1}\sum_{r=1}^{T_1^0-1} x_r u_r - \frac{1}{\hat{T}_1}\sum_{r=\hat{T}_1}^{T_1^0-1} x_r u_r = O_P[\left(I_1^0\right)^{-1/2}+\delta_T]$ if $\hat{T}_1 \leq T_1^0$ and $\frac{1}{\hat{T}_1}\sum_{r=1}^{\hat{T}_1-1} x_r u_r = \frac{1}{\hat{T}_1}\sum_{r=1}^{T_1^0-1} x_r u_r + \frac{1}{\hat{T}_1}\sum_{r=T_1^0}^{\hat{T}_1-1} x_r u_r = O_P[\left(I_1^0\right)^{-1/2}+\delta_T]$ if $\hat{T}_1 > T_1^0$ by using $\hat{T}_1/T = T_1^0/T + O_P(\delta_T)$. In addition, $\frac{1}{I_1^0}\sum_{r=T_1^0}^{\hat{T}_1-1}\|x_r u_r\| = O_P\left(T\delta_T/I_1^0\right)$. It follows that $\left\|\hat{\alpha}_1-\alpha_1^0\right\| = O_P[(\lambda+\delta_T)T/I_1^0 + \left(I_1^0\right)^{-1/2}]$. This completes the proof of part (ii). ■

# C Proof of Theorem 3.2

Given Theorem 3.1, it suffices to show that $P\left[\left\{\mathcal{D}(\hat{\mathcal{T}}_{\hat{m}}, \mathcal{T}_{m^0}^0) > T\delta_T\right\} \cap \left\{m_{\max} \geq \hat{m} > m^0\right\}\right] \to 0$ as $T\to\infty$. Define

$$
\begin{aligned}
F_{m,k,1} &= \left\{\forall l \in \{1,...,m\}, \left|\hat{T}_l - T_k^0\right| > T\delta_T \text{ and } \hat{T}_l < T_k^0\right\},\\
F_{m,k,2} &= \left\{\forall l \in \{1,...,m\}, \left|\hat{T}_l - T_k^0\right| > T\delta_T \text{ and } \hat{T}_l > T_k^0\right\}, \text{ and}\\
F_{m,k,3} &= \left\{\exists l \in \{1,...,m-1\}, \left|\hat{T}_l - T_k^0\right| > T\delta_T,\ \left|\hat{T}_{l+1} - T_k^0\right| > T\delta_T \text{ and } \hat{T}_l < T_k^0 < \hat{T}_{l+1}\right\}.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
& P\left[\left\{\mathcal{D}\left(\hat{\mathcal{T}}_{\hat{m}}, \mathcal{T}_{m^0}^0\right) > T\delta_T\right\} \cap \left\{m_{\max} \geq \hat{m} > m^0\right\}\right]\\
\leq & \sum_{m=m^0+1}^{m_{\max}} P\left[\left\{\mathcal{D}\left(\hat{\mathcal{T}}_m, \mathcal{T}_{m^0}^0\right) > T\delta_T\right\}\right] \leq \sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P\left(\forall l \in \{1,...,m\}, \left|\hat{T}_l - T_k^0\right| > T\delta_T\right)\\
= & \sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0}\left[P\left(F_{m,k,1}\right) + P\left(F_{m,k,2}\right) + P\left(F_{m,k,3}\right)\right].
\end{aligned}
$$

We first bound $\sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P\left(F_{m,k,1}\right)$. Note that $P\left(F_{m,k,1}\right) = P(F_{m,k,1}\cap\{\hat{T}_m > T_{k-1}^0\}) + P(F_{m,k,1}\cap\{\hat{T}_m \leq T_{k-1}^0\})$. Using Lemma A.1 with $t=\hat{T}_m$ and $t=T_k^0$ in the case where $T_k^0 \geq \hat{T}_m > T_{k-1}^0$ yields

$$
\frac{1}{T}\sum_{r=\hat{T}_m}^{T} x_r\left(y_r - x_r'\hat{\beta}_r\right) = \frac{\lambda}{2}\hat{\theta}_{\hat{T}_j}/\left\|\hat{\theta}_{\hat{T}_j}\right\| \text{ and } \frac{1}{T}\left\|\sum_{r=T_k^0}^{T} x_r'\left(y_r - x_r'\hat{\beta}_r\right)\right\| \leq \frac{\lambda}{2},
$$

implying that $\lambda \geq \left\|\frac{1}{T}\sum_{r=\hat{T}_m}^{T_k^0-1} x_r x_r'\left(\hat{\alpha}_{m+1}-\alpha_{k+1}^0\right) + \frac{1}{T}\sum_{r=\hat{T}_m}^{T_k^0-1} x_r x_r'\left(\alpha_{k+1}^0-\alpha_k^0\right) - \frac{1}{T}\sum_{r=\hat{T}_m}^{T_k^0-1} x_r u_r\right\|$. This further implies that the event

$$
G_T \equiv \left\{\left\|\alpha_{k+1}^0-\alpha_k^0\right\| \leq c_{4T,mk}^{-1}\left[\frac{\lambda T}{T_k^0-\hat{T}_m} + \left\|\frac{1}{T_k^0-\hat{T}_m}\sum_{r=\hat{T}_m}^{T_k^0-1} x_r x_r'\left(\hat{\alpha}_{m+1}-\alpha_{k+1}^0\right)\right\| + \left\|\frac{1}{T_k^0-\hat{T}_m}\sum_{r=\hat{T}_m}^{T_k^0-1} x_r u_r\right\|\right]\right\}
$$

occurs with probability one, where $c_{4T,mk} \equiv \mu_{\min}\left(\frac{1}{T_k^0-\hat{T}_m}\sum_{r=\hat{T}_m}^{T_k^0-1} x_r x_r'\right) \geq \underline{c}_{xx}/2$ w.p.a.1 by Lemma A.3(ii). It follows that $\sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P(F_{m,k,1}\cap\{\hat{T}_m > T_{k-1}^0\}) = \sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P(G_T \cap F_{m,k,1}\cap$

$\{\hat{T}_m > T^0_{k-1}\}) \leq F_1(1) + F_1(2) + F_1(3)$, where

$$
\begin{aligned}
F_1(1) &= \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left(c^{-1}_{4T,mk} \lambda \delta_T^{-1} \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/3\right) \\
F_1(2) &= \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left(c^{-1}_{4T,mk} \left\| \frac{1}{T^0_k - \hat{T}_m} \sum_{r=\hat{T}_m}^{T^0_k-1} x_r x_r' \left(\hat{\alpha}_{m+1} - \alpha^0_{k+1}\right) \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/3,\ T^0_k - \hat{T}_m > T\delta_T \right) \\
F_1(3) &= \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left(c^{-1}_{4T,mk} \left\| \frac{1}{T^0_k - \hat{T}_m} \sum_{r=\hat{T}_m}^{T^0_k-1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/3,\ T^0_k - \hat{T}_m > T\delta_T \right).
\end{aligned}
$$

Arguments like those used in the study of (B.8) show that $F_1(1)$ and $F_3(1)$ converge to 0. For $F_1(2)$, we apply Lemma A.1 with $t = T^0_k$ and $t = T^0_{k+1}$ and then the triangle inequality to obtain $\left\|\frac{1}{T}\sum_{r=T^0_k}^{T} x_r (y_r - x_r'\hat{\alpha}_{m+1})\right\| \leq \frac{\lambda}{2}$ and $\frac{1}{T}\left\|\sum_{r=T^0_{k+1}}^{T} x_r'(y_r - x_r'\hat{\alpha}_{m+1}\hat{\beta}_r)\right\| \leq \frac{\lambda}{2}$. This implies that $\lambda \geq \left\|\frac{1}{T}\sum_{r=T^0_k}^{T^0_{k+1}-1} x_r x_r' \left(\hat{\alpha}_{m+1} - \alpha^0_{k+1}\right) - \frac{1}{T}\sum_{r=T^0_k}^{T^0_{k+1}-1} x_r u_r\right\|$ and thus $\left\|\hat{\alpha}_{m+1} - \alpha^0_{k+1}\right\| \leq c^{-1}_{5T,k}[\frac{T\lambda}{T^0_{k+1}-T^0_k} + \frac{1}{T^0_{k+1}-T^0_k}||\sum_{r=T^0_k}^{T^0_{k+1}-1} x_r u_r||]$ where $c_{5T,k} \equiv \mu_{\min}\left(\frac{1}{T^0_{k+1}-T^0_k}\sum_{r=T^0_k}^{T^0_{k+1}-1} x_r x_r'\right) \geq \underline{c}_{xx}/2$ w.p.a.1 by Lemma A.3(ii). It follows that

$$
\begin{aligned}
F_1(2) &= \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left(c^{-1}_{4T,mk} \left\| \frac{1}{T} \sum_{r=\hat{T}_m}^{T^0_k-1} x_r x_r' \left(\hat{\alpha}_{m+1} - \alpha^0_{k+1}\right) \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/3,\ T^0_k - \hat{T}_m > T\delta_T \right) \\
&\leq \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left(c^{-1}_{4T,mk} \bar{c}_{2T,mk} \left\|\hat{\alpha}_{m+1} - \alpha^0_{k+1}\right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/3\right) \\
&\leq \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left(c^{-1}_{5T,k} \frac{\lambda T}{I_{\min}} \geq c_{4T,k}\bar{c}^{-1}_{2T,mk} \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/6\right) \\
&\quad + \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left(c^{-1}_{5T,k} \left\| \frac{1}{T^0_{k+1} - T^0_k} \sum_{r=T^0_k}^{T^0_{k+1}-1} x_r u_r \right\| \geq c_{4T,k}\bar{c}^{-1}_{2T,mk} \left\|\alpha^0_{k+1} - \alpha^0_k\right\|/6\right) \qquad \text{(C.1)} \\
&\to 0
\end{aligned}
$$

where $\bar{c}_{2T,mk} \equiv \mu_{\max}\left(\frac{1}{T^0_k - \hat{T}_m}\sum_{r=\hat{T}_m}^{T^0_k-1} x_r x_r'\right) \leq 2\bar{c}_{xx}$ w.p.a.1 by Lemma A.3(i), the first term in (C.1) converges to zero because $\lambda T/(I_{\min}J_{\min}) = o(1)$ by Assumptions A3(i) and (iii), and the second term converges to zero by the application of Lemma A.2. So we have shown that $\sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P(F_{m,k,1} \cap \{\hat{T}_m > T^0_{k-1}\}) \to 0$. Analogously, we can show that $\sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P(F_{m,k,1} \cap \{\hat{T}_m \leq T^0_{k-1}\}) \to 0$. It follows that $\sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P(F_{m,k,1}) \to 0$ as $T \to \infty$. Similarly, we can show that $\sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P(F_{m,k,2}) \to 0$ as $T \to \infty$.

Now we bound $\sum_{m=m^0+1}^{m_{\max}}\sum_{k=1}^{m^0} P(F_{m,k,3})$. Observe that $P(F_{m,k,3}) = P(F^{(1)}_{m,k,3}) + P(F^{(2)}_{m,k,3}) + P(F^{(3)}_{m,k,3}) + P(F^{(4)}_{m,k,3})$ where $F^{(1)}_{m,k,3} = F_{m,k,3} \cap \{T^0_{k-1} < \hat{T}_l < \hat{T}_{l+1} < T^0_{k+1}\}$, $F^{(2)}_{m,k,3} = F_{m,k,3} \cap \{T^0_{k-1} < \hat{T}_l < T^0_{k+1}, \hat{T}_{l+1} \geq T^0_{k+1}\}$, $F^{(3)}_{m,k,3} = F_{m,k,3} \cap \{\hat{T}_l \leq T^0_{k-1}, T^0_{k-1} < \hat{T}_{l+1} < T^0_{k+1}\}$, and $F^{(4)}_{m,k,3} = F_{m,k,3} \cap \{\hat{T}_l \leq$

$T^0_{k-1}, T^0_{k+1} < \hat{T}_{l+1}\}$. For $F^{(1)}_{m,k,3}$ we apply Lemma A.1 first with $t = T^0_k$ and $t = \hat{T}_l$ to obtain

$$\lambda > \left\| \frac{1}{T} \sum_{r=\hat{T}_l}^{T^0_k - 1} x_r x_r' \left( \hat{\alpha}_{l+1} - \alpha^0_k \right) - \frac{1}{T} \sum_{r=\hat{T}_l}^{T^0_k - 1} x_r u_r \right\|, \tag{C.2}$$

and then with $t = T^0_k$ and $t = \hat{T}_{l+1}$ to obtain

$$\lambda > \left\| \frac{1}{T} \sum_{r=T^0_k}^{\hat{T}_{l+1} - 1} x_r x_r' \left( \hat{\alpha}_{l+1} - \alpha^0_{k+1} \right) - \frac{1}{T} \sum_{r=T^0_k}^{\hat{T}_{l+1} - 1} x_r u_r \right\|. \tag{C.3}$$

Then by triangle inequality $\left\|\alpha^0_{k+1} - \alpha^0_k\right\| \leq \left\|\hat{\alpha}_{l+1} - \alpha^0_k\right\| + \left\|\hat{\alpha}_{l+1} - \alpha^0_{k+1}\right\| \leq c^{-1}_{6T,kl} \left( \left\| \frac{1}{T^0_k - \hat{T}_l} \sum_{r=\hat{T}_l}^{T^0_k - 1} x_r u_r \right\| + \frac{\lambda T}{T^0_k - \hat{T}_l} \right) + c^{-1}_{7T,kl} \left( \frac{\lambda T}{\hat{T}_{l+1} - T^0_k} + \frac{1}{\hat{T}_{l+1} - T^0_k} \sum_{r=T^0_k}^{\hat{T}_{l+1} - 1} x_r u_r \right)$ where $c_{6T,kl} \equiv \mu_{\min} \left( \frac{1}{T^0_k - \hat{T}_l} \sum_{r=\hat{T}_l}^{T^0_k - 1} x_r x_r' \right) \geq \underline{c}_{xx}/2$ w.p.a.1 and $c_{7T,kl} \equiv \mu_{\min} \left( \frac{1}{\hat{T}_{l+1} - T^0_k} \sum_{r=T^0_k}^{\hat{T}_{l+1} - 1} x_r x_r' \right) \geq \underline{c}_{xx}/2$ w.p.a.1 by Lemma A.3(ii). It follows that $\sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P(F^{(1)}_{m,k,3})$ is bounded from above by

$$\begin{aligned}
&\sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( \lambda \delta_T^{-1} \left( c^{-1}_{6T,kl} + c^{-1}_{7T,kl} \right) \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| / 3 \right) \\
&+ \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( c^{-1}_{6T,kl} \left\| \frac{1}{T^0_k - \hat{T}_l} \sum_{r=\hat{T}_l}^{T^0_k - 1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| / 3,\ T^0_k - \hat{T}_l \geq T\delta_T \right) \\
&+ \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( c^{-1}_{7T,kl} \left\| \frac{1}{\hat{T}_{l+1} - T^0_k} \sum_{r=T^0_k}^{\hat{T}_{l+1} - 1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| / 3,\ \hat{T}_{l+1} - T^0_k \geq T\delta_T \right)
\end{aligned}$$

which converges to zero by arguments analogous to those used in the study of (B.8). For $F^{(2)}_{m,k,3}$ we apply Lemma A.1 first with $t = T^0_k$ and $t = \hat{T}_l$ to obtain (C.2) and then with $t = T^0_k$ and $t = T^0_{k+1}$ to obtain

$$\lambda > \left\| \frac{1}{T} \sum_{r=T^0_k}^{T^0_{k+1} - 1} x_r x_r' \left( \hat{\alpha}_{l+1} - \alpha^0_{k+1} \right) - \frac{1}{T} \sum_{r=T^0_k}^{T^0_{k+1} - 1} x_r u_r \right\|. \tag{C.4}$$

Then by the triangle inequality $\left\|\alpha^0_{k+1} - \alpha^0_k\right\| \leq \left\|\hat{\alpha}_{l+1} - \alpha^0_k\right\| + \left\|\hat{\alpha}_{l+1} - \alpha^0_{k+1}\right\| \leq c^{-1}_{6T,kl} \left( \left\| \frac{1}{T^0_k - \hat{T}_l} \sum_{r=\hat{T}_l}^{T^0_k} x_r u_r \right\| + \frac{\lambda T}{T^0_k - \hat{T}_l} \right) + c^{-1}_{8T,k} \left( \frac{\lambda T}{T^0_{k+1} - T^0_k} + \frac{1}{T^0_{k+1} - T^0_k} \sum_{r=T^0_k}^{T^0_{k+1} - 1} x_r u_r \right)$ where $c_{8T,k} \equiv \mu_{\min} \left( \frac{1}{T^0_{k+1} - T^0_k} \sum_{r=T^0_k}^{T^0_{k+1} - 1} x_r x_r' \right) \geq \underline{c}_{xx}/2$ w.p.a.1 by Lemma A.3(ii). It follows that $\sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P(F^{(2)}_{m,k,3})$ is bounded from above by

$$\begin{aligned}
&\sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( \lambda \delta_T^{-1} c^{-1}_{6T,kl} + \frac{\lambda T}{I_{\min}} c^{-1}_{8T,k} \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| / 3 \right) \\
&+ \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( c^{-1}_{6T,kl} \left\| \frac{1}{T^0_k - \hat{T}_l} \sum_{r=\hat{T}_l}^{T^0_k - 1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| / 3,\ T^0_k - \hat{T}_l \geq T\delta_T \right) \\
&+ \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( c^{-1}_{8T,k} \left\| \frac{1}{T^0_{k+1} - T^0_k} \sum_{r=T^0_k}^{T^0_{k+1} - 1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| / 3 \right),
\end{aligned}$$

which converges to zero by arguments analogous to those used in the study of (B.8). For $F^{(3)}_{m,k,3}$ we apply Lemma A.1 first with $t = T^0_{k-1}$ and $t = T^0_k$ to obtain

$$\lambda > \left\| \frac{1}{T} \sum_{r=T^0_{k-1}}^{T^0_k-1} x_r x_r' \left( \hat{\alpha}_{l+1} - \alpha^0_k \right) - \frac{1}{T} \sum_{r=T^0_{k-1}}^{T^0_k-1} x_r u_r \right\|, \tag{C.5}$$

and then with $t = T^0_k$ and $t = \hat{T}_{l+1}$ to obtain

$$\lambda > \left\| \frac{1}{T} \sum_{r=T^0_k}^{\hat{T}_{l+1}-1} x_r x_r' \left( \hat{\alpha}_{l+1} - \alpha^0_{k+1} \right) - \frac{1}{T} \sum_{r=T^0_k}^{\hat{T}_{l+1}-1} x_r u_r \right\|. \tag{C.6}$$

Then by triangle inequality $\left\|\alpha^0_{k+1} - \alpha^0_k\right\| \leq \left\|\hat{\alpha}_{l+1} - \alpha^0_k\right\| + \left\|\hat{\alpha}_{l+1} - \alpha^0_{k+1}\right\| \leq c^{-1}_{9T,k} \left( \left\| \frac{1}{T^0_k - T^0_{k-1}} \sum_{r=T^0_{k-1}}^{T^0_k-1} x_r u_r \right\| + \frac{\lambda T}{T^0_k - T^0_{k-1}} \right) + c^{-1}_{10T,kl} \left( \frac{\lambda T}{\hat{T}_{l+1} - T^0_k} + \frac{1}{\hat{T}_{l+1} - T^0_k} \sum_{r=T^0_k}^{\hat{T}_{l+1}-1} x_r u_r \right)$ where $c_{9T,k} \equiv \mu_{\min} \left( \frac{1}{T^0_k - T^0_{k-1}} \sum_{r=T^0_{k-1}}^{T^0_k-1} x_r x_r' \right) \geq \underline{c}_{xx}/2$ w.p.a.1 and $c_{10T,kl} \equiv \mu_{\min} \left( \frac{1}{\hat{T}_{l+1} - T^0_k} \sum_{r=T^0_k}^{\hat{T}_{l+1}-1} x_r x_r' \right) \geq \underline{c}_{xx}/2$ w.p.a.1 by Lemma A.3(ii). It follows that $\sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P(F^{(3)}_{m,k,3})$ is bounded from above by

$$\begin{aligned}
&\sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( \frac{\lambda T}{I_{\min}} c^{-1}_{9T,k} + \lambda \delta_T^{-1} c^{-1}_{10T,kl} \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| /3 \right) \\
&+ \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( c^{-1}_{9T,k} \left\| \frac{1}{T^0_k - T^0_{k-1}} \sum_{r=T^0_{k-1}}^{T^0_k-1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| /3 \right) \\
&+ \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( c^{-1}_{10T,kl} \left\| \frac{1}{\hat{T}_{l+1} - T^0_k} \sum_{r=T^0_k}^{\hat{T}_{l+1}-1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| /3,\ \hat{T}_{l+1} - T^0_k \geq T\delta_T \right),
\end{aligned}$$

which converges to zero by arguments analogous to those used in the study of (B.8). For $F^{(4)}_{m,k,3}$ we apply Lemma A.1 first with $t = T^0_{k-1}$ and $t = T^0_k$ to obtain (C.5) and then with $t = T^0_k$ and $t = T^0_{k+1}$ to obtain (C.4). Then by the triangle inequality $\left\|\alpha^0_{k+1} - \alpha^0_k\right\| \leq \left\|\hat{\alpha}_{l+1} - \alpha^0_k\right\| + \left\|\hat{\alpha}_{l+1} - \alpha^0_{k+1}\right\| \leq c^{-1}_{9T,k} \left( \frac{\lambda T}{T^0_k - T^0_{k-1}} + \left\| \frac{1}{T^0_k - T^0_{k-1}} \sum_{r=T^0_{k-1}}^{T^0_k-1} x_r u_r \right\| \right) + c^{-1}_{8T,kl} \left( \frac{\lambda T}{T^0_{k+1} - T^0_k} + \frac{1}{T^0_{k+1} - T^0_k} \sum_{r=T^0_k}^{T^0_{k+1}-1} x_r u_r \right)$. It follows that $\sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P(F^{(4)}_{m,k,3})$ is bounded from above by

$$\begin{aligned}
&\sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( \frac{\lambda T}{I_{\min}} \left( c^{-1}_{9T,k} + c^{-1}_{8T,kl} \right) \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| /3 \right) \\
&+ \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( c^{-1}_{9T,k} \left\| \frac{1}{T^0_k - T^0_{k-1}} \sum_{r=T^0_{k-1}}^{T^0_k-1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| /3 \right) \\
&+ \sum_{m=m^0+1}^{m_{\max}} \sum_{k=1}^{m^0} P\left( c^{-1}_{8T,kl} \left\| \frac{1}{T^0_{k+1} - T^0_k} \sum_{r=T^0_k}^{T^0_{k+1}-1} x_r u_r \right\| \geq \left\|\alpha^0_{k+1} - \alpha^0_k\right\| /3 \right)
\end{aligned}$$

which converges to zero by arguments analogous to those used in the study of (B.8). Consequently, $P\left[ \left\{ \mathcal{D}(\hat{\mathcal{T}}_{\hat{m}}, \mathcal{T}^0_{m^0}) > T\delta_T \right\} \cap \left\{ m_{\max} \geq \hat{m} > m^0 \right\} \right] \to 0$ as $T \to \infty$. ■

# D Proof of Theorem 3.3

To avoid confusion of notation, let $\breve{m}$, $\breve{\mathcal{T}}_{\breve{m}}$=$(\breve{T}_1, ..., \breve{T}_{\breve{m}})$, and $\hat{\boldsymbol{\alpha}}_{\breve{m}}(\breve{\mathcal{T}}_{\breve{m}}) = (\hat{\alpha}_1(\breve{\mathcal{T}}_{\breve{m}})', ..., \hat{\alpha}_{\breve{m}+1}(\breve{\mathcal{T}}_{\breve{m}})')'$ be the hypothesized GFL estimates of the number of breaks, the set of break points, and the set of regression coefficient estimates, respectively. Let $\hat{\boldsymbol{\alpha}}_{\breve{\mathcal{T}}_{\breve{m}}}$ be the corresponding set of post-Lasso OLS estimates. Let $Q_{T\lambda}(\cdot;\cdot)$ and $Q_{T,1}(\cdot;\cdot)$ be as defined in (2.5) and 3.1, respectively. Let $\hat{\boldsymbol{\alpha}}_{\mathcal{T}_m} = (\hat{\alpha}'_{1,\mathcal{T}_m}, ..., \hat{\alpha}'_{m+1,\mathcal{T}_m})' \equiv \arg\min_{\boldsymbol{\alpha}_m} Q_{T,1}(\boldsymbol{\alpha}_m; \mathcal{T}_m)$ denote the post-Lasso OLS estimate of $\boldsymbol{\alpha}_m = \left(\alpha'_1, ..., \alpha'_{m+1}\right)'$ for the given set of break dates specified in $\mathcal{T}_m$. We want to show that for any $\breve{m} < m^0$ we have $P(Q_{T\lambda}(\hat{\boldsymbol{\alpha}}_{\breve{m}}(\breve{\mathcal{T}}_{\breve{m}}); \breve{\mathcal{T}}_{\breve{m}}) > Q_{T\lambda}(\hat{\boldsymbol{\alpha}}_{m^0}(\hat{\mathcal{T}}_{m^0}); \hat{\mathcal{T}}_{m^0})) \to 1$. Noting that under Assumption A3(iv)

$$
\begin{aligned}
& \frac{T}{I_{\min} J^2_{\min}} \left[ Q_{T\lambda}\left(\hat{\boldsymbol{\alpha}}_{\breve{m}}\left(\breve{\mathcal{T}}_{\breve{m}}\right); \breve{\mathcal{T}}_{\breve{m}}\right) - Q_{T\lambda}\left(\hat{\boldsymbol{\alpha}}_{m^0}\left(\hat{\mathcal{T}}_{m^0}\right); \hat{\mathcal{T}}_{m^0}\right)\right] \\
= & \frac{T}{I_{\min} J^2_{\min}} \left\{ \frac{1}{T} \sum_{j=1}^{\breve{m}+1} \sum_{t=\breve{T}_{j-1}}^{\breve{T}_j - 1} \left[ y_t - \hat{\alpha}_j\left(\breve{\mathcal{T}}_{\breve{m}}\right)' x_t \right]^2 - \frac{1}{T} \sum_{j=1}^{m^0+1} \sum_{t=\hat{T}_{j-1}}^{\hat{T}_j - 1} \left[ y_t - \hat{\alpha}_j\left(\hat{\mathcal{T}}_{m^0}\right)' x_t \right]^2 \right\} \\
& + \frac{\lambda T}{I_{\min} J^2_{\min}} \left\{ \sum_{j=1}^{\breve{m}} \left\| \hat{\alpha}_{j+1}\left(\breve{\mathcal{T}}_{\breve{m}}\right) - \hat{\alpha}_j\left(\breve{\mathcal{T}}_{\breve{m}}\right) \right\| - \sum_{j=1}^{m^0} \left\| \hat{\alpha}_{j+1}\left(\hat{\mathcal{T}}_{m^0}\right) - \hat{\alpha}_j\left(\hat{\mathcal{T}}_{m^0}\right) \right\| \right\} \\
\geq & \frac{T}{I_{\min} J^2_{\min}} \left[ Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{\breve{\mathcal{T}}_{\breve{m}}}; \breve{\mathcal{T}}_{\breve{m}}\right) - Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{m^0}\left(\hat{\mathcal{T}}_{m^0}\right); \hat{\mathcal{T}}_{m^0}\right)\right] + o_P(1)
\end{aligned}
$$

it suffices to show that for some $c > 0$

$$
P\left( \inf_{0 \leq m < m^0} \inf_{\mathcal{T}_m} \frac{T}{I_{\min} J^2_{\min}} \left[ Q_{T,1}(\hat{\boldsymbol{\alpha}}_{\mathcal{T}_m}; \mathcal{T}_m) - Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{m^0}\left(\hat{\mathcal{T}}_{m^0}\right); \hat{\mathcal{T}}_{m^0}\right)\right] > c + o_P(1) \right) \to 1, \tag{D.1}
$$

where $\mathcal{T}_m = (T_1, ..., T_m)$ with $1 < T_1 < ... < T_m < T$ denotes an arbitrary $m$-dimensional set of potential break dates. We prove (D.1) by showing that (i) $\frac{T}{I_{\min} J^2_{\min}} \left[ Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{m^0}\left(\hat{\mathcal{T}}_{m^0}\right); \hat{\mathcal{T}}_{m^0}\right) - \bar{\sigma}^2_T \right] = o_P(1)$, and (ii) $P\left( \inf_{0 \leq m < m^0} \inf_{\mathcal{T}_m} \frac{T}{I_{\min} J^2_{\min}} \left[ Q_{T,1}(\hat{\boldsymbol{\alpha}}_{\mathcal{T}_m}; \mathcal{T}_m) - \bar{\sigma}^2_T \right] \geq c + o_P(1) \right) \to 1$ as $T \to \infty$, where $\bar{\sigma}^2_T \equiv \frac{1}{T} \sum_{j=1}^{m^0+1} \sum_{t=T^0_{j-1}}^{T^0_j - 1} \left( y_t - \alpha^{0\prime}_j x_t \right)^2 = \frac{1}{T} \sum_{t=1}^{T} u^2_t$.

We first show (i). We make the following decomposition:

$$
Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{m^0}\left(\hat{\mathcal{T}}_{m^0}\right); \hat{\mathcal{T}}_{m^0}\right) - \bar{\sigma}^2_T = \sum_{j=1}^{m^0+1} \frac{1}{T} \sum_{t=\hat{T}_{j-1}}^{\hat{T}_j - 1} \left[ \left( y_t - \hat{\alpha}'_j x_t \right)^2 - u^2_t \right] \equiv \sum_{j=1}^{m^0+1} Q_{T,1j}, \text{ say.}
$$

To study $Q_{T,1j}$, we consider four subcases: (i1) $\hat{T}_{j-1} < T^0_{j-1}$ and $\hat{T}_j < T^0_j$, (i2) $\hat{T}_{j-1} < T^0_{j-1}$ and $\hat{T}_j \geq T^0_j$, (i3) $\hat{T}_{j-1} \geq T^0_{j-1}$ and $\hat{T}_j < T^0_j$, and (i4) $\hat{T}_{j-1} \geq T^0_{j-1}$ and $\hat{T}_j \geq T^0_j$. In subcase (i1), we have

$$
\begin{aligned}
Q_{T,1j} &= \frac{1}{T} \sum_{t=T^0_{j-1}}^{T^0_j - 1} \left[ \left( y_t - \hat{\alpha}'_j x_t \right)^2 - u^2_t \right] + \frac{1}{T} \sum_{t=\hat{T}_{j-1}}^{T^0_{j-1} - 1} \left[ \left( y_t - \hat{\alpha}'_j x_t \right)^2 - u^2_t \right] - \frac{1}{T} \sum_{r=\hat{T}_j}^{T^0_j - 1} \left[ \left( y_t - \hat{\alpha}'_j x_t \right)^2 - u^2_t \right] \\
&\equiv Q_{T,1j}(1) + Q_{T,1j}(2) - Q_{T,1j}(3), \text{ say.}
\end{aligned}
$$

By the fact that $\frac{1}{I^0_j} \sum_{t=T^0_{j-1}}^{T^0_j - 1} x_t u_t = O_P\left( (I^0_j / (\log m^0)^{c_\delta})^{-1/2} \right)$ and $\frac{1}{I^0_j} \sum_{t=T^0_{j-1}}^{T^0_j - 1} x_t x'_t = O_P(1)$ uniformly

in $j$, we have

$$\begin{aligned} Q_{T,1j}(1) &= -2\left(\hat{\alpha}_j-\alpha_j^0\right)'\frac{1}{T}\sum_{t=T_{j-1}^0}^{T_j^0-1}x_tu_t+\left(\hat{\alpha}_j-\alpha_j^0\right)'\frac{1}{T}\sum_{t=T_{j-1}^0}^{T_j^0-1}x_tx_t'\left(\hat{\alpha}_j-\alpha_j^0\right) \\ &= -\left\|\hat{\alpha}_j-\alpha_j^0\right\|O_P\left(((\log m^0)^{c_\delta}/T)^{1/2}\right)+\left\|\hat{\alpha}_j-\alpha_j^0\right\|^2O_P(1) \text{ uniformly in } j. \end{aligned}$$

For $Q_{T,1j}(2)$, we have $Q_{T,1j}(2)=-2\left(\hat{\alpha}_j-\alpha_j^0\right)'\frac{1}{T}\sum_{t=\hat{T}_{j-1}}^{T_j^0-1}x_tu_t+\left(\hat{\alpha}_j-\alpha_j^0\right)'\frac{1}{T}\sum_{t=\hat{T}_{j-1}}^{T_j^0-1}x_tx_t'\left(\hat{\alpha}_j-\alpha_j^0\right)\equiv -2Q_{T,1j}(2,1)+Q_{T,1j}(2,2)$, say. By Theorem 3.1(i) and Markov inequality, w.p.a.1 we have that uniformly in $j$,

$$\begin{aligned} Q_{T,1j}(2,1) &\le \left\|\hat{\alpha}_j-\alpha_j^0\right\|\frac{1}{T}\sum_{t=\hat{T}_{j-1}}^{T_j^0-1}\|x_tu_t\|\le\delta_T\left\|\hat{\alpha}_j-\alpha_j^0\right\|\frac{1}{T\delta_T}\sum_{t=T_j^0-T\delta_T}^{T_j^0-1}\|x_tu_t\|=\delta_T\left\|\hat{\alpha}_j-\alpha_j^0\right\|O_P(1), \\ Q_{T,1j}(2,2) &\le \delta_T\left\|\hat{\alpha}_j-\alpha_j^0\right\|^2\mu_{\max}\left(\frac{1}{T\delta_T}\sum_{t=T_j^0-T\delta_T}^{T_j^0-1}x_tx_t'\right)=\delta_T\left\|\hat{\alpha}_j-\alpha_j^0\right\|^2O_P(1). \end{aligned}$$

It follows that $Q_{T,1j}(2)=\delta_T(\left\|\hat{\alpha}_j-\alpha_j^0\right\|+\left\|\hat{\alpha}_j-\alpha_j^0\right\|^2)O_P(1)$ uniformly in $j$. Similarly, we can show that $Q_{T,1j}(3)=\delta_T(\left\|\hat{\alpha}_j-\alpha_j^0\right\|+\left\|\hat{\alpha}_j-\alpha_j^0\right\|^2)O_P(1)$ uniformly in $j$. Consequently, we have

$$Q_{T,1j}=O_P(1)\left\{(T^{-1/2}(\log m^0)^{c_\delta/2}+\delta_T)\left\|\hat{\alpha}_j-\alpha_j^0\right\|+\delta_T\left\|\hat{\alpha}_j-\alpha_j^0\right\|^2\right\} \text{ in subcase (i1).}$$

Analogously, we can show that this result also holds in subcases (i2)-(i4). Using the bounds for $\left\|\hat{\alpha}_j-\alpha_j^0\right\|$ in the proof of Theorem 3.1(ii), we can readily show that

$$\begin{aligned} \frac{T}{I_{\min}J_{\min}^2}\sum_{j=1}^{m^0+1}Q_{T,1j} &= \frac{T}{I_{\min}J_{\min}^2}\left\{O_P((\log m^0)^{c_\delta/2}T^{-1/2}+\delta_T)\sum_{j=1}^{m^0+1}\left\|\hat{\alpha}_j-\alpha_j^0\right\|+\delta_T\sum_{j=1}^{m^0+1}\left\|\hat{\alpha}_j-\alpha_j^0\right\|^2\right\} \\ &= \frac{Tm^0}{I_{\min}J_{\min}^2}((\log m^0)^{c_\delta/2}T^{-1/2}+\delta_T)O_P\left((\lambda+\delta_T)T/I_{\min}+I_{\min}^{-1/2}\right)=o_P(1) \end{aligned}$$

under Assumption A3(iv) in subcase (i1). It follows that $\frac{T}{I_{\min}J_{\min}^2}\left[Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{m^0}\left(\hat{\mathcal{T}}_{m^0}\right);\hat{\mathcal{T}}_{m^0}\right)-\bar{\sigma}_T^2\right]=o_P(1)$.

We now show (ii). For brevity we assume that $m^0=1$ and $\hat{T}_1<T_1^0$ below as the other cases can be studied analogously. In this case, $m=0$ and $\mathcal{T}_0$ is empty. Then $\hat{\boldsymbol{\alpha}}_{\mathcal{T}_0}$ reduces to the OLS estimate of $y_t$ on $x_t$ using all $T$ observations and we have $\hat{\boldsymbol{\alpha}}_{\mathcal{T}_0}=\hat{\alpha}_{OLS}\equiv\left(\frac{1}{T}\sum_{t=1}^Tx_tx_t'\right)^{-1}\frac{1}{T}\sum_{t=1}^Tx_ty_t$. Using (2.2) with $m_0=1$ yields

$$\begin{aligned} \hat{\alpha}_{OLS} &= \left(\frac{1}{T}\sum_{t=1}^Tx_tx_t'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T_1^0-1}x_tx_t'\alpha_1^0+\left(\frac{1}{T}\sum_{t=1}^Tx_tx_t'\right)^{-1}\frac{1}{T}\sum_{t=T_1^0}^Tx_tx_t'\alpha_2^0+\left(\frac{1}{T}\sum_{t=1}^Tx_tx_t'\right)^{-1}\frac{1}{T}\sum_{t=1}^Tx_tu_t \\ &= \alpha_T^*+O_P\left(T^{-1/2}\right), \text{ say.} \end{aligned}$$

where $\alpha_T^*\equiv\frac{I_1^0}{T}Q_{xx}^{-1}Q_{1xx}\alpha_1^0+\frac{I_2^0}{T}Q_{xx}^{-1}Q_{2xx}\alpha_2^0=O_P(1)$, $Q_{xx}=\frac{1}{T}\sum_{t=1}^Tx_tx_t'$, $Q_{1xx}=\frac{1}{I_1^0}\sum_{t=1}^{T_1^0-1}x_tx_t'$, and $Q_{2xx}=\frac{1}{I_2^0}\sum_{t=T_1^0}^Tx_tx_t'$. Note that $Q_{xx}$, $Q_{1xx}$, and $Q_{2xx}$ are all asymptotically nonsingular by Lemma

A.3. Let $d_1^0 \equiv \alpha_2^0 - \alpha_1^0$. Then $\alpha_T^* - \alpha_1^0 = \frac{I_2^0}{T} Q_{xx}^{-1} Q_{2xx} d_1^0$, and $\alpha_T^* - \alpha_2^0 = -\frac{I_1^0}{T} Q_{xx}^{-1} Q_{1xx} d_1^0$. Using this we can readily show that

$$\begin{aligned} Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{\mathcal{T}_0}\right) - \bar{\sigma}_T^2 &= \frac{1}{T}\sum_{t=1}^{T}\left[\left(y_t - \hat{\alpha}_{OLS}' x_t\right)^2 - u_t^2\right] \\ &= \left(\alpha_1^0 - \hat{\alpha}_{OLS}\right)' \frac{1}{T}\sum_{t=1}^{T_1^0-1} x_t x_t' \left(\alpha_1^0 - \hat{\alpha}_{OLS}\right) + 2\left(\alpha_1^0 - \hat{\alpha}_{OLS}\right)' \frac{1}{T}\sum_{t=1}^{T_1^0-1} x_t u_t \\ &\quad + \left(\alpha_2^0 - \hat{\alpha}_{OLS}\right)' \frac{1}{T}\sum_{t=T_1^0}^{T} x_t x_t' \left(\alpha_2^0 - \hat{\alpha}_{OLS}\right) + 2\left(\alpha_2^0 - \hat{\alpha}_{OLS}\right)' \frac{1}{T}\sum_{t=T_1^0}^{T} x_t u_t \\ &= c_T + O_P\left(T^{-1/2}\left(T^{-1/2} + \left\|d_1^0\right\|\right)\right) = c_T + O_P\left(T^{-1/2}\right), \end{aligned}$$

where the leading term is given by

$$\begin{aligned} c_T &\equiv \frac{I_1^0}{T}\left(\alpha_T^* - \alpha_1^0\right)' Q_{1xx}\left(\alpha_T^* - \alpha_1^0\right) + \frac{I_2^0}{T}\left(\alpha_T^* - \alpha_2^0\right)' Q_{2xx}\left(\alpha_T^* - \alpha_2^0\right) \\ &= \frac{I_1^0 I_2^0}{T^2}\left[\frac{I_2^0}{T} d_1^{0\prime} Q_{2xx} Q_{xx}^{-1} Q_{1xx} Q_{xx}^{-1} Q_{2xx} d_1^0 + \frac{I_1^0}{T} d_1^{0\prime} Q_{1xx} Q_{xx}^{-1} Q_{2xx} Q_{xx}^{-1} Q_{1xx} d_1^0\right] \\ &\geq cI_{\min} J_{\min}^2 / T \text{ for some } c > 0. \end{aligned}$$

It follows that $\frac{T}{I_{\min} J_{\min}^2}\left[Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{\mathcal{T}_m}; \mathcal{T}_m\right) - \bar{\sigma}_T^2\right] \geq c + \frac{T}{I_{\min} J_{\min}^2} O_P\left(T^{-1/2}\right) = c + o_P(1)$ by Assumption A3(iv). This completes the proof of (ii) for the case $m^0 = 1$. Analogous but more tedious arguments show that (ii) also holds for the general case where $m^0 \geq 2$. ■

# E Proof of Theorem 3.4

Denote $\Omega = [0, \lambda_{\max}]$, a bounded interval in $\mathbb{R}^+$. We divide $\Omega$ into three subsets $\Omega_0$, $\Omega_-$ and $\Omega_+$ as follows

$$\Omega_0 = \left\{\lambda \in \Omega : \hat{m}_\lambda = m^0\right\}, \ \Omega_- = \left\{\lambda \in \Omega : \hat{m}_\lambda < m^0\right\}, \text{ and } \Omega_+ = \left\{\lambda \in \Omega : \hat{m}_\lambda > m^0\right\}.$$

Clearly, $\Omega_0, \Omega_-$ and $\Omega_+$ denote the three subsets of $\Omega$ in which the correct-, under- and over-number of breaks are selected by the GFL, respectively. Recall $\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{\hat{m}_\lambda}} = (\hat{\alpha}_{1,\hat{\mathcal{T}}_{\hat{m}_\lambda}}', ..., \hat{\alpha}_{\hat{m}_\lambda+1,\hat{\mathcal{T}}_{\hat{m}_\lambda}}')'$ denotes the set of post-Lasso OLS estimates of the regression coefficients based on the break dates in $\hat{\mathcal{T}}_{\hat{m}_\lambda} = \hat{\mathcal{T}}_{\hat{m}_\lambda}(\lambda) = (\hat{T}_1(\lambda), ..., \hat{T}_{\hat{m}_\lambda}(\lambda))$, where we make the dependence of various estimates on $\lambda$ explicit when necessary. Let $\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}} \equiv Q_{T,1}(\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{\hat{m}_\lambda}}; \hat{\mathcal{T}}_{\hat{m}_\lambda})$. Let $\lambda_T^0$ denote an element in $\Omega_0$ that also satisfies the conditions on $\lambda$ in Assumptions A3(iii)-(iv). For any $\lambda_T^0 \in \Omega_0$, we have $\hat{m}_{\lambda_T^0} = m^0$ and $\left|\hat{T}_j\left(\lambda_T^0\right) - T_j^0\right| \leq T\delta_T$ for $j = 1, ..., m^0$ by Theorem 3.1 as $\lambda_T^0$ also satisfies Assumptions A3(iii)-(iv). By the proof of Theorem 3.3, $\hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}} = \bar{\sigma}_T^2 + (\log m^0)^{c\delta/2} T^{-1/2} + \delta_T) O_P((\lambda + \delta_T) T / I_{\min} + I_{\min}^{-1/2})$, where $\bar{\sigma}_T^2 \equiv \frac{1}{T}\sum_{t=1}^{T} u_t^2 \xrightarrow{P} \sigma_0^2 \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left(u_t^2\right)$ under Assumption A1. Then by Assumption A4 and Slutsky lemma, $IC\left(\lambda_T^0\right) = \log(\hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}) + \rho_T p m^0 = \log(\hat{\sigma}^2_{\mathcal{T}^0_{m^0}}) + o_P(1) \xrightarrow{P} \log\left(\sigma_0^2\right)$. We consider the case of under- and over-fitted models separately.

*Case 1: Under-fitted model.* In this case, $\hat{m}_\lambda < m^0$ and by the proof of Theorem 3.3

$$\frac{T}{I_{\min} J_{\min}^2} \inf_{\lambda\in\Omega_-}\left[\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}} - \hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}\right] = \frac{T}{I_{\min} J_{\min}^2} \inf_{\lambda\in\Omega_-}\left[Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{\hat{m}_\lambda}}; \hat{\mathcal{T}}_{\hat{m}_\lambda}\right) - Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{m^0}\left(\hat{\mathcal{T}}_{m^0}\right); \hat{\mathcal{T}}_{m^0}\right)\right] \geq c + o_P(1)$$

for some $c > 0$. It follows that

$$\begin{aligned} P\left(\inf_{\lambda\in\Omega_-} IC(\lambda) > IC\left(\lambda_T^0\right)\right) &= P\left(\frac{T}{I_{\min}J_{\min}^2}\left[\log\left(\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}}/\hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}\right) + \rho_T p\left(\hat{m}_\lambda - m\right)\right] > 0\right) \\ &= P\left(\frac{T}{I_{\min}J_{\min}^2}\log\left(\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}}/\hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}\right) + o_P(1) > 0\right) \to 1. \end{aligned} \tag{E.1}$$

*Case 2: Over-fitted model.* Let $\lambda \in \Omega_+$. By the fact that $\log(1+x) = x + O\left(x^2\right)$ for $x$ in the neighborhood of 0, and Lemma E.1,

$$\begin{aligned} \delta_T^{-1}\left[IC(\lambda) - IC\left(\lambda_T^0\right)\right] &= \delta_T^{-1}\log\left(\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}}/\hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}\right) + \delta_T^{-1}\rho_T p\left(\hat{m}_\lambda - m^0\right) \\ &= \left(\hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}\right)^{-1}\delta_T^{-1}\left(\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}} - \hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}\right) + \delta_T^{-1}\rho_T p\left(\hat{m}_\lambda - m^0\right) + o_P(1). \end{aligned}$$

Noting that $\delta_T^{-1}(\hat{\sigma}^2_{\hat{\mathcal{T}}_{\hat{m}_\lambda}} - \hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}) = O_P(1)$ by Lemma E.1, $\hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}} = \sigma_0^2 + o_P(1)$, and $\delta_T^{-1}\rho_T \to \infty$ by Assumption A4(ii), we have

$$\begin{aligned} & P\left(\inf_{\lambda\in\Omega_+} IC(\lambda) > IC\left(\lambda_T^0\right)\right) \\ \geq\ & P\left(\left(\sigma_0^2\right)^{-1}\min_{m^0<m\leq m_{\max}}\inf_{\mathcal{T}_m:\mathcal{D}\left(\mathcal{T}_m,\mathcal{T}_{m^0}^0\right)\leq T\delta_T}\left[\delta_T^{-1}\left(\hat{\sigma}^2_{\mathcal{T}_m} - \hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}\right) + \delta_T^{-1}\rho_T p\left(m - m^0\right)\right] + o_P(1) > 0\right) \\ \to\ & 1 \text{ as } T\to\infty. \end{aligned} \tag{E.2}$$

Combining (E.1) with (E.2) yields

$$P\left(\inf_{\lambda\in\Omega_-\cup\Omega_+} IC(\lambda) > IC\left(\lambda_T^0\right)\right) \to 1 \text{ as } T\to\infty. \tag{E.3}$$

This implies that the minimizer $\hat{\lambda}$ of $IC(\lambda)$ cannot belong to either $\Omega_-$ or $\Omega_+$. Consequently, we have $P(\hat{\lambda}\in\Omega_0) = P(\hat{m}_{\hat{\lambda}} = m_0) \to 1$ as $T\to\infty$. ■

**Lemma E.1** $\max_{m^0<m\leq m_{\max}}\sup_{\mathcal{T}_m\in\mathbb{T}_m}\delta_T^{-1}\left|\hat{\sigma}^2_{\mathcal{T}_m} - \hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}\right| = O_P(1)$, *where* $\mathbb{T}_m = \{\mathcal{T}_m = (T_1,...,T_m) : 1 < T_1 < ... < T_m < T,\ \mathcal{D}\left(\mathcal{T}_m,\mathcal{T}_{m^0}^0\right) \leq T\delta_T\}$.

**Proof.** Noting that $|\hat{\sigma}^2_{\mathcal{T}_m} - \hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}}| \leq |\hat{\sigma}^2_{\mathcal{T}_m} - \bar{\sigma}^2_T| + |\hat{\sigma}^2_{\hat{\mathcal{T}}_{m^0}} - \bar{\sigma}^2_T|$ and $\mathcal{D}(\hat{\mathcal{T}}_{m^0},\mathcal{T}_{m^0}^0) \leq T\delta_T$ w.p.a.1 by Theorem 3.1(i), it suffices to show that $\max_{m^0\leq m\leq m_{\max}}\sup_{\mathcal{T}_m\in\mathbb{T}_m}\delta_T^{-1}\left|\hat{\sigma}^2_{\mathcal{T}_m} - \bar{\sigma}^2_T\right| = O_P(1)$.

Let $m \in \left\{m^0,...,m_{\max}\right\}$. Given $\mathcal{T}_m = (T_1,...,T_m)\in\mathbb{T}_m$, let $\hat{\boldsymbol{\alpha}}_{\mathcal{T}_m} = \left(\hat{\alpha}'_{1,\mathcal{T}_m},...,\hat{\alpha}'_{m+1,\mathcal{T}_m}\right)' = \arg\min_{\boldsymbol{\alpha}_m} Q_{T,1}(\boldsymbol{\alpha}_m;\mathcal{T}_m)$ denote the post-Lasso estimate of $\boldsymbol{\alpha}_m = (\alpha_1,...,\alpha_{m+1})$. Let $\hat{\sigma}^2_{\mathcal{T}_m} \equiv Q_{T,1}(\hat{\boldsymbol{\alpha}}_{\mathcal{T}_m};\mathcal{T}_m)$. Note that we do not impose the condition that $\min_{0\leq j\leq m}(T_{j+1} - T_j) \geq I_{\min} \to \infty$. It is possible to have $T_{j+1} - T_j < p$ for some $j$, in which case the solution $\{\hat{\alpha}_{j,\mathcal{T}_m}, j = 1,...,m+1\}$ is not unique despite its existence. We can treat $\mathcal{T}_m$ and $\mathcal{T}_{m^0}^0 = (T_1^0,...,T_{m^0}^0)$ as two sets with $m$ and $m^0$ break dates, respectively. Let $\bar{\mathcal{T}}_{m+m^0} = (\bar{T}_1,\bar{T}_2,...,\bar{T}_{m+m^0})$ denote the union of $\mathcal{T}_m$ and $\mathcal{T}_{m^0}^0$ with elements ordered in non-descending order: $1 < \bar{T}_1 \leq \bar{T}_2 \leq \cdots \leq \bar{T}_{m+m^0} < T$. In view of the fact that $\hat{\sigma}^2_{\mathcal{T}_{m^0}^0} \geq \hat{\sigma}^2_{\bar{\mathcal{T}}_{m+m^0}}$ and $\hat{\sigma}^2_{\mathcal{T}_{m^0}^0} = \bar{\sigma}^2_T + O_P(T^{-1})$, we have

$$0 \leq \hat{\sigma}^2_{\mathcal{T}_{m^0}^0} - \hat{\sigma}^2_{\bar{\mathcal{T}}_{m+m^0}} = \bar{\sigma}^2_T - \hat{\sigma}^2_{\bar{\mathcal{T}}_{m+m^0}} + O_P\left(T^{-1}\right) \leq \left(m + m^0 + 1\right)J_T + O_P\left(T^{-1}\right), \tag{E.4}$$

where

$$J_T \equiv \sup_{1\le s<r\le T+1,(s,r-1) \text{ does not contain any break points}} T^{-1}\left|\inf_{\alpha}\sum_{t=s}^{r-1}\left(y_t-\alpha' x_t\right)^2-u_t^2\right|.$$

Let $X_{sr}=(x_s,...,x_{r-1})'$, $Y_{sr}=(y_s,...,y_{r-1})'$, and $U_{sr}=(u_s,...,u_{r-1})'$. By standard least squares regression results, if the time interval $(s,r-1)$ does not contain any break points, then $\left|\inf_\alpha \sum_{t=s}^{r-1}\left(y_t-\alpha' x_t\right)^2-u_t^2\right|$ $=U_{sr}'P_{X_{sr}}U_{sr}$, where $P_{X_{sr}}=X_{sr}\left(X_{sr}'X_{sr}\right)^{+}X_{sr}'$ and $A^{+}$ denotes the Moore-Penrose generalized inverse of $A$. Let $v_T=T\delta_T$. Then

$$\begin{aligned}J_T &\le \sup_{1\le s<r\le T+1} T^{-1}U_{sr}'P_{X_{sr}}U_{sr}\\ &= \sup_{1\le s<r\le T+1,r-s\ge v_T} T^{-1}U_{sr}'P_{X_{sr}}U_{sr}+\sup_{1\le s<r\le T,\ r-s<v_T} T^{-1}U_{sr}'P_{X_{sr}}U_{sr}\equiv J_{T1}+J_{T2},\ \text{say.}\end{aligned}$$

For $J_{T1}$, by Lemmas A.3 and A.4 and Assumption A3(ii) we have that w.p.a.1

$$\begin{aligned}J_{T1} &= \sup_{1\le s<r\le T+1,r-s\ge v_T} T^{-1}U_{sr}'X_{sr}\left(X_{sr}'X_{sr}\right)^{-1}X_{sr}'U_{sr}\\ &\le T^{-1}\left[\sup_{1\le s<r\le T+1,r-s\ge v_T}\mu_{\max}\left(\frac{1}{r-s}X_{sr}'X_{sr}\right)\right]^{-1}\sup_{1\le s<r\le T+1,r-s\ge v_T}\left\|\frac{1}{\sqrt{r-s}}X_{sr}'U_{sr}\right\|^2\\ &= T^{-1}O_P(1)\,O_P((\log T)^{c_\delta})=o_P(\delta_T).\end{aligned}$$

For $J_{T2}$, noting that $\mu_{\max}\left(P_{X_{sr}}\right)=1$, we have by analogous arguments as used in the proof of Lemma A.4 and Assumption A3(ii)

$$\begin{aligned}J_{T2} &\le \sup_{1\le s<r\le T,\ r-s<v_T} T^{-1}\sum_{t=s}^{r-1}u_t^2\le T^{-1}\sup_{1\le s\le T-v_T}\sum_{t=s}^{s+v_T-1}\left[u_t^2-E\left(u_t^2\right)\right]+T^{-1}\sup_{1\le s\le T-v_T}\sum_{t=s}^{s+v_T-1}E\left(u_t^2\right)\\ &\le T^{-1}O_P\left(\sqrt{v_T(\log T)}^{c_\delta}\right)+T^{-1}O\left(v_T\right)=O_P\left(T^{-1}v_T\right)=O_P\left(\delta_T\right).\end{aligned}$$

It follows that $J_T=O_P\left(\delta_T\right)$. This, in conjunction with (E.4), implies that $-O_P\left(\delta_T\right)\le\hat\sigma^2_{\mathcal{T}_{m+m^0}}-\bar\sigma^2_T\le O_P\left(T^{-1}\right)$, which holds for all $m$ and $\mathcal{T}_m=(T_1,...,T_m)$. It follows that uniformly in $m$ and $\mathcal{T}_m$ we have

$$\delta_T^{-1}(\hat\sigma^2_{\mathcal{T}_m}-\bar\sigma^2_T)\ge\delta_T^{-1}(\hat\sigma^2_{\mathcal{T}_{m+m^0}}-\bar\sigma^2_T)\ge -O_P(1). \tag{E.5}$$

Next, we want to show

$$\max_{m^0+1\le m\le m_{\max}}\sup_{\mathcal{T}_m\in\mathbb{T}_m}\delta_T^{-1}(\hat\sigma^2_{\mathcal{T}_m}-\bar\sigma^2_T)\le O_P(1). \tag{E.6}$$

Since $\mathcal{T}_m=(T_1,...,T_m)\in\mathbb{T}_m$, for each $T_j^0\in\mathcal{T}^0$ there exists $T_j^*\in\mathcal{T}_m$ such that $\left|T_j^*-T_j^0\right|\le T\delta_T$. This, in conjunction with Assumption A3(i), also ensures that $T_j^*<T_{j+1}^*$ for $j=0,1,...,m^0$, where by default $T_0^*=1$ and $T_{m^0+1}^*=T+1$. Let $\mathcal{T}_{m^0}^*=\left(T_1^*,...,T_{m^0}^*\right)$. Note that

$$\hat\sigma^2_{\mathcal{T}_m}-\bar\sigma^2_T\le Q_{T,1}\left(\hat{\boldsymbol{\alpha}}_{\mathcal{T}_{m^0}^*};\mathcal{T}_{m^0}^*\right)-\bar\sigma^2_T=\sum_{j=1}^{m^0+1}\bar Q_{T,1j},$$

where $\bar Q_{T,1j}\equiv\frac{1}{T}\sum_{t=T_{j-1}^*}^{T_j^*-1}[(y_t-\hat\alpha'_{j,\mathcal{T}_{m^0}^*}x_t)^2-u_t^2]$. In addition, $\min_{1\le j\le m^0}\left|T_j^*-T_j^0\right|\le T\delta_T$, $\min_{0\le j\le m^0}$ $\left|T_{j+1}^0-T_j^0\right|=I_{\min}$, and the fact that $T\delta_T=o\left(I_{\min}\right)$ ensure that $T_j^*-T_{j-1}^*=I_j^0+O(T\delta_T)=I_j^0+o\left(I_{\min}\right)$

for $j = 1, ..., m^0 + 1$. As a result, $\hat{\alpha}_{j,\mathcal{T}^*_{m^0}}$ is uniquely defined in large samples and given by $\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} = (X'_{T^*_{j-1}T^*_j} X_{T^*_{j-1}T^*_j})^{-1} X'_{T^*_{j-1}T^*_j} Y_{T^*_{j-1}T^*_j}$. It is straightforward to show that $\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j = O_P\left[(I^0_j)^{-1/2} + \delta_T\right]$ and

$$\sum_{j=1}^{m^0+1} \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\|^s = m^0 O_P(I_{\min}^{-s/2} + \delta_T^s) \text{ for } s = 1, 2. \tag{E.7}$$

To study $\bar{Q}_{T,1j}$, we consider four subcases: (i1) $T^*_{j-1} < T^0_{j-1}$ and $T^*_j < T^0_j$, (i2) $T^*_{j-1} < T^0_{j-1}$ and $T^*_j \geq T^0_j$, (i3) $T^*_{j-1} \geq T^0_{j-1}$ and $T^*_j < T^0_j$, and (i4) $T^*_{j-1} \geq T^0_{j-1}$ and $T^*_j \geq T^0_j$. In subcase (i1), we have

$$\begin{aligned}
\bar{Q}_{T,1j} &= \frac{1}{T}\sum_{t=T^0_{j-1}}^{T^0_j-1} [(y_t - \hat{\alpha}'_{j,\mathcal{T}^*_{m^0}} x_t)^2 - u_t^2] + \frac{1}{T}\sum_{t=T^*_{j-1}}^{T^0_{j-1}-1} [(y_t - \hat{\alpha}'_{j,\mathcal{T}^*_{m^0}} x_t)^2 - u_t^2] - \frac{1}{T}\sum_{r=T^*_j}^{T^0_j-1} [(y_t - \hat{\alpha}'_{j,\mathcal{T}^*_{m^0}} x_t)^2 - u_t^2] \\
&\equiv \bar{Q}_{T,1j}(1) + \bar{Q}_{T,1j}(2) - \bar{Q}_{T,1j}(3), \text{ say.}
\end{aligned}$$

By Theorem 3.2(ii) and the fact that $\frac{1}{I^0_j}\sum_{t=T^0_{j-1}}^{T^0_j-1} x_t u_t = O_P\left((I^0_j/(\log m^0)^{c_\delta})^{-1/2}\right)$ and $\frac{1}{I^0_j}\sum_{t=T^0_{j-1}}^{T^0_j-1} x_t x'_t = O_P(1)$ uniformly in $j$, we have

$$\begin{aligned}
\bar{Q}_{T,1j}(1) &= -2\left(\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right)' \frac{1}{T}\sum_{t=T^0_{j-1}}^{T^0_j-1} x_t u_t + \left(\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right)' \frac{1}{T}\sum_{t=T^0_{j-1}}^{T^0_j-1} x_t x'_t \left(\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right) \\
&= \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\| O_P\left(T^{-1/2}(\log m^0)^{c_\delta/2}\right) + \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\|^2 \text{ uniformly in } j.
\end{aligned}$$

For $\bar{Q}_{T,1j}(2)$, we have $\bar{Q}_{T,1j}(2) = -2(\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j)' \frac{1}{T}\sum_{t=T^*_{j-1}}^{T^0_{j-1}-1} x_t u_t + (\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j)' \frac{1}{T}\sum_{t=T^*_{j-1}}^{T^0_{j-1}-1} x_t x'_t (\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j) \equiv -2\bar{Q}_{T,1j}(2,1) + \bar{Q}_{T,1j}(2,2)$, say. Noting that uniformly in $j$

$$\begin{aligned}
\bar{Q}_{T,1j}(2,1) &\leq \delta_T \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\| \frac{1}{T\delta_T} \sum_{t=T^0_j - T\delta_T}^{T^0_{j-1}-1} \|x_t u_t\| = \delta_T \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\| O_P(1), \text{ and} \\
\bar{Q}_{T,1j}(2,2) &\leq \delta_T \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\|^2 \mu_{\max}\left(\frac{1}{T\delta_T}\sum_{t=T^0_j - T\delta_T}^{T^0_j-1} x_t x'_t\right) = \delta_T \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\|^2 O_P(1),
\end{aligned}$$

we have $\bar{Q}_{T,1j}(2) = O_P(\delta_T)\left(\left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\| + \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\|^2\right)$ uniformly in $j$. Analogously, we can show that $\bar{Q}_{T,1j}(3) = O_P(\delta_T)\left(\left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\| + \left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\|^2\right)$ uniformly in $j$. It follows that $\bar{Q}_{T,1j} = O_P\left(\delta_T + T^{-1/2}(\log m^0)^{c_\delta/2}\right)\left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\| + O_P(1)\left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\|^2$ uniformly in $j$ in subcase (i1). The same probability order holds in subcases (i2)-(i4). Then by (E.7), we have

$$\begin{aligned}
\delta_T^{-1}\sum_{j=1}^{m^0+1} \bar{Q}_{T,1j} &= O_P\left(1 + T^{-1/2}(\log m^0)^{c_\delta/2}\delta_T^{-1}\right)\sum_{j=1}^{m^0+1}\left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\| + O_P\left(\delta_T^{-1}\right)\sum_{j=1}^{m^0+1}\left\|\hat{\alpha}_{j,\mathcal{T}^*_{m^0}} - \alpha^0_j\right\|^2 \\
&= \left(1 + T^{-1/2}(\log m^0)^{c_\delta/2}\delta_T^{-1}\right) m^0 O_P(I_{\min}^{-1/2} + \delta_T) + \delta_T^{-1} m^0 O_P(I_{\min}^{-1} + \delta_T^2) \\
&= O_P(1),
\end{aligned}$$

and (E.6) follows. Combining (E.5) with (E.6) yields $\max_{m^0+1 \leq m \leq m_{\max}} \sup_{\mathcal{T}_m \in \mathbb{T}_m} \delta_T^{-1}\left|\hat{\sigma}^2_{\mathcal{T}_m} - \bar{\sigma}^2_T\right| = O_P(1)$. ■

# F Proof of Theorem 3.5

Despite the presence of the Lasso penalty term, the proof follows from the same idea as used in the literature on break estimation; see, e.g., Bai (1995, Theorem 1), Bai (1997a, Proposition 3), and Su et al. (2013, Theorem 4.4). The main difference is that these early papers focus on the case of a single break whereas we allow the number of breaks ($m^0$) to diverge to infinity. [Bai and Perron (1998) stated that the limiting distribution in the (fixed) multiple break case is the same as the single break case, but did not give a formal proof.] By Theorem 3.4, $m^0 = \hat{m}_\lambda$ w.p.a.1 so that we can treat $m^0$ as if it were known in large samples. We reformulate the GFL objective function as

$$S_{T\lambda}(\boldsymbol{\alpha},r) = \frac{1}{T}\sum_{j=1}^{m^0+1}\sum_{t=T_{j-1}}^{T_j-1}\left(y_t - x_t'\alpha_j\right)^2 + \lambda\sum_{j=1}^{m^0}\|\alpha_{j+1}-\alpha_{j-1}\| \tag{F.1}$$

where $\boldsymbol{\alpha} = (\alpha_1', \alpha_2', ..., \alpha_{m^0+1}')'$ and $r = (T_1, ..., T_{m^0})$. Let $\hat{\boldsymbol{\alpha}}(r) \equiv \arg\min_{\boldsymbol{\alpha}} S_{T\lambda}(\boldsymbol{\alpha}, r)$, $\hat{r} \equiv (\hat{T}_1, ..., \hat{T}_{m^0}) = \arg\min_r S_{T\lambda}(\hat{\boldsymbol{\alpha}}(r), r)$, and $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}(\hat{r}) = (\hat{\alpha}_1', \hat{\alpha}_2', ..., \hat{\alpha}_{m^0+1}')'$. Let $r^0 = (T_1^0, ..., T_{m^0}^0)$. To study the asymptotic distributions of the Lasso estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{r}$, we can evaluate the global behavior of $S_{T\lambda}(\boldsymbol{\alpha}, r)$ over the whole parameter space for $\boldsymbol{\alpha}$ and $r$ via reparametrization. Define

$$V_{T\lambda}(\mathbf{a},\upsilon) = T\left[S_{T\lambda}(\boldsymbol{\alpha}^0 + D_{m^0}^{-1}\mathbf{a}, r(\upsilon)) - S_{T\lambda}\left(\boldsymbol{\alpha}^0, r^0\right)\right] \tag{F.2}$$

where $r(\upsilon) = (r_1(\upsilon_1), ..., r_{m^0}(\upsilon_{m^0}))$ with $r_j(\upsilon_j) = \lfloor T_j^0 + c_{T,j}\upsilon_j \rfloor$, $c_{T,j} = O(\bar{d}_{T,j}^{-2})$, $\upsilon = (\upsilon_1, ...., \upsilon_{m^0}) \in \mathbb{R}^{m^0}$, $\mathbf{a} = (a_1', a_2', ..., a_{m^0+1}')'$ is a $p\left(m^0+1\right) \times 1$ vector, and $D_{m^0}$ is as defined in Section 3.3. Assume that $r_j(\upsilon_j) = 1$ if $r_j(\upsilon_j) \leq 1$ and $r_j(\upsilon_j) = T$ if $r_j(\upsilon_j) \geq T$. Apparently, the reparametrization in (F.2) conforms with the anticipated rates of pointwise convergence for $\hat{\boldsymbol{\alpha}}$ and $\hat{r}$. Let $\hat{\mathbf{a}}$ and $\hat{\upsilon}$ minimize $V_{T\lambda}(\mathbf{a},\upsilon)$. Then $\hat{\mathbf{a}} = D_{m^0}^{-1}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0)$ and $\lfloor c_{T,j}\hat{\upsilon}_j \rfloor = \hat{T}_j - T_j^0$ for $j = 1, ..., m^0$.

For notational simplicity, we focus on the case where $\upsilon_j \leq 0\ \forall j \in \{1, ..., m^0\}$ as the other $2^{m^0} - 1$ cases can be analyzed analogously. Noting that $\alpha_{j+1}^0 - \alpha_j^0 = d_{T,j}^0$, we have

$$\begin{aligned}
V_{T\lambda}(\mathbf{a},\upsilon) &= \sum_{j=1}^{m^0}\sum_{t=T_{j-1}^0}^{r_j(\upsilon_j)-1}\left\{\left[u_t - (I_j^0)^{-1/2}a_j'x_t\right]^2 - u_t^2\right\} \\
&\quad + \sum_{j=1}^{m^0}\sum_{t=r_j(\upsilon_j)}^{T_j^0-1}\left\{\left[u_t - (I_{j+1}^0)^{-1/2}a_{j+1}'x_t - d_{T,j}^{0\prime}x_t\right]^2 - u_t^2\right\} \\
&\quad + \sum_{t=T_{m^0}^0}^{T}\left\{\left[u_t - (I_{m^0+1}^0)^{-1/2}a_{m^0+1}'x_t\right]^2 - u_t^2\right\} \\
&\quad + T\lambda\sum_{j=1}^{m^0}\left\{\left\|(I_{j+1}^0)^{-1/2}a_{j+1} - (I_j^0)^{-1/2}a_j + d_{T,j}^0\right\| - \left\|d_{T,j}^0\right\|\right\} \\
&\equiv V_{T\lambda,1}(\mathbf{a},\upsilon) + V_{T\lambda,2}(\mathbf{a},\upsilon) + V_{T\lambda,3}(\mathbf{a}) + V_{T\lambda,4}(\mathbf{a}), \text{ say.}
\end{aligned}$$

We shall prove the weak convergence of $V_{T\lambda}(\mathbf{a},\upsilon)$ on the compact set $\mathbb{S}_K \equiv \{(\mathbf{a},\upsilon) : \|\mathbf{a}\| \leq \sqrt{m^0}K, \|\upsilon\| \leq \sqrt{m^0}K\}$ where $K$ is fixed positive constant. By the triangle inequality and Assumption A6,

$|V_{T\lambda,4}(\mathbf{a})| \leq T\lambda\sum_{j=1}^{m^0}\left\|(I_{j+1}^0)^{-1/2}a_{j+1}-(I_j^0)^{-1/2}a_j\right\| = O(m^0T\lambda I_{\min}^{-1/2}) = o(1)$ uniformly in $\mathbf{a}$. It is straightforward to show that uniformly in $(\mathbf{a},\upsilon)\in\mathbb{S}_L$,

$$V_{T\lambda,1}(\mathbf{a},\upsilon) = -2\sum_{j=1}^{m^0}(I_j^0)^{-1/2}a_j'\sum_{t=T_{j-1}^0}^{T_j^0-1}x_tu_t+\sum_{j=1}^{m^0}(I_j^0)^{-1}a_j'\sum_{t=T_{j-1}^0}^{T_j^0-1}x_tx_t'a_j+o_P(1),$$

and

$$V_{T\lambda,2}(\mathbf{a},\upsilon) = -2\sum_{j=1}^{m^0}d_{T,j}^{0\prime}\sum_{t=\lfloor T_j^0+c_T\upsilon_j\rfloor}^{T_j^0-1}x_tu_t+\sum_{j=1}^{m^0}d_{T,j}^{0\prime}\sum_{t=\lfloor T_j^0+c_T\upsilon_j\rfloor}^{T_j^0-1}x_tx_t'd_{T,j}^0+o_P(1).$$

In addition, $V_{T\lambda,3}(\mathbf{a}) = -2(I_{m^0+1}^0)^{-1/2}a_{m^0+1}'\sum_{t=T_{m^0}^0}^{T}x_tu_t+(I_{m^0+1}^0)^{-1}a_{m^0+1}'\sum_{t=T_{m^0}^0}^{T}x_tx_t'a_{m^0+1}$. It follows that uniformly in $(\mathbf{a},\ \upsilon)\in\mathbb{S}_L$ we have

$$V_{T\lambda}(\mathbf{a},\upsilon) = \bar{V}_{1T}(\mathbf{a})+\bar{V}_{2T}(\upsilon)+o_P(1)$$

where $\bar{V}_{1T}(\mathbf{a}) = -2\mathbf{a}'D_{m^0}^{-1}\mathbb{X}'U+\mathbf{a}'D_{m^0}^{-1}\mathbb{X}'\mathbb{X}D_{m^0}^{-1}\mathbf{a}$, and $\bar{V}_{2T}(\upsilon) = \sum_{j=1}^{m^0}[-2d_{T,j}^{0\prime}\sum_{t=\lfloor T_j^0+c_T\upsilon_j\rfloor}^{T_j^0-1}x_tu_t+d_{T,j}^{0\prime}\sum_{t=\lfloor T_j^0+c_T\upsilon_j\rfloor}^{T_j^0-1}x_tx_t'd_{T,j}^0]$. Noting that $\bar{V}_{1T}(\mathbf{a})$ converges weakly on a compact set to $\bar{V}_{1T}^{(0)}(\mathbf{a}) = -2\mathbf{a}'\mathbf{\Phi}^{1/2}Z+\mathbf{a}'\mathbf{\Psi}\mathbf{a}$, where $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are as defined in Section 3.3, and $Z$ is a $p\left(m^0+1\right)\times 1$ vector of independent standard normal variables. By the continuous mapping theorem (CMT),

$$S\hat{\mathbf{a}} = SD^{-1}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}^0)\xrightarrow{d}S\arg\min_{\mathbf{a}}\bar{V}_{1T}^{(0)}(\mathbf{a}) = N\left(0,S\mathbf{\Psi}^{-1}\mathbf{\Phi}\mathbf{\Psi}^{-1}S'\right).$$

This proves part (i) in Theorem 3.5.

Let $c_{T,j} = (\bar{d}_{T,j}'\Psi_j\bar{d}_{T,j})^{-1}$ for $j=1,...,m^0$. By the invariance principle for heterogenous mixing processes (e.g., White (2001, Theorem 7.18)), $d_{T,j}^{0\prime}\sum_{t=\lfloor T_j^0+c_{T,j}\upsilon_j\rfloor}^{T_j^0-1}x_tu_t = \frac{1}{\sqrt{c_{T,j}}}\sum_{t=\lfloor T_j^0+c_{T,j}\upsilon_j\rfloor}^{T_j^0-1}c_{T,j}^{1/2}d_{T,j}^{0\prime}x_tu_t$ $\Rightarrow\phi_{j,1}W_{j,1}(-\upsilon)$. Because $c_{T,j}\to\infty$, we have $d_{T,j}^{0\prime}\sum_{t=\lfloor T_j^0+c_{T,j}\upsilon_j\rfloor}^{T_j^0-1}x_tx_t'd_{T,j}^0 = \frac{-c_{T,j}\upsilon_j}{-c_{T,j}\upsilon_j}d_{T,j}^{0\prime}\sum_{t=\lfloor T_j^0+c_{T,j}\upsilon_j\rfloor}^{T_j^0-1}x_tx_t'd_{T,j}^0\xrightarrow{P}|\upsilon_j|$ by Assumptions A1 and A5(ii). It follows that $\bar{V}_{2T}(\upsilon)\Rightarrow\sum_{j=1}^{m^0}[-2\phi_{j,1}W_{j,1}(-\upsilon_j)+|\upsilon_j|]$ when $\upsilon_j<0\ \forall j\in\{1,...,m^0\}$. For the case $\upsilon_j>0\ \forall j\in\{1,...,m^0\}$, the counter part of $\bar{V}_{2T}(\upsilon)$ is $\bar{V}_{2T}^*(\upsilon) = \sum_{j=1}^{m^0}[-2d_{T,j}^{0\prime}\sum_{t=T_j^0}^{\lfloor T_j^0+c_{T,j}\upsilon_j\rfloor}x_tu_t+d_{T,j}^{0\prime}\sum_{t=T_j^0}^{\lfloor T_j^0+c_{T,j}\upsilon_j\rfloor}x_tx_t'd_{T,j}^0]$, which converges weakly to $\sum_{j=1}^{m^0}[-2\sqrt{\xi_j}\phi_{j,2}W_{j,2}(\upsilon_j)+\xi_j\upsilon_j]$. The cases where elements of $\upsilon$ have different signs can be derived analogously by discussing the signs of $(T_j-T_j^0)$'s as in the proof of Theorem 3.4. The independence between $W_{j,1}(\cdot)$ and $W_{j,2}(\cdot)$ arises because by a simple application of Davydov's inequality for strong mixing processes (see, e.g., Hall and Heyde (1980, Corollary A.2)) and Assumptions A1 and A6(i), for any $\upsilon_j<0$, $\bar{\upsilon}_j>0$, and small $\epsilon>0$,

$$\left|E\left[d_{T,j}^{0\prime}\sum_{t=\lfloor T_j^0+c_{T,j}\upsilon_j\rfloor}^{T_j^0-1}x_tu_t\sum_{s=T_j^0}^{\lfloor T_j^0+c_{T,j}\bar{\upsilon}_j\rfloor}u_sx_s'd_{T,j}^0\right]\right|\leq 8\left\|d_{T,j}^0\right\|^2\sup_{t\geq 1}E\left[\left\|x_tu_t\right\|^{2+\epsilon}\right]\sum_{\tau=1}^{\infty}\tau\alpha(\tau)^{\epsilon/(2+\epsilon)} = o(1).$$

By the same reason, $W_{j,l}$ and $W_{i,k}$ are independent for all $j\neq i$ and $l,k=1,2$. Consequently, we have $\bar{V}_{2T}(\upsilon)\Rightarrow\sum_{j=1}^{m^0}-2Z_j(\upsilon_j)$, $(\Delta_j'\Psi_j\Delta_j)\bar{d}_{T,j}^2(\hat{T}_j-T_j^0)\xrightarrow{d}\arg\max_r Z_j(r)$ by CMT, and $\hat{T}_j-T_j^0$ are asymptotically independent of $\hat{T}_i-T_i^0$ for all $j\neq i$. This completes the proof of part (ii). ■

# G Proof of Theorem 3.6

We prove the theorem by showing that $\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{m^0}}$ shares the same asymptotic distribution as $\hat{\boldsymbol{\alpha}}_{\mathcal{T}^0_{m^0}}$ and the asymptotic distribution of $\hat{\boldsymbol{\alpha}}_{\mathcal{T}^0_{m^0}}$ is as given in the theorem. The latter can be verified easily under our assumptions by a simple application of the central limit theorem for heterogenous strong mixing processes; see, e.g., White (2001, Theorem 5.2). For notational simplicity, we shall suppress the dependence of $D_{m^0}$ and $\hat{D}_{m^0}$ on $m^0$ and write them as $D$ and $\hat{D}$, respectively. Noting that $\hat{D}(\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{m^0}} - \boldsymbol{\alpha}^0) = (\hat{D}^{-1}\hat{\mathbb{X}}'\hat{\mathbb{X}}\hat{D}^{-1})^{-1}\hat{D}^{-1}\hat{\mathbb{X}}'[(\mathbb{X}-\hat{\mathbb{X}})\boldsymbol{\alpha}^0 + U]$ and $D(\hat{\boldsymbol{\alpha}}_{\mathcal{T}^0_{m^0}} - \boldsymbol{\alpha}^0) = (D^{-1}\mathbb{X}'\mathbb{X}D^{-1})^{-1}D^{-1}\mathbb{X}'U$ by (3.3)-(3.4), we have

$$\begin{aligned} S[\hat{D}(\hat{\boldsymbol{\alpha}}_{\hat{\mathcal{T}}_{m^0}} - \boldsymbol{\alpha}^0) - D(\hat{\boldsymbol{\alpha}}_{\mathcal{T}^0_{m^0}} - \boldsymbol{\alpha}^0)] &= S(\hat{A}^{-1}\hat{B} - A^{-1}B) + S\hat{A}^{-1}\hat{C} \\ &= S\hat{A}^{-1}(\hat{B} - B) + S(\hat{A}^{-1} - A^{-1})B + S\hat{A}^{-1}\hat{C}, \end{aligned}$$

where $\hat{A} = \hat{D}^{-1}\hat{\mathbb{X}}'\hat{\mathbb{X}}\hat{D}^{-1}$, $A = D^{-1}\mathbb{X}'\mathbb{X}D^{-1}$, $\hat{B} = \hat{D}^{-1}\hat{\mathbb{X}}'U$, $B = D^{-1}\mathbb{X}'U$, and $\hat{C} = \hat{D}^{-1}\hat{\mathbb{X}}'(\mathbb{X}-\hat{\mathbb{X}})\boldsymbol{\alpha}^0$. We prove the theorem by showing that (i) $S_{1T} \equiv SA^{-1}(\hat{B}-B) = o_P(1)$, (ii) $S_{2T} \equiv S(\hat{A}^{-1}-A^{-1})\hat{B} = o_P(1)$, and (iii) $S_{3T} \equiv S\hat{A}^{-1}\hat{C} = o_P(1)$.

To proceed, we first show that: (a) $\lambda_{\min}(A) \geq \underline{c}_{xx}/2$ and $\lambda_{\max}(A) \leq 2\bar{c}_{xx}$ w.p.a.1, (b) $\left\|\hat{A} - A\right\|^2 = o_P\left(1/m^0\right)$, and (c) $\lambda_{\min}(\hat{A}) \geq \underline{c}_{xx}/4$ and $\lambda_{\max}(\hat{A}) \leq 4\bar{c}_{xx}$ w.p.a.1. By Weyl inequality, $\lambda_{\min}(A) \geq \lambda_{\min}(E(A)) - \lambda_{\max}(A - E(A)) \geq \lambda_{\min}(E(A)) - \|A - E(A)\|$. Assumption A2(i) ensures that $\lambda_{\min}(E(A)) \geq \underline{c}_{xx}$. By Assumption A1 and Davydov inequality, we can readily verify that $E\|A - E(A)\|^2 = O\left(m^0/I_{\min}\right) = o(1)$. Thus $\|A - E(A)\| = o_P(1)$ by Chebyshev inequality and the first part of (a) follows. Analogously, we can prove the second part of (a). For (b), we have

$$\begin{aligned} \hat{A} - A &= \hat{D}^{-1}(\hat{\mathbb{X}} - \mathbb{X})'\hat{\mathbb{X}}\hat{D}^{-1} + \hat{D}^{-1}\mathbb{X}'(\hat{\mathbb{X}} - \mathbb{X})\hat{D}^{-1} + D^{-1}\mathbb{X}'\mathbb{X}(\hat{D}^{-1} - D^{-1}) + (\hat{D}^{-1} - D^{-1})\mathbb{X}'\mathbb{X}\hat{D}^{-1} \\ &\equiv A_1 + A_2 + A_3 + A_4, \text{ say.} \end{aligned}$$

Write $A_1$ as a partitioned matrix: $A_1 = (A_{1,ij})_{i,j=1}^{m^0+1}$ where $A_{1,ij}$'s are $p \times p$ matrices. Note that $\hat{\mathbb{X}}'\hat{\mathbb{X}} = \text{diag}(\hat{\mathbb{X}}_1'\hat{\mathbb{X}}_1, ..., \hat{\mathbb{X}}_{m^0+1}'\hat{\mathbb{X}}_{m^0+1})$ and $\mathbb{X}'\hat{\mathbb{X}}$ is a block tridiagonal matrix w.p.a.1:

$$\mathbb{X}'\hat{\mathbb{X}} = \begin{pmatrix} \sum_{t=1}^{T_1^0 \wedge \hat{T}_1 - 1} \xi_t & \sum_{t=\hat{T}_1}^{T_1^0 - 1} \xi_t & 0 & \cdots & 0 & 0 \\ \sum_{t=T_1^0}^{\hat{T}_1 - 1} \xi_t & \sum_{t=T_1^0 \vee \hat{T}_1}^{T_2^0 \wedge \hat{T}_2 - 1} \xi_t & \sum_{t=\hat{T}_2}^{T_2^0 - 1} \xi_t & \cdots & 0 & 0 \\ 0 & \sum_{t=T_2^0}^{\hat{T}_2 - 1} \xi_t & \sum_{t=T_2^0 \vee \hat{T}_2}^{T_3^0 \wedge \hat{T}_3 - 1} \xi_t & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sum_{t=T^0_{m^0-1} \vee \hat{T}_{m^0-1}}^{T^0_{m^0} \wedge \hat{T}_{m^0} - 1} \xi_t & \sum_{t=\hat{T}_{m^0}}^{T_{m^0} - 1} \xi_t \\ 0 & 0 & 0 & \cdots & \sum_{t=T^0_{m^0}}^{\hat{T}_{m^0} - 1} \xi_t & \sum_{t=T^0_{m^0} \vee \hat{T}_{m^0}}^{T} \xi_t \end{pmatrix}$$

where $\xi_t = x_t x_t'$ and $\sum_{t=a}^{b} \xi_t = 0$ if $b > a$ and we use the fact that when $P(T_{j-1}^0 < \hat{T}_j < T_{j+1}^0) \to 1$ because w.p.a.1 $\left|\hat{T}_j - T_j^0\right| \leq T\delta_T = o(I_{\min})$ by Theorem 3.1(i) and Assumption A3(i). We can analyze $A_{1,ij}$ for $i, j = 1, ..., m^0 + 1$. For example, if $T_1^0 \geq \hat{T}_1$, then $A_{1,11} = 0$, $A_{1,21} = 0$, and

$$\|A_{1,12}\| = \hat{I}_1^{-1/2}\hat{I}_2^{-1/2}\left\|\sum_{t=\hat{T}_1}^{T_1^0 - 1} x_t x_t'\right\| \leq T\delta_T \hat{I}_1^{-1/2}\hat{I}_2^{-1/2}\frac{1}{T\delta_T}\sum_{t=T_1 - T\delta_T}^{T_1^0 - 1}\|x_t x_t'\| = O_P\left(T\delta_T I_{\min}^{-1}\right);$$

and if $T_1^0 < \hat{T}_1$, then $A_{1,12} = 0$, $\|A_{1,11}\| = \hat{I}_1^{-1} \left\| \sum_{t=T_1^0}^{\hat{T}_1 - 1} x_t x_t' \right\| \leq T\delta_T \hat{I}_1^{-1} \frac{1}{T\delta_T} \sum_{t=T_1^0}^{T_1^0 + T\delta_T - 1} \|x_t x_t'\| = O_P\left(T\delta_T I_{\min}^{-1}\right)$, and analogously, $\|A_{1,21}\| = \hat{I}_1^{-1/2} \hat{I}_2^{-1/2} \left\| \sum_{t=T_1^0}^{\hat{T}_1 - 1} x_t x_t' \right\| = O_P\left(T\delta_T I_{\min}^{-1}\right)$. By the same token, we can show that for those $A_{1,ij}$'s that are nonzero, their Frobenius norm are uniformly bounded from above by $O_P\left(T\delta_T I_{\min}^{-1}\right)$. Consequently, $\|A_1\|^2 = \sum_{i=1}^{m^0+1} \sum_{j=1,|j-i|\leq 1}^{m^0+1} \|A_{ij}\|^2 = O_P(m^0 \left(T\delta_T I_{\min}^{-1}\right)^2) = o_P\left(1/m^0\right)$. For $A_3$, we have

$$\begin{aligned}
\|A_3\|^2 &= \operatorname{tr}\left(D^{-1}\mathbb{X}'\mathbb{X}D^{-1}D(\hat{D}^{-1} - D^{-1})(\hat{D}^{-1} - D^{-1})DD^{-1}\mathbb{X}'\mathbb{X}D^{-1}\right) \\
&\leq \max_{1\leq j\leq m^0+1} I_j^0 \left(\hat{I}_j^{-1/2} - I_j^{0-1/2}\right)^2 \operatorname{tr}\left(D^{-1}\mathbb{X}'\mathbb{X}D^{-1}D^{-1}\mathbb{X}'\mathbb{X}D^{-1}\right) \\
&\leq \max_{1\leq j\leq m^0+1} I_j^0 \left(\hat{I}_j^{-1/2} - I_j^{0-1/2}\right)^2 \lambda_{\max}(A) \operatorname{tr}(A) \\
&= O_P\left(T^2\delta_T^2 I_{\min}^{-2}\right) O_P(1) O_P\left(m^0\right) = O_P\left(m^0 \left(T\delta_T I_{\min}^{-1}\right)^2\right) = o_P\left(1/m^0\right),
\end{aligned}$$

where we use the fact that $I_j^0 \left(\hat{I}_j^{-1/2} - I_j^{0-1/2}\right)^2 = \frac{(\hat{I}_j - I_j^0)^2}{\hat{I}_j \left(\hat{I}_j^{1/2} + (I_j^0)^{1/2}\right)^2} = O_P\left(T^2\delta_T^2 I_{\min}^{-2}\right)$ uniformly in $j$ by Theorem 3.1(i). Analogously, we can show that $\|A_s\| = o_P\left(1/m^0\right)$ for $s = 2, 4$. Thus we have shown that $\left\|\hat{A} - A\right\|^2 = o_P\left(1/m^0\right)$. For part (c), we apply Weyl inequality to obtain w.p.a.1, $\lambda_{\min}(\hat{A}) \geq \lambda_{\min}(A) - \lambda_{\max}(A - \hat{A}) \geq \lambda_{\min}(A) - \left\|A - \hat{A}\right\| \geq \underline{c}_{xx}/2 - o_P(1) \geq \underline{c}_{xx}/4$. Analogously, we can show the second part of (c) holds.

To show (i), we first make the following decomposition $S_{1T} = S\hat{A}^{-1}\hat{D}^{-1}(\hat{\mathbb{X}} - \mathbb{X})'U + S\hat{A}^{-1}(\hat{D}^{-1} - D^{-1})\mathbb{X}'U \equiv S_{1T,1} + S_{1T,2}$. By Theorem 3.1(i) and Assumption A3(i), $\left|\hat{T}_j - T_j^0\right| \leq \delta_T T = o(I_{\min})$. This ensures that w.p.a.1 $\hat{T}_j$ lies between $T_{j-1}^0$ and $T_{j+1}^0$ for $j = 1, ..., m^0$. Let $\bar{S}_{1T,1} \equiv \hat{D}^{-1}(\hat{\mathbb{X}} - \mathbb{X})'U$. Write $\bar{S}_{1T,1} = (\bar{S}_{1T,1,1}', ..., \bar{S}_{1T,1,m^0+1}')'$, where $S_{1T,j}$'s are $p \times 1$ vectors. $\bar{S}_{1T,1,1} = 0$ if $T_1^0 \geq \hat{T}_1$ and $\bar{S}_{1T,1,1} = \hat{I}_j^{-1/2} \sum_{t=T_1^0}^{\hat{T}_1 - 1} x_t u_t$ if $T_1^0 < \hat{T}_1$, we have w.p.a.1, $\left\|\bar{S}_{1T,1,1}\right\| \leq T\delta_T \hat{I}_1^{-1/2} \frac{1}{T\delta_T} \sum_{t=T_1^0}^{T_1^0 + T\delta_T - 1} \|x_t u_t\| = O_P(T\delta_T I_{\min}^{-1/2}) O_P(1) = o_P\left((m^0)^{-1/2}\right)$. Analogously, we can show that $\left\|\bar{S}_{1T,j,m^0+1}\right\| = o_P\left((m^0)^{-1/2}\right)$ for $j = 2, ..., m^0$, $m^0 + 1$ and $\left\|\bar{S}_{1T,1}\right\|^2 = \sum_{j=1}^{m^0+1} \left\|\bar{S}_{1T,1,j}\right\|^2 = o_P(1)$. Consequently, we have

$$\|S_{1T,1}\|^2 \leq \left\|S\hat{A}^{-1}\right\|^2 \left\|\bar{S}_{1T,1}\right\|^2 = \operatorname{tr}\left(S\hat{A}^{-1}\hat{A}^{-1}S'\right) \left\|\bar{S}_{1T,1}\right\|^2 \leq \left[\lambda_{\min}(\hat{A})\right]^{-2} \|S\|^2 \left\|\bar{S}_{1T,1}\right\|^2 = o_P(1).$$

Noting that $\left\|D^{-1}\mathbb{X}'U\right\|^2 = O_P\left(m^0\right)$ by Markov inequality and $\operatorname{tr}(S\hat{A}^{-1}\hat{A}^{-1}S') \leq [\lambda_{\min}(\hat{A})]^{-2} \|S\|^2 = O_P(1)$, we have

$$\begin{aligned}
\|S_{1T,2}\|^2 &\leq \left\|S\hat{A}^{-1}\hat{D}^{-1}(\hat{D} - D)\right\|^2 \left\|D^{-1}\mathbb{X}U\right\|^2 \\
&= \operatorname{tr}\left(S\hat{A}^{-1}\hat{D}^{-1}(\hat{D} - D)(\hat{D} - D)\hat{D}^{-1}\hat{A}^{-1}S'\right) \left\|D^{-1}\mathbb{X}U\right\|^2 \\
&\leq \max_{1\leq j\leq m^0+1} \hat{I}_j^{-1} \left(\hat{I}_j^{1/2} - I_j^{01/2}\right)^2 \operatorname{tr}\left(S\hat{A}^{-1}\hat{A}^{-1}S'\right) \left\|D^{-1}\mathbb{X}U\right\|^2 \\
&= O_P(T^2\delta_T^2 I_{\min}^{-2}) O_P(1) O_P\left(m^0\right) = O_P(m^0 \left(T\delta_T I_{\min}^{-1}\right)^2) = o_P(1),
\end{aligned}$$

where we use the fact that $\hat{I}_j^{-1} \left(\hat{I}_j^{1/2} - I_j^{01/2}\right)^2 \leq \frac{(\hat{I}_j - I_j^0)^2}{\hat{I}_j \left(\hat{I}_j^{1/2} + I_j^{01/2}\right)^2} = O_P\left(T^2\delta_T^2 I_{\min}^{-2}\right)$ uniformly in $j$ by Theorem 3.1(i). Thus, we have $S_{1T} = o_P(1)$.

To show (ii), we apply the above results in (a)-(c) and the fact that $\|B\|^2 = O_P\left(m^0\right)$ to obtain

$$\begin{aligned}
\|S_{2T}\|^2 &= \left\|S\hat{A}^{-1}(A-\hat{A})A^{-1}B\right\|^2 \leq \left\|S\hat{A}^{-1}(A-\hat{A})A^{-1}\right\|^2 \|B\|^2 \\
&= \operatorname{tr}\left(S\hat{A}^{-1}(A-\hat{A})A^{-1}A^{-1}(A-\hat{A})S'\right)\|B\|^2 \\
&\leq \left[\lambda_{\min}(A)\right]^{-2}\left[\lambda_{\min}(\hat{A})\right]^{-2}\left\|\hat{A}-A\right\|^2\|S\|^2\|B\|^2 \\
&= O_P(1)\,O_P(1)\,o_P\left(1/m^0\right)O(1)\,O_P\left(m^0\right) = o_P(1).
\end{aligned}$$

We now show (iii). We write $\hat{C} = (\hat{C}_1', ..., \hat{C}_{m^0+1}')'$ where $\hat{C}_j$'s are $p \times 1$ vectors. For $\hat{C}_1$, we have

$$\hat{C}_1 = \begin{cases} 0 & \text{if } T_1^0 \geq \hat{T}_1 \\ \hat{I}_j^{-1/2}\sum_{t=T_1^0}^{\hat{T}_1-1} x_t x_t'\left(\alpha_2^0 - \alpha_1^0\right) & \text{if } T_1^0 < \hat{T}_1 \end{cases} \quad \text{w.p.a.1,}$$

where we use the fact that when $\hat{T}_1 > T_1^0$, $P(\hat{T}_1 < T_2^0) \to 1$ because w.p.a.1 $\hat{T}_1 - T_1^0 \leq T\delta_T = o(T_2^0 - T_1^0)$ by Theorem 3.1(i) and Assumption A3(i). It follows that

$$\left\|\hat{C}_1\right\| \leq T\delta_T\hat{I}_1^{-1/2}\frac{1}{T\delta_T}\sum_{t=T_1^0}^{T_1^0+T\delta_T-1}\|x_t\|^2\left\|\alpha_2^0-\alpha_1^0\right\| = O_P\left(T\delta_T I_{\min}^{-1/2}\right).$$

Analogously, we can show that $\left\|\hat{C}_{m^0+1}\right\| = O_P(T\delta_T I_{\min}^{-1/2})$. For the $\hat{C}_j$ with $j = 2, ..., m^0$, we can discuss four subcases according to the signs of $\hat{T}_{j-1} - T_{j-1}^0$ and $\hat{T}_j - T_j^0$ as in the proof of Theorem 3.4, and show that $\left\|\hat{C}_j\right\| = O_P(T\delta_T I_{\min}^{-1/2})$ uniformly in $j$ for each subcase. Consequently, we have $\left\|\hat{C}\right\|^2 = \sum_{j=1}^{m^0+1}\left\|\hat{C}_j\right\|^2 = m^0 O_P\left(T^2\delta_T^2 I_{\min}^{-1}\right) = o_P(1)$ and

$$\|S_{3T}\|^2 \leq \operatorname{tr}\left(S\hat{A}^{-1}\hat{A}^{-1}S'\right)\left\|\hat{C}\right\|^2 \leq \left[\lambda_{\min}(\hat{A})\right]^{-2}\|S\|^2\left\|\hat{C}\right\|^2 = O_P(1)\,O(1)\,o_P(1) = o_P(1).$$

This completes the proof of the theorem. ■