THE MAXWELL SCHOOL

CENTER FOR POLICY RESEARCH

WORKING PAPER SERIES

# Estimation and Identification of Change Points in Panel Models with Nonstationary or Stationary Regressors and Error Term

Badi H. Baltagi, Chihwa Kao and Long Liu

ISSN: 1525-3066

426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020

http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

Paper No. 178
January 2015

## CENTER FOR POLICY RESEARCH –Spring 2015

**Leonard M. Lopoo, Director**
**Associate Professor of Public Administration and International Affairs (PAIA)**

### Associate Directors

Margaret Austin
Associate Director
Budget and Administration

John Yinger
Trustee Professor of Economics and PAIA
Associate Director, Metropolitan Studies Program

### SENIOR RESEARCH ASSOCIATES

Badi H. Baltagi............................................ Economics
Robert Bifulco..........................................................PAIA
Thomas Dennison ..................................................PAIA
Alfonso Flores-Lagunes ............................. Economics
Sarah Hamersma ...................................................PAIA
William C. Horrace .................................... Economics
Yilin Hou ..................................................................PAIA
Duke Kao.................................................... Economics
Sharon Kioko ........................................................ PAIA
Jeffrey Kubik.............................................. Economics
Yoonseok Lee ............................................ Economics
Amy Lutz....................................................... Sociology
Yingyi Ma...................................................... Sociology
Jerry Miner .................................................Economics
Cynthia Morrow ................................................. PAIA
Jan Ondrich..................................................Economics
John Palmer......................................................... PAIA
David Popp .......................................................... PAIA
Stuart Rosenthal .........................................Economics
Ross Rubenstein.................................................. PAIA
Rebecca Schewe ..........................................Sociology
Amy Ellen Schwartz ......................... PAIA/Economics
Perry Singleton………………………….....Economics
Abbey Steele........................................................ PAIA
Michael Wasylenko... ………………………Economics
Peter Wilcoxen…………………………………………PAIA

### GRADUATE ASSOCIATES

Emily Cardon......................................................... PAIA
Hannah Dalager ..................................................... PAIA
Maidel De La Cruz.................................................. PAIA
Carlos Diaz ...................................................Economics
Vantiel Elizabeth Duncan ...................................... PAIA
Alex Falevich ................................................Economics
Lincoln Groves ...................................................... PAIA
Ruby Jennings........................................................ PAIA
Yusun Kim ............................................................. PAIA
Bridget Lenkiewicz ................................................ PAIA
Michelle Lofton ...................................................... PAIA
Roberto Martinez.................................................... PAIA
Qing Miao ..............................................................PAIA
Nuno Abreu Faro E Mota............................. Economics
Judson Murchie .....................................................PAIA
Sun Jung Oh .......................................... Social Science
Brian Ohl................................................................PAIA
Laura Rodriquez-Ortiz ...........................................PAIA
Timothy Smilnak ....................................................PAIA
Kelly Stevens .........................................................PAIA
Rebecca Wang ............................................. Sociology
Pengju Zhang ............................................. Economics
Xirui Zhang ................................................. Economics

### STAFF

Kelly Bogart…..………….……Administrative Specialist
Karen Cimilluca.....…………………..…Office Coordinator
Kathleen Nasto........................Administrative Assistant
Candi Patterson……………….……Computer Consultant
Mary Santy…………………….....Administrative Assistant
Katrina Wingle……………….Administrative Assistant

## Abstract

This paper studies the estimation of change point in panel models. We extend Bai (2010) and Feng, Kao and Lazarová (2009) to the case of stationary or nonstationary regressors and error term, and whether the change point is present or not. We prove consistency and derive the asymptotic distributions of the Ordinary Least Squares (OLS) and First Difference (FD) estimators. We find that the FD estimator is robust for all cases considered.

**JEL No.** C12, C13, C22

**Keywords:** Panel Data, Change Point, Consistency, Nonstationarity

We dedicate this paper in honor of Peter Schmidt's many contributions to econometrics and in particular non-stationary time series analysis like Amsler, Schmidt and Vogelsang (2009) and panel data econometrics including his extensive work on dynamic panel data like Ahn and Schmidt (1995). We would like to thank the Associate Editor and three referees for their helpful comments and suggestions.

Address correspondence to: Badi H. Baltagi, Department of Economics, Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020; tel: 315-443-1630; fax: 315-443-1081; e-mail: bbaltagi@maxwell.syr.edu.

Chihwa Kao: Department of Economics, Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020; tel: 315-443-3233; fax: 315-443-1081; e-mail: cdkao@maxwell.syr.edu.

Long Liu: Department of Economics, College of Business, University of Texas at San Antonio, One UTSA Circle, TX 78249-0633; tel: 210-458-6169; fax: 210-458-5837; e-mail: long.liu@utsa.edu. Long Liu gratefully acknowledges the summer research grant from the College of Business at the University of Texas at San Antonio.

# Estimation and Identification of Change Points in Panel Models with Nonstationary or Stationary Regressors and Error Term*

Badi H. Baltagi†, Chihwa Kao‡, Long Liu§

This version: December 31, 2014

**Abstract**

This paper studies the estimation of change point in panel models. We extend Bai (2010) and Feng, Kao and Lazarová (2009) to the case of stationary or nonstationary regressors and error term, and whether the change point is present or not. We prove consistency and derive the asymptotic distributions of the Ordinary Least Squares (OLS) and First Difference (FD) estimators. We find that the FD estimator is robust for all cases considered.

**Keywords:** Panel Data, Change Point, Consistency, Nonstationarity.
**JEL Classification:** C12, C13, C22.

## 1 Introduction

Testing and estimation of change points in time series models have been widely studied, see Picard (1985), Nunes, Kuan and Newbold (1995), Hsu and Kuan (2008), Bai (1996, 1997, 1998) and Perron and Zhu (2005), to mention a few. Zeileis, Kleiber, Krämer and Hornik (2003) incorporate testing and dating of structural changes in the package `strucchange` in the `R` system for statistical computing. One important issue in the time series change point literature is that the estimate of the break date can not be consistently estimated, no matter how large the sample. Recently, this change point literature has been extended to panel data, see Feng, Kao and Lazarová (2009), Bai (2010), Hsu and Lin (2011), and Kim (2011), to mention a few. For panel data, the number of cross-sectional units $n$ can be much larger than the number of time series observations $T$. Bai

*We dedicate this paper in honour of Peter Schmidt's many contributions to econometrics and in particular nonstationary time series analysis like Amsler, Schmidt and Vogelsang (2009) and panel data econometrics including his extensive work on dynamic panel data like Ahn and Schmidt (1995). We would like to thank the Associate Editor and three referees for their helpful comments and suggestions.

†Address correspondence to: Badi H. Baltagi, Department of Economics, Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020; tel: 315-443-1630; fax: 315-443-1081; e-mail: bbaltagi@maxwell.syr.edu.

‡Chihwa Kao: Department of Economics, Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, NY 13244-1020; tel: 315-443-3233; fax: 315-443-1081; e-mail: cdkao@maxwell.syr.edu.

§Long Liu: Department of Economics, College of Business, University of Texas at San Antonio, One UTSA Circle, TX 78249-0633; tel: 210-458-6169; fax: 210-458-5837; e-mail: long.liu@utsa.edu. Long Liu gratefully acknowledges the summer research grant from the College of Business at the University of Texas at San Antonio.

(2010) shows that it is possible to obtain consistent estimates of the break point as $n$ goes to infinity. Consistency is obtained even when a regime contains a single observation, making it possible to quickly identify the onset of a new regime. Feng et al. (2009) extend Bai (2010) to a multiple regression model in a panel data setting where a break occurs at an unknown common date. They show that the break date estimate is consistent and derive its asymptotic distribution without the shrinking break assumption. In a pure time series framework, Perron and Zhu (2005) analyze structural breaks with a deterministic time trend regressor. Bai, Lumsdaine and Stock (1998) consider a dynamic model in multivariate time series including I(0), I(1), and deterministically trending regressors. Kim (2011) extends the Perron and Zhu (2005) paper to large $(n, T)$ panel data with cross-sectional dependence. There are two potential limitations of these papers. First, a break point is assumed to exist. Second, in most papers, both regressors and error term are assumed to be stationary. Exceptions are Bai, Lumsdaine and Stock (1998) who discuss both stationary and nonstationary regressors and Perron and Zhu (2005) and Kim (2011) who discuss both stationary and nonstationary error terms. However, Bai, Lumsdaine and Stock (1998) and Perron and Zhu (2005) are concerned with the time series case. In addition, Perron and Zhu (2005) and Kim (2011) only discuss the case where the regressor is a time trend. In a pure time series framework, Nunes et al. (1995) and Bai (1998) show that when the disturbances follow an $I(1)$ process, there is a tendency to spuriously estimate a break point, in the middle of the sample, even though a break point does not exist. Recently, Hsu and Lin (2011) show that the spurious break still exists when a fixed effects estimator is used in panel data.

This paper studies the estimation of a change point in a panel data model with an autocorrelated regressor and an autocorrelated error (both of which can be stationary or nonstationary). This is done in case a change point is present or not present in the model. We focus on the change point estimation using the Ordinary Least Squares (OLS) and First Difference (FD) estimators. We establish the consistency and rate of convergence of these change point estimators. The assumption of the shrinking magnitude of the break is relaxed. More formally, the magnitude of the break in panel data is allowed to shrink to zero slower than in pure time series. The limiting distributions of the change point estimators are derived. We find that the FD estimator of the change point is robust to stationary or nonstationary regressors and error term, no matter whether a change point is present or not.

The paper is organized as follows: Section 2 introduces the model and assumptions. Section 3 proves the consistency of the change point using an OLS estimator. In addition, the limiting distribution of the OLS change point estimator is derived. Section 4 derives the consistency and limiting distribution of the change point using a FD estimator. Simulation results are presented in Section 5, while Section 6 provides the concluding remarks. Mathematical proofs and more simulation results are contained in the supplemental appendix, Baltagi, Kao and Liu (2012), and are available upon request from the authors. We use $L$ denotes a lag operator and $E$ a mathematical expectation, $\xrightarrow{d}$ to denote convergence in distribution, $\xrightarrow{p}$ convergence in probability and $[x]$ the largest integer $\leq x$. We write $W(r)$ as $W$ and the integral $\int_{\tau_1}^{\tau_2} W(r)\, dr$ as $\int_{\tau_1}^{\tau_2} W$ when there is no ambiguity over limits.

## 2 The Model and Assumptions

Consider the following panel regression with a change point at $k_0$ in the slope parameter,

$$y_{it} = \begin{cases} \alpha_1 + \beta_1 x_{it} + u_{it} & \text{for } t = 1, ..., k_0 \\ \alpha_2 + \beta_2 x_{it} + u_{it} & \text{for } t = k_0 + 1, ..., T \end{cases} \tag{1}$$

for $i = 1, ..., n$, where $y_{it}$ is the dependent variable and $x_{it}$ is the explanatory variable. For simplicity, we consider the case of one regressor besides a constant, but our results can be extended to the multiple regressors case. $\alpha_1$ and $\alpha_2$ are unknown intercept parameters and $\beta_1$ and $\beta_2$ are unknown slope parameters. $u_{it}$ is the disturbance term. The general case with fixed effects will be discussed in Section 4. Define $X_{it} = (1, x_{it})'$, $\gamma = (\alpha_1, \beta_1)'$ and $\delta = (\alpha_2 - \alpha_1, \beta_2 - \beta_1)'$. Equation (1) can be rewritten as

$$y_{it} = X_{it}'\gamma + X_{it}'\delta \cdot 1\{t > k_0\} + u_{it}, \qquad i = 1, \cdots, n; t = 1, \cdots, T, \tag{2}$$

where $1(\cdot)$ is an indicator function. If $\delta \neq 0$, there is a change at an *unknown* date $k_0$ where $k_0 = [\tau_0 T]$ for some $\tau_0 \in (0, 1)$. If $\delta = 0$, there is no change in the model and hence $k_0 = [\tau_0 T]$ for $\tau_0 = 0$ or 1. We aim to estimate the change point $k_0$. Following Joseph and Wolfson (1992) and Bai (2010), we assume the common break point $k_0$ that is the same for all $i = 1, ..., n$. As discussed in Bai (2010), "Theoretically, common break is a more restrictive assumption than the random breaks of Joseph and Wolfson (1993). Nevertheless, when break points are indeed common, as a result of common shocks or policy shift affecting every individual, imposing the constraint gives a more precise estimation. Computationally, common break model is much simpler. Furthermore, even if each series has its own break point, the common break method can be considered as estimating the mean of the random break points, which can be useful." This common break assumption has been used in empirical research such as Murray and Papell (2000). Bai, Lumsdaine and Stock (1998) is another important paper on common breaks in the multivariate time series literature.

Following Baltagi, Kao and Liu (2008), we consider the case where $x_{it}$ and $u_{it}$ are AR(1) processes, i.e.,

$$x_{it} = \lambda x_{i,t-1} + \varepsilon_{it} \tag{3}$$

and

$$u_{it} = \rho u_{it-1} + e_{it}, \tag{4}$$

where $-1 < \lambda \leq 1$ and $-1 < \rho \leq 1$. Clearly, $u_{it}$ is stationary when $|\rho| < 1$, and nonstationary when $\rho = 1$. Similar to the assumptions in Kim (2011), we assume that $\varepsilon_{it}$ and $e_{it}$ are linear processes that satisfy the following assumptions:

**Assumption 1** *For each $i$, $\varepsilon_{it}$ is such that $E\left|\varepsilon_{it}\right|^{2+\phi} < \infty$, $\phi > 0$ and $\varepsilon_{it} = c\left(L\right)\eta_{it}$, where $\eta_{it} \sim iid\left(0, \sigma_\eta^2\right)$ and $c\left(L\right) = \sum_{j=0}^{\infty} c_j L^j$ with $\sum_{j=0}^{\infty} j\left|c_j\right| < M$ and $c(z) \neq 0$ for all $|z| \leq 1$. $M$ is a generic finite positive number which depends on neither $T$ nor $n$.*

**Assumption 2** *For each $i$, $e_{it}$ is such that $E\left|e_{it}\right|^{2+\phi} < \infty$, $\phi > 0$ and $e_{it} = d\left(L\right)\zeta_{it}$, where $\zeta_{it} \sim iid\left(0, \sigma_\zeta^2\right)$ and $d\left(L\right) = \sum_{j=0}^{\infty} d_j L^j$ with $\sum_{j=0}^{\infty} j\left|d_j\right| < M$ and $d(z) \neq 0$ for all $|z| \leq 1$. $M$ is a generic finite positive number which depends on neither $T$ nor $n$.*

**Assumption 3** *We also assume $\varepsilon_{it}$ and $e_{it}$ are independent.*

Assumptions 1 and 2 imply that the partial sum processes $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\varepsilon_{it}$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}e_{it}$ satisfy the following multivariate invariance principle:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\varepsilon_{it} \xrightarrow{d} \sigma_\varepsilon W_{\varepsilon i}$$

and

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}e_{it} \xrightarrow{d} \sigma_e W_{ei}$$

as $T \to \infty$ for all $i$, where $[W_{ei}, W_{\varepsilon i}]'$ is a standardized Brownian motion. $\sigma_\varepsilon^2 = \sigma_\eta^2 \sum_{j=0}^{\infty} c_j^2$ and $\sigma_e^2 = \sigma_\zeta^2 \sum_{j=0}^{\infty} d_j^2$ are the long-run variances of $\varepsilon_{it}$ and $e_{it}$, respectively. When $n = 1$, $[W_{ei}, W_{\varepsilon i}]'$ reduces to $[W_e, W_\varepsilon]'$. Assumptions 1 and 2 imply cross-sectional independence. Together with Assumptions 3, we know that the innovation for the regressor $\varepsilon_{it}$ and the regression error $e_{js}$ are independent for all $i \neq j$ and $t \neq s$. The regressor is independent of the error at all leads and lags and hence is completely exogenous. We introduce this stringent assumption to simplify various technical difficulties arising from the complexity of panel model with a structural break. In practical applications, this independence assumption is restrictive and may not hold. To relax this assumption, one could follow Kim (2011) and include a factor loadings structure in the error term. Kim (2011) estimates a common deterministic time trend break in panel data, and finds that the strong cross sectional dependence generated by the common factors reduces the rate of convergence and thus eliminates some of the benefits of panel data. While this is beyond the scope of this study, we nevertheless conducted some simulations in Section 5 to investigate the impact of cross sectional dependence.

# 3 Ordinary Least Squares Estimator

Let $Y^{(i)} = (y_{i1}, \cdots, y_{iT})'$, $X^{(i)} = (X_{i1}, \cdots, X_{iT})'$, $Z_0^{(i)} = (0, \cdots, 0, X_{i,k_0+1}, \cdots, X_{iT})'$, and $U^{(i)} = (u_{i1}, \cdots, u_{iT})'$ denote the stacked data and error for individual $i$ over the time periods observed. Stacking the data over all individuals, we get $Y = \begin{pmatrix} Y^{(1)\prime}, & \cdots, & Y^{(n)\prime} \end{pmatrix}'$, $X = \begin{pmatrix} X^{(1)\prime}, & \cdots, & X^{(n)\prime} \end{pmatrix}'$, $Z_0 = \begin{pmatrix} Z_0^{(1)\prime}, & \cdots, & Z_0^{(n)\prime} \end{pmatrix}'$ and $U = \begin{pmatrix} U^{(1)\prime}, & \cdots, & U^{(n)\prime} \end{pmatrix}'$. All of these vectors are of dimension $Tn \times 1$. Using this notation, (2) can be written in matrix form as

$$Y = X\gamma + Z_0\delta + U. \tag{5}$$

For any possible change point $k$, we define the matrices $Z_k^{(i)} = (0, \cdots, 0, x_{i,k+1}, \cdots, x_{iT})'$ and $Z_k = \begin{pmatrix} Z_k^{(1)\prime}, & \cdots, & Z_k^{(n)\prime} \end{pmatrix}'$. The OLS estimator of the slope parameters which depend upon $k$, is given by:

$$\begin{pmatrix} \hat{\gamma}_k \\ \hat{\delta}_k \end{pmatrix} = \begin{bmatrix} X'X & X'Z_k \\ Z_k'X & Z_k'Z_k \end{bmatrix}^{-1} \begin{bmatrix} X'Y \\ Z_k'Y \end{bmatrix} \tag{6}$$

and the corresponding OLS sum of squared residuals is given by:

$$SSR_{OLS}(k) = (Y - X\hat{\gamma}_k - Z_k\hat{\delta}_k)'(Y - X\hat{\gamma}_k - Z_k\hat{\delta}_k).$$

The OLS estimate of the change point is obtained as follows:

$$\hat{k}_{OLS} = \arg\min_{1 \le k \le T} SSR_{OLS}(k).$$

Define $V_{OLS}(k) = SSR_{OLS} - SSR_{OLS}(k) = \hat{\delta}_k^2 Z_k M Z_k$, where $SSR_{OLS}$ is the sum of squared residuals for the case of no break, i.e., $k_0 = T$ and $M = I - X(X'X)^{-1}X'$. As shown in Bai (1997), to minimize $SSR_{OLS}(k)$ is equivalent to maximizing $V_{OLS}(k)$. This implies that

$$\hat{k}_{OLS} = \arg\min_{1 \le k \le T} SSR_{OLS}(k) = \arg\max_{1 \le k \le T} V_{OLS}(k)$$

and

$$\hat{\tau}_{OLS} = \frac{\hat{k}_{OLS}}{T}.$$

## 3.1 When there is a break point

In this section, we show the consistency of the change point estimator and derive its rate of convergence when a break point exists, i.e., $\tau_0 \in (0, 1)$ and $\delta$ is a nonzero constant. We assume

**Assumption 4** $\beta_2 - \beta_1 \neq 0$.

Assumption 4 ensures that we have a one time break in the systematic part and that the pre and post break samples are not asymptotically negligible, which is a standard assumption needed to derive any useful asymptotic result.[1] In pure time series, an assumption of shrinking magnitude of the break is needed to derive the limiting distribution of the change point estimator. Otherwise not enough information is provided by the time series data to identify the true change point. As discussed in Feng, Kao and Lazarova (2009), this shrinking magnitude assumption of the break is not needed in panel data. For example, when many economies are observed each year, one can identify the structural change by simply examining the activities of these countries over years. Therefore, the large number of observations on countries can be used to derive the asymptotics around the true break date. Following Feng et al. (2009), the magnitude of the break $\beta_2 - \beta_1$ is assumed to be fixed throughout the paper. The following theorem shows that $\hat{\tau}_{OLS}$ may not be always consistent in the pure time series case.

**Theorem 1** *Under Assumptions 1-4, in the pure time series case, where $n = 1$, for $\tau_0 \in (0, 1)$ and as $T \to \infty$, we have the following results:*

*(a) when $|\rho| < 1$ and $|\lambda| < 1$,*

$$\hat{\tau}_{OLS} \xrightarrow{p} \tau_0;$$

*(b) when $\rho = 1$ and $|\lambda| < 1$,*

$$\hat{\tau}_{OLS} \xrightarrow{d} \arg \max_{\tau \in [\underline{\tau}, \bar{\tau}]} \frac{1}{\tau (1 - \tau)} \left[ \int_{\tau}^{1} W_e - (1 - \tau) \int_{0}^{1} W_e \right]^2;$$

*(c) when $|\rho| < 1$ and $\lambda = 1$,*

$$\hat{\tau}_{OLS} \xrightarrow{p} \tau_0;$$

*(d) when $\lambda = \rho = 1$,*

$$\hat{\tau}_{OLS} \xrightarrow{d} \arg \max_{\tau \in [\underline{\tau}, \bar{\tau}]} \left[ \tilde{F} (\tau, \tau_0) + S (\tau) \right]' F (\tau)^{-1} \left[ \tilde{F} (\tau, \tau_0) + S (\tau) \right],$$

*where*

$$F (\tau) = P (\tau) - P (\tau) P (0)^{-1} P (\tau),$$

[1] In general, for any $x_{it}$ with nonzero mean $\theta$, the model in Equation (1) can be rewritten as

$$y_{it} = \begin{cases} (\alpha_1 + \beta_1 \theta) + \beta_1 (x_{it} - \theta) + u_{it} & \text{for } t = 1, ..., k_0 \\ (\alpha_2 + \beta_2 \theta) + \beta_2 (x_{it} - \theta) + u_{it} & \text{for } t = k_0 + 1, ..., T \end{cases},$$

where the new regressor is zero mean again. From the equation above, we can see that a change in the slope implies a change in the intercept, as long as the initial regressor $x_{it}$ has nonzero mean.

$$\tilde{F}(\tau,\tau_0) = \begin{cases} [P(0) - P(\tau)]\, P(0)^{-1} P(\tau_0) \begin{pmatrix} 0 \\ \beta_2 - \beta_1 \end{pmatrix} & \text{if } \tau \leq \tau_0 \\ P(\tau)\, P(0)^{-1} [P(0) - P(\tau_0)] \begin{pmatrix} 0 \\ \beta_2 - \beta_1 \end{pmatrix} & \text{if } \tau > \tau_0 \end{cases},$$

$$S(\tau) = \begin{pmatrix} \sigma_e \int_\tau^1 W_e \\ \sigma_\varepsilon \sigma_e \int_\tau^1 W_e W_\varepsilon \end{pmatrix} - P(\tau)\, P(0)^{-1} \begin{pmatrix} \sigma_e \int_0^1 W_e \\ \sigma_\varepsilon \sigma_e \int_0^1 W_e W_\varepsilon \end{pmatrix},$$

$$\text{and } P(\tau) = \begin{pmatrix} 1-\tau & \sigma_\varepsilon \int_\tau^1 W_\varepsilon \\ \sigma_\varepsilon \int_\tau^1 W_\varepsilon & \sigma_\varepsilon^2 \int_\tau^1 W_\varepsilon^2 \end{pmatrix}.$$

Theorem 1 implies that $\hat{\tau}_{OLS}$ is consistent when $|\rho| < 1$ but inconsistent when $\rho = 1$ if there is a break in the pure time series case. To be more specific, when $|\rho| < 1$ and $|\lambda| < 1$, this is consistent with the findings in Nunes et al. (1995). When $|\rho| < 1$ and $\lambda = 1$, as discussed in Bai (1996), both $\hat{\tau}_{OLS}$ and $\hat{k}_{OLS}$ are consistent in this cointegration model. When $\rho = 1$ and $|\lambda| < 1$, $\hat{\tau}_{OLS}$ converges to a function that does not depend on the true value of the break fraction $\tau_0$. Similarly, when $\rho = \lambda = 1$, $\hat{\tau}_{OLS}$ converges to a function that includes $\tilde{F}(\tau,\tau_0)' F(\tau)^{-1} \tilde{F}(\tau,\tau_0)$ and $S(\tau)' F(\tau)^{-1} S(\tau)$. One can show that the function $\tilde{F}(\tau,\tau_0)' F(\tau)^{-1} \tilde{F}(\tau,\tau_0)$ will be maximized at $\tau_0$, but the function $S(\tau)' F(\tau)^{-1} S(\tau)$ does not depend on $\tau_0$ at all. This implies that $\hat{\tau}_{OLS}$ is inconsistent when $\rho = 1$, whether $|\lambda| < 1$ or $\lambda = 1$. Overall, using the relationship that $\hat{\tau}_{OLS} = \hat{k}_{OLS}/T$, we know that $\hat{k}_{OLS}$ is inconsistent except in a cointegration model. In fact, if the magnitude of the break $\delta$ is fixed, the asymptotic distribution of $\hat{k}_{OLS} - k_0$ depends upon the underlying distribution of the regressors and error, e.g., Picard (1985) and Bai (1997). This difficulty can be overcome with panel data. As shown in Bai (2010), the consistency of $\hat{k}_{OLS}$ can be established in a mean shift panel data model. The theorem below extends Bai (2010)'s results to the case where the regressor and the error term are allowed to be $I(0)$ or $I(1)$ processes.

**Theorem 2** *Under Assumptions 1-4, for $\tau_0 \in (0,1)$ and as $(n,T) \to \infty$, we have the following results:*

*(a) when $|\rho| < 1$ and $|\lambda| < 1$, $\hat{k}_{OLS} - k_0 = O_p\left(\frac{1}{n}\right)$;*

*(b) when $\rho = 1$ and $|\lambda| < 1$, $\hat{k}_{OLS} - k_0 = O_p\left(\frac{T}{n}\right)$ if $\frac{T}{n} \to 0$;*

*(c) when $|\rho| < 1$ and $\lambda = 1$, $\hat{k}_{OLS} - k_0 = O_p\left(\frac{1}{nT}\right)$;*

*(d) when $\lambda = \rho = 1$, $\hat{k}_{OLS} - k_0 = O_p\left(\frac{1}{nT}\right)$.*

Theorem 2 shows that the consistency of $\hat{k}_{OLS}$ can be achieved even for a fixed $\delta$, as long as $n$ is large. That is, large cross-sectional dimension will create enough information to identify the true change point. Unlike the time series set-up, for fixed $\delta$, $\hat{k}_{OLS}$ is *consistent* with convergence speed of $n$ when $|\rho| < 1$ and $|\lambda| < 1$. However, when $\rho = 1$ and $|\lambda| < 1$, consistency of $\hat{k}_{OLS}$ needs $\frac{T}{n} \to 0$. When $\lambda = 1$, no matter whether $|\rho| < 1$ or $\rho = 1$, $\hat{k}_{OLS}$ is consistent with $nT$ convergence speed. This is because when $|\rho| < 1$ and $\lambda = 1$, $x_{it}$ is an $I(1)$ process that is strong enough to dominate the $I(0)$ error term. When $\lambda = \rho = 1$, large $n$ helps to reduce the noise caused by the $I(1)$ error term as in the panel spurious regression (e.g., Kao, 1999). Besides, the consistency of the estimator of $k_0$ in panel data has a different meaning. For fixed $T$, the fixed $k_0$ can be regarded as a parameter. Theorem 2 shows that as $n \to \infty$, $\hat{k}_{OLS} \to k_0$. For large $T$, $k_0$ increases with $T$, however, the distance between the estimate and the true value vanishes, i.e. $\hat{k}_{OLS} - k_0 \to 0$ as $(n, T) \to \infty$. This consistency concept is different from the one in the standard textbooks. Since $\hat{\tau}_{OLS} = \hat{k}_{OLS}/T$, Theorem 2 implies the following proposition.

**Proposition 1** *Under Assumptions 1-4, for $\tau_0 \in (0, 1)$ and as $(n, T) \to \infty$, we have the following results:*

*(a) when $|\rho| < 1$ and $|\lambda| < 1$,* $\hat{\tau}_{OLS} - \tau_0 = O_p\left(\frac{1}{nT}\right)$;

*(b) when $\rho = 1$ and $|\lambda| < 1$,* $\hat{\tau}_{OLS} - \tau_0 = O_p\left(\frac{1}{n}\right)$ *if* $\frac{T}{n} \to 0$;

*(c) when $|\rho| < 1$ and $\lambda = 1$,* $\hat{\tau}_{OLS} - \tau_0 = O_p\left(\frac{1}{nT^2}\right)$;

*(d) when $\lambda = \rho = 1$,* $\hat{\tau}_{OLS} - \tau_0 = O_p\left(\frac{1}{nT^2}\right)$.

Proposition 1 shows that the fraction estimate $\hat{\tau}_{OLS}$ is always *consistent* with a convergence speed of at least $n$. Comparing this result with that in Theorem 1, it is clear that $\hat{\tau}_{OLS}$ in a panel data setting is robust to different values of $\lambda$ and $\rho$. This highlights the difference between the results in the panel data case and those in the pure time series case.

Define $D_{nT} = \begin{pmatrix} nT & 0 \\ 0 & nT^2 \end{pmatrix}$. With the estimator $\hat{k}$, the asymptotics of $\hat{\gamma}_{\widehat{k}_{OLS}} = \hat{\gamma}\left(\widehat{k}_{OLS}\right)$ and $\widehat{\delta}_{\widehat{k}_{OLS}} = \widehat{\delta}\left(\widehat{k}_{OLS}\right)$ can be established as follows:

**Theorem 3** *Under Assumptions 1-4, as $(n, T) \to \infty$, we have the following results:*

*(a) when $|\rho| < 1$ and $|\lambda| < 1$,*

$$\begin{pmatrix} \sqrt{nT}\left(\hat{\gamma}_{\widehat{k}_{OLS}} - \gamma\right) \\ \sqrt{nT}\left(\hat{\delta}_{\widehat{k}_{OLS}} - \delta\right) \end{pmatrix} \xrightarrow{d} N\left(0, \sigma_e^2 \begin{pmatrix} \frac{1}{\tau_0} & 0 & -\frac{1}{\tau_0} & 0 \\ 0 & \frac{\left(1-\lambda^2\right)^2}{\tau_0(1-\lambda\rho)^2\sigma_\varepsilon^2} & 0 & -\frac{\left(1-\lambda^2\right)^2}{\tau_0(1-\lambda\rho)^2\sigma_\varepsilon^2} \\ -\frac{1}{\tau_0} & 0 & \frac{1}{\tau_0(1-\tau_0)} & 0 \\ 0 & -\frac{\left(1-\lambda^2\right)^2}{\tau_0(1-\lambda\rho)^2\sigma_\varepsilon^2} & 0 & \frac{\left(1-\lambda^2\right)^2}{\tau_0(1-\tau_0)(1-\lambda\rho)^2\sigma_\varepsilon^2} \end{pmatrix}\right);$$

*(b) when* $\rho = 1$ *and* $|\lambda| < 1$,

$$\begin{pmatrix} T^{-1} D_{nT}^{1/2} \left( \hat{\gamma}_{\widehat{k}_{OLS}} - \gamma \right) \\ T^{-1} D_{nT}^{1/2} \left( \hat{\delta}_{\widehat{k}_{OLS}} - \delta \right) \end{pmatrix} \xrightarrow{d} N \left( 0, \sigma_e^2 \begin{pmatrix} \frac{1}{3}\tau_0 & 0 & \frac{1}{6}\tau_0 & 0 \\ 0 & \frac{(\lambda+1)^2}{2\sigma_\varepsilon^2} & 0 & -\frac{(\lambda+1)^2}{2\sigma_\varepsilon^2} \\ \frac{1}{6}\tau_0 & 0 & \frac{1}{3} & 0 \\ 0 & -\frac{(\lambda+1)^2}{2\sigma_\varepsilon^2} & 0 & \frac{(\lambda+1)^2}{\sigma_\varepsilon^2(1-\tau_0)} \end{pmatrix} \right);$$

*(c) when* $|\rho| < 1$ *and* $\lambda = 1$,

$$\begin{pmatrix} D_{nT}^{1/2} \left( \hat{\gamma}_{\widehat{k}_{OLS}} - \gamma \right) \\ D_{nT}^{1/2} \left( \hat{\delta}_{\widehat{k}_{OLS}} - \delta \right) \end{pmatrix} \xrightarrow{d} N \left( 0, \sigma_e^2 \begin{pmatrix} \frac{1}{\tau_0} & 0 & -\frac{1}{\tau_0} & 0 \\ 0 & \frac{2}{\tau_0^2(1-\rho)^2\sigma_\varepsilon^4} & 0 & -\frac{2}{\tau_0^2(1-\rho)^2\sigma_\varepsilon^4} \\ -\frac{1}{\tau_0} & 0 & \frac{1}{\tau_0(1-\tau_0)} & 0 \\ 0 & -\frac{2}{\tau_0^2(1-\rho)^2\sigma_\varepsilon^4} & 0 & \frac{2}{\tau_0^2\left(1-\tau_0^2\right)(1-\rho)^2\sigma_\varepsilon^4} \end{pmatrix} \right);$$

*(d) when* $\lambda = \rho = 1$,

$$\begin{pmatrix} T^{-1} D_{nT}^{1/2} \left( \hat{\gamma}_{\widehat{k}_{OLS}} - \gamma \right) \\ T^{-1} D_{nT}^{1/2} \left( \hat{\delta}_{\widehat{k}_{OLS}} - \delta \right) \end{pmatrix} \xrightarrow{d} N \left( 0, \sigma_e^2 \begin{pmatrix} \frac{1}{3}\tau_0 & 0 & \frac{1}{6}\tau_0 & 0 \\ 0 & \frac{2}{3\sigma_\varepsilon^4} & 0 & -\frac{2(1-\tau_0)}{3(1+\tau_0)\sigma_\varepsilon^4} \\ \frac{1}{6}\tau_0 & 0 & \frac{1}{3} & 0 \\ 0 & -\frac{2(1-\tau_0)}{3(1+\tau_0)\sigma_\varepsilon^4} & 0 & \frac{4}{3(1+\tau_0)^2\sigma_\varepsilon^4} \end{pmatrix} \right).$$

As shown in Theorem 2 and Proposition 1, $\hat{k}_{OLS}$ and $\hat{\tau}_{OLS}$ have faster speed so that $\hat{\gamma}_{\widehat{k}_{OLS}}$ and $\widehat{\delta}_{\widehat{k}_{OLS}}$ have the same distributions as $\hat{\gamma}_{k_0}$ and $\widehat{\delta}_{k_0}$. Hence Theorem 3 implies that the asymptotic distributions of $\hat{\gamma}_{\widehat{k}_{OLS}}$ and $\widehat{\delta}_{\widehat{k}_{OLS}}$ can be established as if the change point were *known*. Let $W$ be a two-sided Brownian motion on the real line, that is $W(0) = 0$ and $W(r) = W_1(r)$ for $r > 0$ and $W(r) = W_2(r)$ for $r < 0$, where $W_1$ and $W_2$ are two independent Brownian motions, e.g., see Picard (1985). We have the following Theorem:

**Theorem 4** *Under Assumptions 1-4, for* $\tau_0 \in (0, 1)$, *and as* $(n, T) \to \infty$, *we have the following results:*

*(a) when* $|\rho| < 1$ *and* $|\lambda| < 1$,

$$b_1 n(\hat{k}_{OLS} - k_0) \xrightarrow{d} \arg\max_r \left[ W(r) - \frac{|r|}{2} \right],$$

$$\text{where } b_1 = \frac{\left[ (\alpha_1-\alpha_2)^2 + \frac{\sigma_\varepsilon^2}{1-\lambda^2}(\beta_1-\beta_2)^2 \right]^2}{\sigma_e^2 \left[ (\alpha_1-\alpha_2)^2 + \frac{\sigma_\varepsilon^2}{(1-\rho\lambda)^2}(\beta_1-\beta_2)^2 \right]};$$

*(b) when* $\rho = 1$ *and* $|\lambda| < 1$,

$$\frac{b_2 n}{T} \left( \hat{k}_{OLS} - k_0 \right) \xrightarrow{d} \arg\max_r \left[ W(r) - \frac{|r|}{2} \right],$$

$$\text{where } b_2 = \frac{\left[ (\alpha_1-\alpha_2)^2 + \frac{\sigma_\varepsilon^2}{1-\lambda^2}(\beta_1-\beta_2)^2 \right]^2}{\frac{\tau_0 \sigma_\varepsilon^2 \sigma_e^2}{(1-\lambda)^2}(\beta_1-\beta_2)^2};$$

(c) *when* $|\rho| < 1$ *and* $\lambda = 1$,

$$b_3 nT\left(\hat{k}_{OLS} - k_0\right) \xrightarrow{d} \arg\max_r \left[W(r) - \frac{|r|}{2}\right],$$

*where* $b_3 = \tau_0 (1-\rho)^2 (\beta_1 - \beta_2)^2 \sigma_\varepsilon^2 / \sigma_e^2$;

(d) *when* $\lambda = \rho = 1$,

$$b_4 nT\left(\hat{k}_{OLS} - k_0\right) \xrightarrow{p} 0,$$

*where* $b_4 = \tau_0 (\beta_1 - \beta_2)^2 \sigma_\varepsilon^2$.

Theorem 4 shows the asymptotic distribution for $\hat{k}_{OLS}$ in the panel data case. The density function of $\arg\max_r \left[W(r) - \frac{|r|}{2}\right]$ has been well studied by Bhattacharya (1987), Picard (1985), Yao (1987), and Bai (1997). It is symmetric about the origin. For pure time series, Bai (1997) found that the asymptotic distribution of the change point estimator is non-symmetric when regressors or disturbances are nonstationary.[2] Theorem 4 finds the same result for panel data. More specifically, in Theorem 4, the convergence speed depends upon the true break fraction $\tau_0$ except for the first case when $|\rho| < 1$ and $|\lambda| < 1$. For example, in case (b), $b_2$ is proportional to $1/\tau_0$. This means that a smaller $\tau_0$ is easier to identify than a larger $\tau_0$. Similarly, in cases (c) and (d), $b_3$ and $b_4$ are proportional to $\tau_0$. It implies that a larger $\tau_0$ is easier to identify than a smaller $\tau_0$. These findings will be further discussed in the Monte Carlo simulation section. For case (d) when $\lambda = \rho = 1$, Theorem 4 shows that $nT\left(\hat{k}_{OLS} - k_0\right) = o_p(1)$. This is an improvement on the result $nT\left(\hat{k}_{OLS} - k_0\right) = O_p(1)$ in Theorem 2. However, the limiting distribution in this spurious regression case can not be derived. Overall, Theorem 4 implies that the distribution of $\hat{k}_{OLS}$ is not robust to different values of $\lambda$, $\rho$ and $\tau_0$. Realizing this disadvantage, we consider a FD-based robust break point estimator in Section 4.

### 3.2 When there is no break point

In this section, we discuss the consistency of the break fraction estimate when there is no break point, i.e., $\delta = 0$.

**Theorem 5** *Under Assumptions 1-3, in the pure time series case, where* $n = 1$ *and when* $\tau_0 = 0$ *or* 1, *as* $T \to \infty$, *we have the following results:*

(a) *when* $|\rho| < 1$ *and* $|\lambda| < 1$,

$$\hat{\tau}_{OLS} \xrightarrow{p} \{0, 1\};$$

[2] We thank a referee pointing this out.

*(b) when $\rho = 1$ and $|\lambda| < 1$,*

$$\hat{\tau}_{OLS} \xrightarrow{d} \arg\max_{\tau \in [\underline{\tau}, \bar{\tau}]} \frac{1}{\tau(1-\tau)} \left[ \int_{\tau}^{1} W_e - (1-\tau) \int_{0}^{1} W_e \right]^2;$$

*(c) when $|\rho| < 1$ and $\lambda = 1$,*

$$\hat{\tau}_{OLS} \xrightarrow{p} \{0, 1\};$$

*(d) when $\lambda = \rho = 1$,*

$$\hat{\tau}_{OLS} \xrightarrow{d} \arg\max_{\tau \in [\underline{\tau}, \bar{\tau}]} S(\tau)' F(\tau)^{-1} S(\tau),$$

*where $S(\tau)$ and $F(\tau)$ are defined in Theorem 1.*

Theorem 5 implies that $\hat{\tau}_{OLS}$ is consistent when $|\rho| < 1$, but not consistent when $\rho = 1$ in the pure time series case. This is consistent with the findings in the previous literature. To be more specific, when $|\rho| < 1$, as discussed in Nunes et al. (1995), $\hat{\tau}_{OLS}$ converges to 0 or 1 when a break point does not exist. However, when $\rho = 1$, Nunes et al. (1995) and Bai (1998) show that there is a tendency to spuriously estimate a break point in the middle of the sample when the disturbances follow an $I(1)$ process, even though a break point does not actually exist. For the panel data case, we have the following theorem:

**Theorem 6** *Under Assumptions 1-3, for $\tau_0 = 0$ or $1$, $\hat{\tau}_{OLS} \xrightarrow{d} \arg\max M^*(\tau)$ as $(n, T) \to \infty$, where $M^*(\tau) = R'(\tau) Q^{-1}(\tau) R(\tau) + [R(1) - R(\tau)]' [Q(1) - Q(\tau)]^{-1} [R(1) - R(\tau)]$, where $Q(\tau)$ and $R(\tau)$ are defined below for each case.*

*(a) If $|\rho| < 1$ and $|\lambda| < 1$, $Q(\tau) = \begin{pmatrix} \tau & 0 \\ 0 & \frac{\tau\sigma_\varepsilon^2}{1-\lambda^2} \end{pmatrix}$, $Q(1) - Q(\tau) = \begin{pmatrix} 1-\tau & 0 \\ 0 & \frac{(1-\tau)\sigma_\varepsilon^2}{1-\lambda^2} \end{pmatrix}$, $R(\tau) = N\left(0, \sigma_\varepsilon^2 \sigma_e^2 \begin{pmatrix} \tau & 0 \\ 0 & \frac{\tau}{(1-\rho\lambda)^2} \end{pmatrix}\right)$ and $R(1) - R(\tau) = N\left(0, \sigma_\varepsilon^2 \sigma_e^2 \begin{pmatrix} 1-\tau & 0 \\ 0 & \frac{1-\tau}{(1-\rho\lambda)^2} \end{pmatrix}\right)$. With probability 1, $M^*(\tau) < M^*(0)$, and $M^*(\tau) < M^*(1)$ for every $0 < \tau < 1$.*

*(b) If $\rho = 1$ and $|\lambda| < 1$, $Q(\tau) = \begin{pmatrix} \tau & 0 \\ 0 & \frac{\tau\sigma_\varepsilon^2}{1-\lambda^2} \end{pmatrix}$, $Q(1) - Q(\tau) = \begin{pmatrix} 1-\tau & 0 \\ 0 & \frac{(1-\tau)\sigma_\varepsilon^2}{1-\lambda^2} \end{pmatrix}$, $R(\tau) = N\left(0, \sigma_\varepsilon^2 \sigma_e^2 \begin{pmatrix} \frac{k\tau^2}{3} & 0 \\ 0 & \frac{\tau^2}{2(1-\lambda)^2} \end{pmatrix}\right)$ and $R(1) - R(\tau) = N\left(0, \sigma_\varepsilon^2 \sigma_e^2 \begin{pmatrix} \frac{(T-k)(1-\tau)(1+2\tau)}{3} & 0 \\ 0 & \frac{(1-\tau)(1+\tau)}{2(1-\lambda)^2} \end{pmatrix}\right)$. With probability 1, $M^*(0) < M^*(\tau) < M^*(1)$ for every $0 < \tau < 1$.*

*(c) If $|\rho| < 1$ and $\lambda = 1$, $Q(\tau) = \begin{pmatrix} \tau & 0 \\ 0 & \frac{\tau^2\sigma_\varepsilon^2}{2} \end{pmatrix}$, $Q(1) - Q(\tau) = \begin{pmatrix} 1-\tau & 0 \\ 0 & \frac{(1-\tau^2)\sigma_\varepsilon^2}{2} \end{pmatrix}$, $R(\tau) = N\left(0, \sigma_\varepsilon^2 \sigma_e^2 \begin{pmatrix} \tau & 0 \\ 0 & \frac{\tau^2}{2(1-\rho)^2} \end{pmatrix}\right)$ and $R(1) - R(\tau) = N\left(0, \sigma_\varepsilon^2 \sigma_e^2 \begin{pmatrix} 1-\tau & 0 \\ 0 & \frac{1-\tau^2}{2(1-\rho)^2} \end{pmatrix}\right)$. With probability 1, $M^*(\tau) < M^*(0)$, and $M^*(\tau) < M^*(1)$ for every $0 < \tau < 1$.*

(d) *If* $\rho = \lambda = 1$, $Q(\tau) = \begin{pmatrix} \tau & 0 \\ 0 & \frac{\tau^2\sigma_\varepsilon^2}{2} \end{pmatrix}$, $Q(1) - Q(\tau) = \begin{pmatrix} 1-\tau & 0 \\ 0 & \frac{(1-\tau^2)\sigma_\varepsilon^2}{2} \end{pmatrix}$,
$R(\tau) = N\left(0, \sigma_\varepsilon^2\sigma_e^2 \begin{pmatrix} \frac{\tau^3}{3} & 0 \\ 0 & \frac{\tau^4}{6} \end{pmatrix}\right)$ *and* $R(1) - R(\tau) = N\left(0, \sigma_\varepsilon^2\sigma_e^2 \begin{pmatrix} \frac{(1-\tau)^2(1+2\tau)}{3} & 0 \\ 0 & \frac{(1-\tau)^2(1+2\tau+3\tau^2)}{6} \end{pmatrix}\right)$.
*With probability 1,* $M^*(\tau) > M^*(0)$ *and* $M^*(\tau) > M^*(1)$ *for every* $0 < \tau < 1$.

First of all, Theorem 6 shows that when there is no break point in the model, $\hat{\tau}_{OLS} \xrightarrow{p} \{0, 1\}$ if $|\rho| < 1$, whether $|\lambda| < 1$ or $\lambda = 1$. Secondly, $\hat{\tau}_{OLS} \xrightarrow{p} 1$ if $\rho = 1$ and $|\lambda| < 1$. As we will see in the simulation results in Section 5, the empirical distribution of $\hat{k}$ is not symmetric and the highest probability mass of $\hat{k}$ occurs at the right tail. This is because the signal $x_{it}$ is a stationary $I(0)$ process and the error term $u_{it}$ is a nonstationary $I(1)$ process when $\rho = 1$ and $|\lambda| < 1$. The error term dominates the signal and we are actually checking if there is a break in the error term. And of course, the answer is no. For an $I(1)$ process, the variation increases as $t$ increases. Hence at the right tail where $k$ is close to $T$, it looks more like a jump. Finally, when there is no break in the model and $\rho = \lambda = 1$, the spurious break problem that is found in Nunes et al. (1995) and Bai (1998) in the time series case also exists in the panel data case. This is consistent with the findings by Hsu and Lin (2011).

# 4 First Difference Estimator

With individual effects, the panel regression model in Equation (1) becomes

$$y_{it} = \begin{cases} \alpha_1 + \beta_1 x_{it} + \mu_i + u_{it} & \text{for } t = 1, ..., k_0 \\ \alpha_2 + \beta_2 x_{it} + \mu_i + u_{it} & \text{for } t = k_0 + 1, ..., T \end{cases} \tag{7}$$

for $i = 1, ..., n$. After the within transformation, Equation (7) becomes

$$\tilde{y}_{it} = \begin{cases} (\alpha_1 - \alpha_2)\frac{T-k_0}{T} + \beta_1 \tilde{x}_{it} + \eta_{1i} + \tilde{u}_{it} & \text{for } t = 1, ..., k_0 \\ (\alpha_2 - \alpha_1)\frac{k_0}{T} + \beta_2 \tilde{x}_{it} + \eta_{2i} + \tilde{u}_{it} & \text{for } t = k_0 + 1, ..., T, \end{cases}$$

where $\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\tilde{x}_{it} = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\tilde{u}_{it} = u_{it} - \frac{1}{T}\sum_{t=1}^{T} u_{it}$, $\eta_{1i} = (\beta_1 - \beta_2)\left(\frac{1}{T}\sum_{t=k_0+1}^{T} x_{it}\right)$ and $\eta_{2i} = (\beta_2 - \beta_1)\left(\frac{1}{T}\sum_{t=1}^{k_0} x_{it}\right)$. We can see that although the within transformation wipes out the fixed effects $\mu_i$, it creates new fixed effects $\eta_{1i}$ and $\eta_{2i}$ which are due to subtracting the average of the regressor in a different regime with a different coefficient. To solve this problem, one could interact the individual dummies with the time change dummy. However this method is infeasible without knowing the true change point $k_0$. It implies that the parameters $\alpha_1$, $\alpha_2$, $\beta_1$ and $\beta_2$ cannot be identified using within estimation without knowing the true break date $k_0$. Hence we focus on the FD estimator instead to wipe out the individual effects. Applying the FD transformation, Equation (7) becomes

$$\Delta y_{it} = \begin{cases} \beta_1 \Delta x_{it} + \Delta u_{it} & \text{for } t = 2, ..., k_0 \\ (\alpha_2 - \alpha_1) + (\beta_2 - \beta_1) x_{it} + \Delta u_{it} & \text{for } t = k_0 + 1 \\ \beta_2 \Delta x_{it} + \Delta u_{it} & \text{for } t = k_0 + 2, ..., T \end{cases}, \tag{8}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta x_{it} = x_{it} - x_{i,t-1}$ and $\Delta u_{it} = u_{it} - u_{i,t-1}$. As we can see, the form of Equation (8) for cases $t = k_0 + 1$ and $t \geq k_0 + 2$ are different. As discussed in Feng et al. (2009), ignoring this difference will not change the estimation result for a large $T$. We apply the least squares estimate using the FD data $\{\Delta y_{it}, \Delta x_{it}\}_{t=2}^{T}$ to obtain $\hat{k}$. Therefore, Equation (8) can be approximately written as

$$\Delta y_{it} = \gamma_2 \Delta x_{it} + \delta_2 \Delta x_{it} \cdot 1\{t > k_0\} + \Delta u_{it}, \qquad i = 1, \cdots, n; t = 1, \cdots, T,$$

where $\gamma_2$ and $\delta_2$ are the second elements in $\gamma$ and $\delta$, i.e., $\gamma_2 = \beta_1$ and $\delta_2 = \beta_2 - \beta_1$.

Let $DY^{(i)} = (\Delta y_{i2}, \cdots, \Delta y_{iT})'$, $DX^{(i)} = (\Delta x_{i2}, \cdots, \Delta x_{iT})'$, $DZ_0^{(i)} = (0, \cdots, 0, \Delta x_{i,k_0+1}, \cdots, \Delta x_{iT})'$, and $DU^{(i)} = (\Delta u_{i2}, \cdots, \Delta u_{iT})'$ denote the stacked data and error for individual $i$ over the time periods observed. Stacking the data over all individuals, we get: $DY = \begin{pmatrix} DY^{(1)\prime}, & \cdots, & DY^{(n)\prime} \end{pmatrix}'$, $DX = \begin{pmatrix} DX^{(1)\prime}, & \cdots, & DX^{(n)\prime} \end{pmatrix}'$, $DZ_0 = \begin{pmatrix} DZ_0^{(1)\prime}, & \cdots, & DZ_0^{(n)\prime} \end{pmatrix}'$, and $DU = \begin{pmatrix} DU^{(1)\prime}, & \cdots, & DU^{(n)\prime} \end{pmatrix}'$. All of these vectors are of dimension $n(T-1) \times 1$. Using this notation, The model in Equation (8) can be rewritten in matrix form as:

$$DY = DX\gamma_2 + DZ_0\delta_2 + DU. \tag{9}$$

For any possible change point $k$, we define the matrices $DZ_k^{(i)} = (0, \cdots, 0, \Delta x_{i,k+1}, \cdots, \Delta x_{iT})'$ and $DZ_k = \begin{pmatrix} DZ_k^{(1)\prime}, & \cdots, & DZ_k^{(n)\prime} \end{pmatrix}'$. The FD estimator, which depends upon $k$, is given by:

$$\begin{pmatrix} \tilde{\gamma}_{2,k} \\ \tilde{\delta}_{2,k} \end{pmatrix} = \begin{bmatrix} DX'DX & DX'DZ_k \\ DZ_k'DX & DZ_k'DZ_k \end{bmatrix}^{-1} \begin{bmatrix} DX'DY \\ DZ_k'DY \end{bmatrix}$$

and the corresponding sum of squared residuals is given by:

$$SSR_{FD}(k) = (DY - DX\tilde{\gamma}_{2k} - DZ_k\tilde{\delta}_{2,k})'(DY - DX\tilde{\gamma}_{2k} - DZ_k\tilde{\delta}_{2,k}).$$

Define $V_{FD}(k) = SSR_{FD} - SSR_{FD}(k) = \tilde{\delta}_{2k}^2 DZ_k M_{DX} DZ_k$, where $SSR_{FD}$ is the sum of squared residuals for the case of no break, i.e., $k_0 = T$ and $M_{DX} = I - DX(DX'DX)^{-1}DX'$. Similar to the argument for the OLS estimator in Section 3, the FD estimator of the change point is given by

$$\hat{k}_{FD} = \arg\min_{2 \leq k \leq T} SSR_{FD}(k) = \arg\max_{2 \leq k \leq T} V_{FD}(k) \tag{10}$$

and

$$\hat{\tau}_{FD} = \frac{\hat{k}_{FD}}{T}.$$

One can see that $\Delta x_{it} = (\lambda - 1)\, x_{i,t-1} + \varepsilon_{it}$ and $\Delta u_{it} = (\rho - 1)\, u_{it-1} + e_{it}$ are $I(0)$ processes. This implies that first differencing will always transform the data into a case with stationary regressor and error term.

### 4.1 When a change exists

In this section, we show the consistency of the change point estimator and derive its rate of convergence when a break point exists, i.e., $\tau_0 \in (0,1)$. After the FD transformation, the regressors and the error term will always be $I(0)$ processes. For comparison purposes, let us discuss the result in the pure time series case, i.e., $n = 1$.

**Theorem 7** *Under Assumptions 1-4, In the pure time series case, where $n = 1$, as $T \to \infty$, we have $\hat{\tau}_{FD} \xrightarrow{p} \tau_0$.*

Theorem 7 shows that $\hat{\tau}_{FD}$ is always consistent in the pure time series case. However, $\hat{k}_{FD}$ may not be consistent. As shown in Bai (1996), $\hat{k}_{OLS}$ is even consistent for the cointegration case in pure time series. After first differencing, only $\hat{\tau}_{FD}$ is still consistent but $\hat{k}_{FD}$ is not consistent anymore. In the panel data case, the FD-based estimator of $k$ can always guarantee consistency. Similar to case (a) in Theorem 2, we have the following Theorem.

**Theorem 8** *Under Assumptions 1-4, for $\tau_0 \in (0,1)$, we have $\hat{k}_{FD} - k_0 = O_p\left(\frac{1}{n}\right)$ as $(n,T) \to \infty$.*

Theorem 8 shows that the FD-based estimator $\hat{k}_{FD}$ is always $n$ consistent, no matter whether the regressor and error term are stationary or nonstationary. Since $\hat{k}_{FD}$ has a faster convergence speed than $\sqrt{n}$ consistency of the slope parameters, the asymptotics of $\tilde{\gamma}_{2,\widehat{k}_{FD}} = \tilde{\gamma}_2\left(\hat{k}_{FD}\right)$ and $\tilde{\delta}_{2,\widehat{k}_{FD}} = \tilde{\delta}_2\left(\hat{k}_{FD}\right)$ can be established as if the change point were known. The asymptotics results are given in the following theorem.

**Theorem 9** *Under Assumptions 1-4, as $(n,T) \to \infty$, we have*

$$\sqrt{nT}\begin{pmatrix} \tilde{\gamma}_{2,\widehat{k}_{FD}} - \gamma_2 \\ \tilde{\delta}_{2,\widehat{k}_{FD}} - \delta_2 \end{pmatrix} \xrightarrow{d} N\left(0, \frac{\omega}{\psi^2}\begin{pmatrix} 1 & 1-\tau_0 \\ 1-\tau_0 & 1-\tau_0 \end{pmatrix}\right),$$

*where $\psi = \left[1 + (\lambda-1)^2 \sum_{j=0}^{\infty} \lambda^{2j}\right]\sigma_\varepsilon^2$ and*

$$\omega = \left\{\left[1 + (\lambda-1)(\rho-1)\sum_{j=0}^{\infty}(\rho\lambda)^j\right]^2 + \left[\sum_{j=0}^{\infty}(\rho\lambda)^j\right]^2 \sum_{r=1}^{\infty}\left[(\rho-1)^2\rho^{2(r-1)} + (\lambda-1)^2\lambda^{2(r-1)}\right]\right\}\sigma_\varepsilon^2\sigma_e^2.$$

Theorem 9 shows that the convergence speed of $\tilde{\gamma}_{2\widehat{k}_{FD}}$ and $\tilde{\delta}_{2\widehat{k}_{FD}}$ is always $\sqrt{nT}$, whether the regressor and the error term are stationary or nonstationary.

**Theorem 10** *Under Assumptions 1-4, for $\tau_0 \in (0,1)$, as $(n,T) \to \infty$, we have*

$$\frac{(\beta_2 - \beta_1)^2 \psi^2}{\omega} n(\hat{k}_{FD} - k_0) \xrightarrow{d} \arg\max_r \left[ W(r) - \frac{|r|}{2} \right],$$

*where $\psi$ and $\omega$ are defined in Theorem 9.*

Different from $\hat{k}_{OLS}$ in Theorem 4, the asymptotic distribution of $\hat{k}_{FD}$ is robust to different values of $\rho$, $\lambda$, and $\tau_0$. As discussed in Yao (1987) and Bai (1997), the distribution function $\arg\max_r \left[ W(r) - \frac{|r|}{2} \right]$ is symmetric with critical values $\pm 7$ and $\pm 11$ for 10% and 5% significance levels, respectively. Confidence intervals for $k_0$ can be therefore constructed. $\delta_2 = \beta_2 - \beta_1$ can be estimated using $\tilde{\delta}_{2\widehat{k}_{FD}}$. By Theorem 9, we know $\tilde{\delta}_{2,\widehat{k}_{FD}} = \delta_2 + o_p(1)$. $\psi$ and $\omega$ can be estimated by $\hat{\psi} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \Delta x_{it}^2$ and $\hat{\omega} = \frac{1}{nT} \sum_{i=1}^n \sum_{s,t=1}^T \Delta x_{is} \Delta x_{it} \Delta \hat{u}_{is} \Delta \hat{u}_{it}$, where $\Delta \hat{u}_{it}$ is the residual from the FD regression. Similar to Lemma 16 in Baltagi et al. (2012), it can be shown that $\hat{\psi} = \psi + o_p(1)$ and $\hat{\omega} = \omega + o_p(1)$ and hence $\frac{\tilde{\delta}^2_{2,\widehat{k}_{FD}} \hat{\psi}^2}{\hat{\omega}} - \frac{\delta_2^2 \psi^2}{\omega} = o_p(1)$. Theorem 10 implies that $n\left(\hat{k} - k_0\right) = O_p(1)$. Hence

$$\left( \frac{\tilde{\delta}^2_{2,\widehat{k}_{FD}} \hat{\psi}^2}{\hat{\omega}} - \frac{\delta_2^2 \psi^2}{\omega} \right) n(\hat{k}_{FD} - k_0) = o_p(1).$$

It means that Theorem 10 yields the same results by replacing $\delta_2$, $\psi$ and $\omega$ with their consistent estimates. It is worth pointing out that we prove consistency and derive the asymptotic distributions of the FD estimator even when the regression is spurious, i.e., when the regressor is nonstationary with nonstationary error term. As one referee points out: This robust estimator of the unknown change point using FD could be useful in the context of testing for cointegration or no-cointegration in a panel data context. For example, after estimating the unknown change point using FD, one can split the data and compute the regression residuals from the sub-samples. Then one can use the mixture of residuals from the sub-samples to perform the usual residual based test for panel cointegration.

## 4.2 When a change does not exist

In the time series set-up, as shown in Theorem 5, the OLS-based estimator is inconsistent when $\rho = 1$. In fact, Nunes et al. (1995) and Bai (1998) show that there is a tendency to spuriously estimate a break point in the middle of the sample using the OLS-based estimator when the disturbances follow an $I(1)$ process, even though a break point does not actually exist. This spurious break problem can be solved by using the FD-based estimator.

**Theorem 11** *Under Assumptions 1-3, in the pure time series case, where $n = 1$, and as $T \to \infty$, for $\tau_0 = 0$ or $1$, we have $\hat{\tau}_{FD} \xrightarrow{p} \{0, 1\}$.*

Similarly, the FD-based estimator $\hat{\tau}_{FD}$ is also consistent in the panel data case.

**Theorem 12** *Under Assumptions 1-3, for* $\tau_0 = 0$ *or* $1$*, we have* $\hat{\tau}_{FD} \xrightarrow{p} \{0, 1\}$ *as* $(n, T) \to \infty$.

Theorem 12 shows that $\hat{\tau}_{FD}$ always converges to 0 or 1. After the FD transformation, the regressor and the error term will always be $I(0)$ processes. Different from the results in Theorem 6, the spurious break problem will not happen if we use the FD based estimator of $\tau$. Hence the FD-based estimator is robust whether a break point exists or not.

# 5 Finite Sample Performance

In this section, Monte Carlo simulations are conducted to study the finite sample properties of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$. We consider a simple model

$$y_{it} = 1 + x_{it} + \delta \cdot 1\{t > k_0\} + \delta x_{it} \cdot 1\{t > k_0\} + u_{it}, \qquad i = 1, \cdots, n; t = 1, \cdots, T,$$

where $x_{it}$ and $u_{it}$ follow an AR(1) process given in (3) and (4), respectively. $\lambda$ and $\rho$ are varied over the range $(0, 0.2, 0.5, 0.8, 1)$ and $\sigma_\varepsilon^2 = \sigma_e^2 = 5$. The sample size $T$ is fixed at 50, and $n$ is varied over the range $(1, 10, 50, 100)$. For each experiment, we perform $1,000$ replications. We consider two cases: $\delta = 0$ when there is no break point; and $\delta = 0.2$ when there is a break point in the sample at $k_0 = 15$ and 35. For each replication, the break point is estimated using OLS and FD. Due to limited space, we only present the results of cases for $(\lambda = 0, \rho = 0)$, $(\lambda = 0, \rho = 1)$ and $(\lambda = 1, \rho = 0)$ and $(\lambda = 1, \rho = 1)$ and provide the rest of these results in the appendix available in the working paper version of this paper. Basically when the values $\lambda$ and $\rho$ are small, the findings are similar to the value of 0. Similarly, when the values $\lambda$ and $\rho$ are large, the findings are similar to the value of 1. Overall, for all these $\lambda$ and $\rho$ combinations, we find the same conclusion reported in the paper: The FD estimator remains robust to all cases considered.

Figures 1-4 show the empirical distributions of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$. When there is no break point, the highest probability mass of $\hat{k}$ occurs at both tails for $(\lambda = 0, \rho = 0)$ and $(\lambda = 1, \rho = 0)$. The highest probability mass of $\hat{k}_{OLS}$ occurs at the right tail for $(\lambda = 0, \rho = 1)$. However, the mass of the distribution is more concentrated in the middle than in the tails for $(\lambda = 1, \rho = 1)$. It means that a spurious break still happens in panel data, even for a large $n$. When there is a break point at $k_0 = 15$ or 35, the estimator is not concentrated around the true break point when $n = 1$ except the case $(\lambda = 1, \rho = 0)$. As $n$ increases, the estimate of $k$ is improved with the increase in the number of cross-sectional observations $n$. This indicates that in the panel data set-up, if the cross-sectional dimension is large, the weak signal can be strengthened by the repeating regression across the cross-sectional dimension. This argument is similar in spirit to the argument of establishing consistency for the panel spurious regression, see for example Phillips and Moon (1999) and Kao (1999). It is worth pointing out that the empirical distribution is not symmetric for $k_0 = 15$ and 35 except for case (a) in Figure 1. Our simulations suggest that Theorem 4 makes predictions about finite sample behavior that are reasonable.

When there is no break point, the highest probability mass of $\hat{k}_{FD}$ occurs at both tails for all combinations of $\lambda$ and $\rho$. Different from $\hat{k}_{OLS}$, the spurious break problem does not appear for $(\lambda = 1, \rho = 1)$. When there is a break point at $k_0 = 15$ or 35, the estimator is concentrated around the true break point as $n$ increases. Moreover, the estimate of $k$ using FD converges to the true break faster than the one using OLS except for the cointegration case.

To investigate the effect of cross-sectional dependence, following Kim (2011), we consider the following model

$$y_{it} = 1 + x_{it} + \delta \cdot 1\{t > k_0\} + \delta x_{it} \cdot 1\{t > k_0\} + h_i' F_t + u_{it}, \qquad i = 1, \cdots, n; t = 1, \cdots, T,$$

where $F_t = 0.6F_{t-1} + \eta_t$ with $\eta_t \overset{iid}{\sim} N(0,1)$, and $h_i$ is 0.5 for $n = 1$, and drawn from $U(0,1)$ for the other values of $n$.

Figures 5-8 show the empirical distributions of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, that ignore the cross-sectional dependence, when in fact there are common factors. Comparing results with Figures 1-4, we see that the highest probability mass under cross-sectionally dependent factors is lower than those without cross-sectional dependent factors. In fact, in Figure 8 where $(\lambda = 1, \rho = 1)$, $\hat{k}_{FD}$ cannot find the true break point $k_0 = 15$ anymore. The asymptotic properties of the change point estimate depend upon the specification of the error process, the specification of the regressors, whether there is serial correlation and cross sectional dependence, among other things. In this paper, we only focused on robustness with respect to serial correlation. Deriving the asymptotic properties of the change point estimate by allowing for cross-sectional dependence is an interesting research question, however, we believe it is beyond the scope of this paper. From our limited Monte Carlo results, we know that the FD may do more harm than good if we allow for strong cross-sectional dependence. In fact, our limited experiments indicate that the FD estimator of the change point is not robust with respect to strong cross-sectional dependence generated by a factor structure. A thorough investigation for this problem is needed following the work of Kim (2011, 2014).

# 6 Conclusion

In this paper, we discuss the estimation and inference of the change point in a panel regression not knowing whether the regressor and error term are stationary or nonstationary. Also, the change point may be present or not present in the model. We consider the change point estimation using the OLS and FD estimators. Different from the results in the pure time series case, consistency of the change point estimator can be established. The distribution of the OLS-based estimator of the change point varies over different values of $\rho$ and $\lambda$. However, the FD-based estimator of the change point is robust to stationary or nonstationary regressors and error term, no matter whether a change point is present or not. Based on these results, we recommend the FD-based estimator of the change point.

## References


[1] Ahn, S.C., and Schmidt, P. (1995), "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5-28.

[2] Amsler, C., Schmidt, P., and Vogelsang, T. (2009), "The KPSS Test Using Fixed-b Critical Values: Size and Power in Highly Autocorrelated Time Series," *Journal of Time Series Econometrics*, 1, 1-42.

[3] Arnold, L. (1974), "*Stochastic Differential Equations: Theory and Applications*," New York: Wiley.

[4] Bai, J. (1996), "A Note on Spurious Break and Regime Shift in Cointegrating Relationship," Working paper 96-13, Department of Economics, MIT.

[5] Bai, J. (1997), "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics,* 79, 551-563.

[6] Bai, J. (1998), "A Note on Spurious Break," *Econometric Theory*, 14, 663-669.

[7] Bai, J. (2010), "Common Breaks in Means and Variances for Panel Data," *Journal of Econometrics*, 157, 78–92.

[8] Bai, J., Lumsdaine, R. L., and Stock, J. H. (1998), "Testing for and Dating Common Breaks in Multivariate Time Series," *The Review of Economic Studies*, 65, 3, 395-432.

[9] Baltagi, B. H., Kao, C., and Liu, L. (2008), "Asymptotic Properties of Estimators for the Linear Panel Regression Model with Random Individual Effects and Serially Correlated Errors: the Case of Stationary and Non-stationary Regressors and Residuals," *Econometrics Journal*, 11, 554–572.

[10] Baltagi, B. H., Kao, C., and Liu, L. (2012), "Supplementary Appendix to Estimation and Identification of Change Points in Panel Models with Nonstationary or Stationary Regressors and Error Term," Working paper, Syracuse University.

[11] Baltagi, B. H., Kao, C., and Na, S. (2011), "Test of Hypotheses in Panel Data Models when the Regressor and Disturbances are Possibly Nonstationary," *Advances in Statistical Analysis*, 95, 4, 329-350.

[12] Bhattacharya, P. K. (1987), "Maximum Likelihood Estimation of a Change-Point in the Distribution of Independent Random Variables, General Multiparameter Case," *Journal of Multivariate Analysis* 23, 183–208.

[13] Feng, Q., Kao, C., and Lazarová, S. (2009), "Estimation and Identification of Change Points in Panel Models," Syracuse University Working Paper.

[14] Hsu, C., and Lin, C. (2011), "Change-Point Estimation for Nonstationary Panel Data," Working Paper, Department of Economics, National Central University, Jhongli 32001, Taiwan.

[15] Hsu Y. C., and Kuan, C. M. (2008), "Change-point Estimation of Nonstationary I(d) Process," *Economics Letters*, 98, 115-121.

[16] Joseph, L., and Wolfson, D. B. (1992), "Estimation in Multi-path Change-point Problems," *Communications in Statistics-Theory and Methods*, 21, 4, 897-913.

[17] Kao, C. (1999), "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data," *Journal of Econometrics*, 90, 1-44.

[18] Kim, D. (2011), "Estimating a Common Deterministic Time Trend Break in Large Panels with Cross Sectional Dependence," *Journal of Econometrics*, 164, 310-330.

[19] Kim, D. (2014), "Common Breaks in Time Trends for Large Panel Data with a Factor Structure," *Econometrics Journal,* 17, 301–337.

[20] Murray, C. J., and Papel, D. H. (2000), "Testing for Unit Roots in Panels in the Presence of Structural Change with an Application to OECD Unemployment," *Advances in Econometrics,* 15, 223-238.

[21] Nunes, L. C., Kuan, C. M., and Newbold, P. (1995), "Spurious Breaks," *Econometrics Theory*, 11, 736-749.

[22] Perron, P., and Zhu, X. (2005), "Structural Breaks with Deterministic and Stochastic Trends," *Journal of Econometrics,* 129, 65–119.

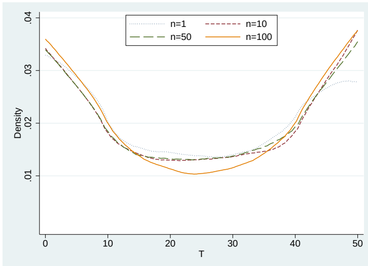
[23] Phillips, P. C. B., and Moon, H. (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057-1111.

[24] Picard, D. (1985), "Testing and Estimating Change-Points in Time Series," *Advances in Applied Probability*, 17, 841-867.

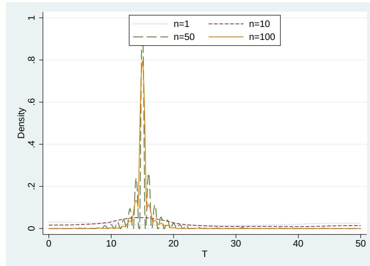
[25] Yao, Y. (1987), "Approximating the Distribution of the ML Estimation of the Change-Point in a Sequence of Independent R.V.'S," *Annals of Statistics*, 3, 1321-1328.

[26] Zeileis, A., Kleiber, C., Krämer, W., and Hornik, K. (2003), "Testing and Dating of Structural Changes in Practice," *Computational Statistics & Data Analysis*, 44, 109-123.

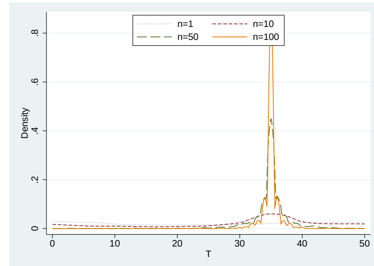
Figure 1: Empirical Distribution of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, no common component, $\lambda = 0$, $\rho = 0$



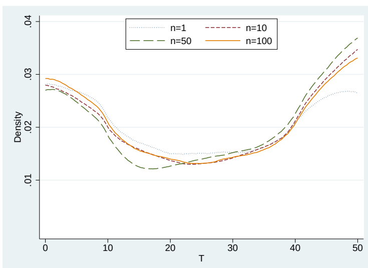

(a) OLS, No Break



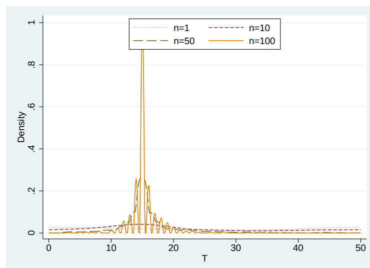

(b) OLS, Break at $k_0 = 15$



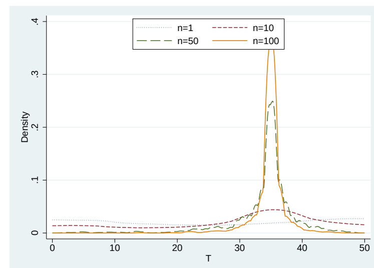

(c) OLS, Break at $k_0 = 35$




(d) FD, No Break

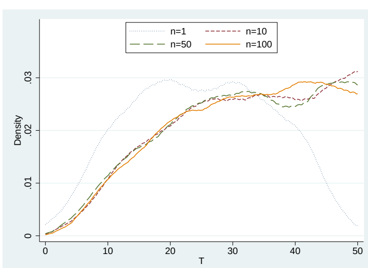



(e) FD, Break at $k_0 = 15$

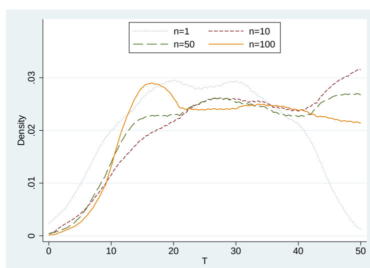



(f) FD, Break at $k_0 = 35$

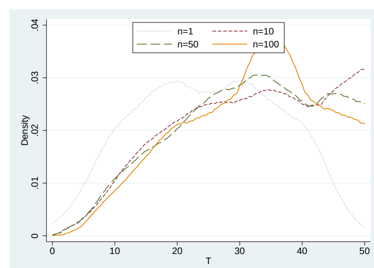
Figure 2: Empirical Distribution of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, no common component, $\lambda = 0$, $\rho = 1$



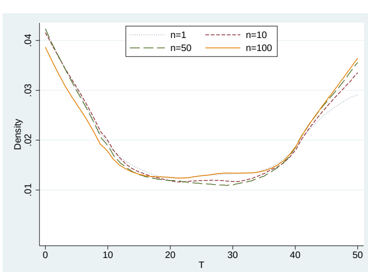

(a) OLS, No Break



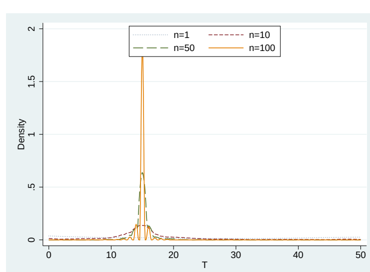

(b) OLS, Break at $k_0 = 15$



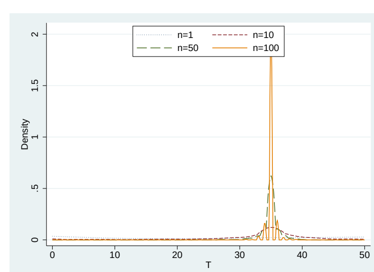

(c) OLS, Break at $k_0 = 35$




(d) FD, No Break

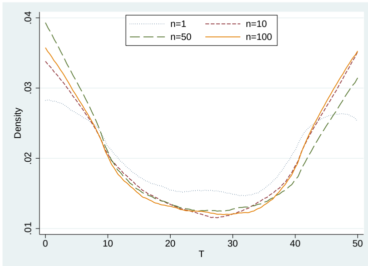



(e) FD, Break at $k_0 = 15$

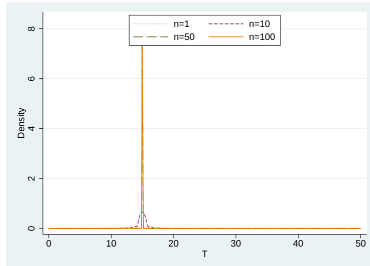



(f) FD, Break at $k_0 = 35$

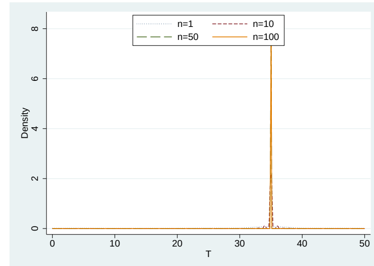
Figure 3: Empirical Distribution of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, no common component, $\lambda = 1$, $\rho = 0$



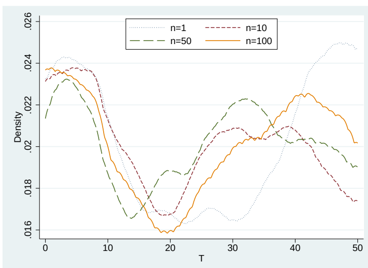

(a) OLS, No Break



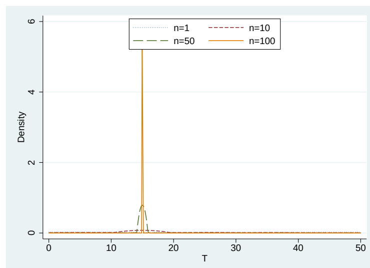

(b) OLS, Break at $k_0 = 15$



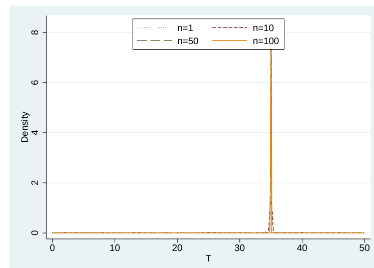

(c) OLS, Break at $k_0 = 35$




(d) FD, No Break

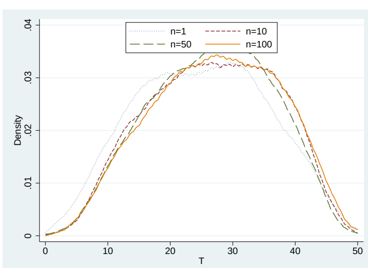



(e) FD, Break at $k_0 = 15$

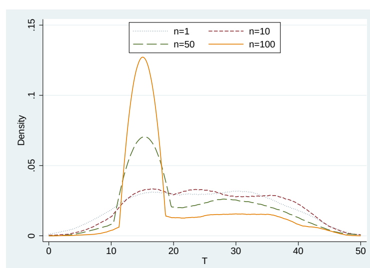



(f) FD, Break at $k_0 = 35$

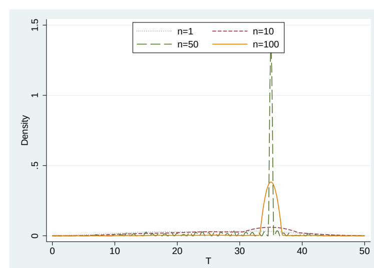
Figure 4: Empirical Distribution of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, no common component, $\lambda = 1$, $\rho = 1$



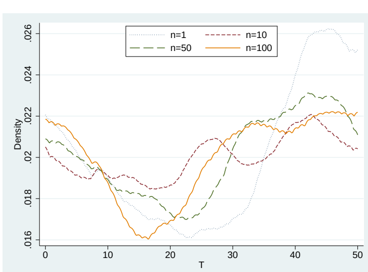

(a) OLS, No Break



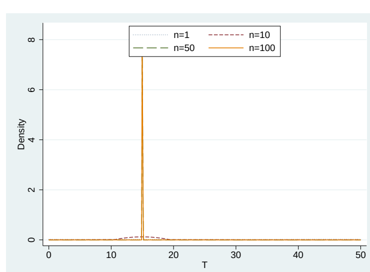

(b) OLS, Break at $k_0 = 15$



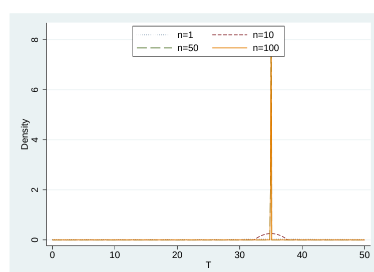

(c) OLS, Break at $k_0 = 35$




(d) FD, No Break

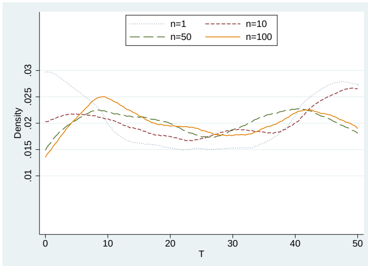



(e) FD, Break at $k_0 = 15$

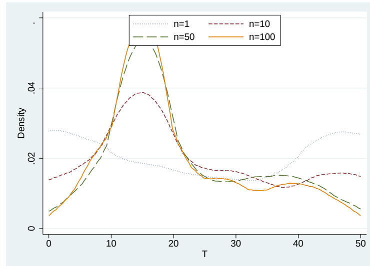



(f) FD, Break at $k_0 = 35$

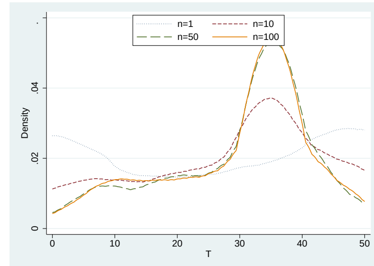
Figure 5: Empirical Distribution of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, with common component, $\lambda = 0$, $\rho = 0$



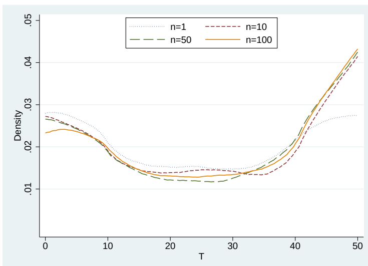

(a) OLS, No Break



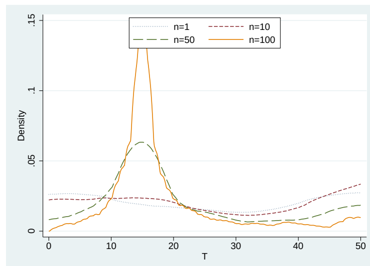

(b) OLS, Break at $k_0 = 15$



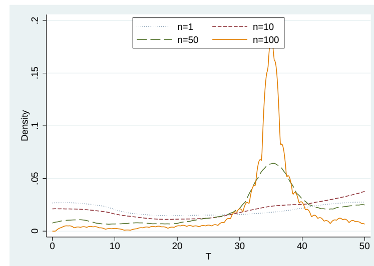

(c) OLS, Break at $k_0 = 35$




(d) FD, No Break

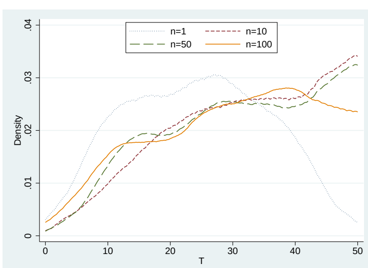



(e) FD, Break at $k_0 = 15$

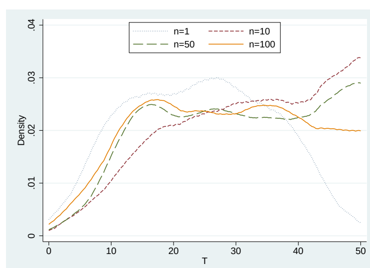



(f) FD, Break at $k_0 = 35$

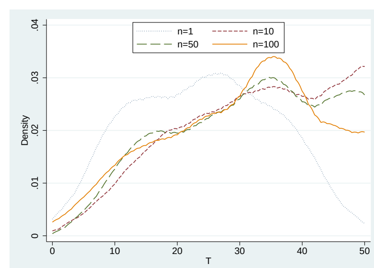
Figure 6: Empirical Distribution of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, with common component, $\lambda = 0$, $\rho = 1$



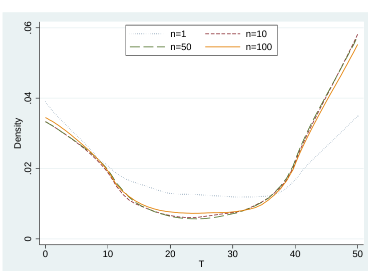

(a) OLS, No Break



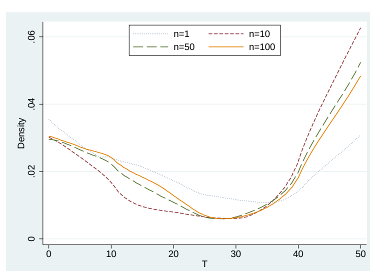

(b) OLS, Break at $k_0 = 15$



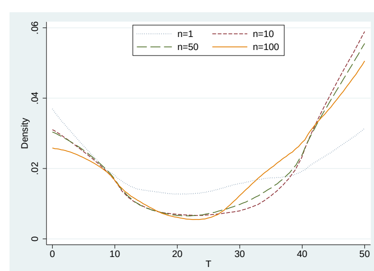

(c) OLS, Break at $k_0 = 35$




(d) FD, No Break



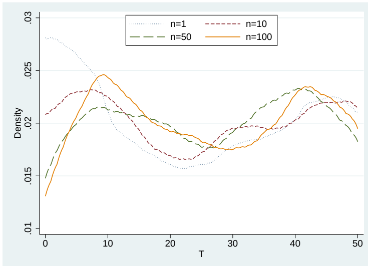

(e) FD, Break at $k_0 = 15$



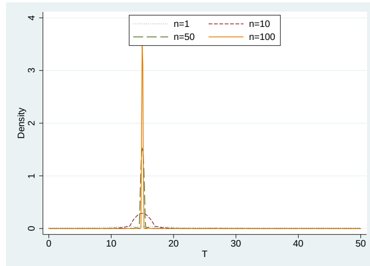

(f) FD, Break at $k_0 = 35$

90

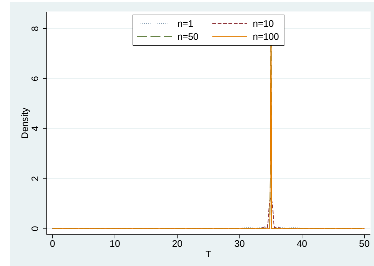
Figure 7: Empirical Distribution of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, with common component, $\lambda = 1$, $\rho = 0$



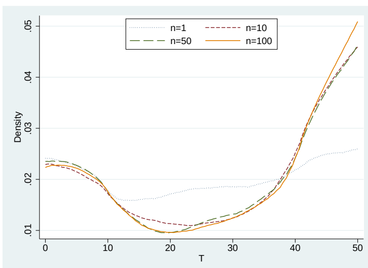

(a) OLS, No Break



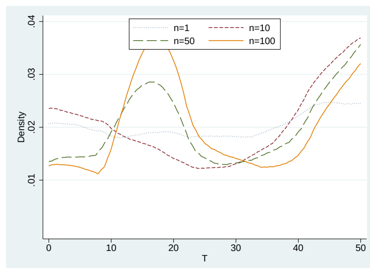

(b) OLS, Break at $k_0 = 15$



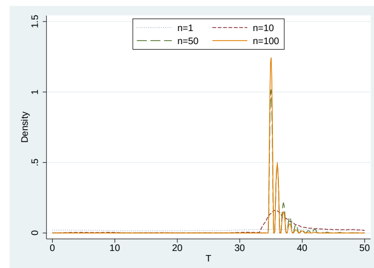

(c) OLS, Break at $k_0 = 35$




(d) FD, No Break

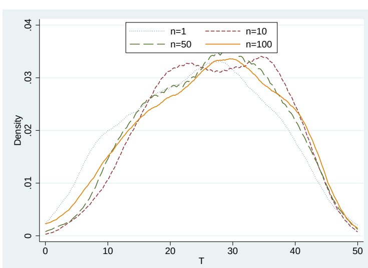



(e) FD, Break at $k_0 = 15$

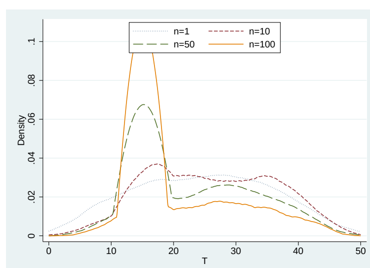



(f) FD, Break at $k_0 = 35$

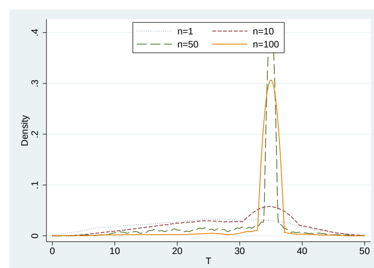
Figure 8: Empirical Distribution of $\hat{k}_{OLS}$ and $\hat{k}_{FD}$, with common component, $\lambda = 1$, $\rho = 1$



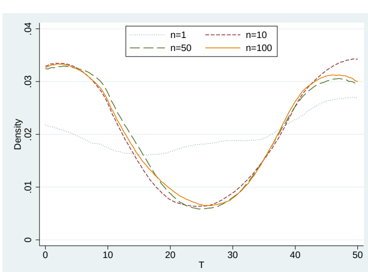

(a) OLS, No Break



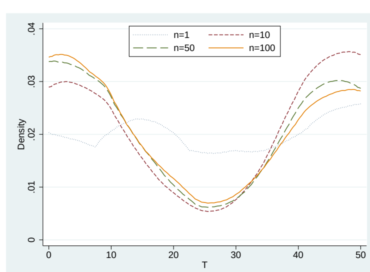

(b) OLS, Break at $k_0 = 15$



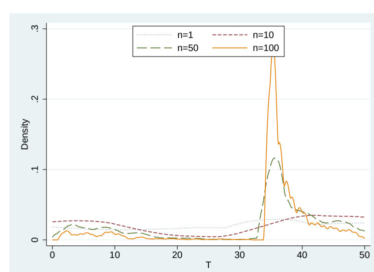

(c) OLS, Break at $k_0 = 35$




(d) FD, No Break




(e) FD, Break at $k_0 = 15$




(f) FD, Break at $k_0 = 35$