1 Introduction

This paper extends Pesaran (2006) and Baltagi, Feng and Kao (2016) (BFK hereafter) by allowing for endogenous regressors in large heterogeneous panels with unknown common structural changes in slopes and error factor structure. Since endogenous regressors and structural breaks are often encountered in empirical studies with large panels, this extension makes the Pesaran's (2006) CCE approach empirically more appealing.

Structural changes in time series regression models with endogenous regressors have been recently studied by Boldea, Hall and Han (2012), Hall, Han and Boldea (2012), Perron and Yamamoto (2014, 2015) and Chen (2015). An important finding of Perron and Yamamoto (2015) is that the ordinary least squares (OLS) estimates of break fractions are still consistent even in the presence of endogenous regressors. The intuition is that changes in the slope parameters imply changes in the probability limits of the OLS estimates. We extend this intuition to heterogeneous panels with an error factor structure when estimating common break points. In turn, the consistency of the OLS estimator of the break dates, established in BFK, can be extended to the model with endogenous regressors.

We show that the CCE approach is still valid when dealing with cross-sectional dependence due to unobservable factors even in the presence of endogeneity and structural changes in slopes and error factor loadings. As in Pesaran (2006), after the unobservable factors are controlled for by cross-sectional averages of observable variables, common break points can be consistently estimated using least squares as proposed by Bai (1997, 2010) even in the presence of endogeneity. Conditional on the estimated break points, slope parameters can be consistently estimated by instrumental variable (IV) estimation using augmented data in each regime defined by the estimated break point.

We also show that our break date estimator is robust to potential structural changes in the factor loadings, a phenomenon recently considered by Stock and Watson (2009), Breitung and Eickmeier (2011), Chen, Dolado and Gonzalo (2014), Yamamoto and Tanaka (2015), Cheng, Liao and Schorfheide (2016), Bai, Han and Shi (2017) and Ma and Su (2018), to name a few. Since the CCE approach used in this paper wipes out factors instead of estimating them directly, structural changes in factor loadings do not affect the consistency of our estimators. In this sense, our methodology differs from other papers on structural changes in panels with interactive fixed effects, see for example Li, Qian and Su (2016).

The paper is organized as follows. Section 2 introduces a heterogeneous panel data model allowing for endogenous regressors and structural changes. To estimate parameters of interests, Section 3 starts with a simple case, followed by a formal discussion of the general model with common correlated effects. Section 4 provides concluding remarks. The online Appendix contains all the proofs and the technical materials. Monte Carlo simulations are also included to shed some light on the performance of the common break point estimators.

2 Model

Consider a heterogeneous panel data model with a multifactor error structure, see Pesaran (2006):

$$y_{it} = x_{it}'\beta_i + e_{it}, \tag{1}$$

$$e_{it} = \gamma_i' f_t + \varepsilon_{it}, \tag{2}$$

$i = 1, ..., N$; $t = 1, ..., T$. x_{it} is a $p \times 1$ vector of explanatory variables, and the error term e_{it} is cross-sectionally correlated, modelled by a multifactor structure, where f_t is an $m \times 1$ vector of unobserved factors and γ_i is the corresponding loading vector. ε_{it} is the idiosyncratic error independent of x_{it}. However, x_{it} could be affected by the unobservable common effects f_t,

$$x_{it} = \Gamma_i' f_t + v_{it}, \tag{3}$$

$i = 1, ..., N$; $t = 1, ..., T$, where Γ_i is an $m \times p$ factor loading matrix. v_{it} is a $p \times 1$ vector of disturbances. Given the correlation between x_{it} and e_{it} due to the unobservable factors f_t, OLS for each individual series could be inconsistent. Pesaran (2006) develops the CCE estimator of β_i by using cross-section averages as observable proxies for the f_t.

Harding and Lamarche (2011) extend this model by allowing for endogenous regressors and correlation between x_{it} and the factor loadings γ_i in errors in a homogeneous panel data model. They find that the Pesaran's CCE approach can be easily modified to accommodate these situations. Recently, Forchini, Jiang and Peng (2015) also study the case of endogenous regressors in Pesaran's (2006) model of heterogeneous panels. Instead of using IV estimation, they control for endogeneity by considering reduced form equations, in which the reduced form parameters can be consistently estimated by the CCE approach. Then, the structural parameters $\beta_i' s$ (and their mean β) can be inferred from the estimated reduced form parameters. In addition, Neal (2015) extends the CCE approach of Pesaran (2006) and Chudik and Pesaran (2015) in the dynamic heterogeneous panels to the case of endogenous regressors using lags as instruments.

BFK extend Pesaran (2006) by allowing for structural changes in some or all components of β_i, which may be due to macro policy shocks or technological progress. Assume a structural break at a common unknown date k_0:

$$y_{it} = x_{it}' \beta_i(k_0) + e_{it}, \tag{4}$$

$i = 1, ..., N; t = 1, ..., T$, where $\beta_i(k_0)$ are different before and after the date k_0, i.e.,

$$\beta_i(k_0) = \begin{cases} \beta_{1i}, & t = 1, ..., k_0, \\ \beta_{2i} \neq \beta_{1i}, & t = k_0 + 1, ..., T. \end{cases}$$

In this model, there are two regimes in the time dimension with a break at k_0. BFK show that the common break date k_0 can be consistently estimated as in Bai (2010) and Kim (2011) in a panel mean-shift model and a panel deterministic time trend model, respectively. Also, the slopes $\beta_i' s$ and their cross-section mean can be consistently estimated by the CCE approach in each regime.

In this paper, we consider both endogenous regressors and structural changes in Pesaran's model (1). Specifically, ε_{it} is allowed to be correlated with v_{it} (thus x_{it}). In addition, we also allow for a common break in the error factor structure, i.e.,

$$\gamma_i(k_1) = \begin{cases} \gamma_{1i}, & t = 1, ..., k_1, \\ \gamma_{2i} \neq \gamma_{1i}, & t = k_1 + 1, ..., T, \end{cases}$$

where the common break k_1 can be the same as or different from k_0. It is important to consider the instability of the error factor structure when the structural changes in slopes are present. As indicated in the Monte Carlo experiments in the online Appendix, a break in the error factor loadings could lead to a spurious break in slope parameters. Such error factor structure instability could affect the estimation procedure proposed in similar panel change point models like Li, Qian, Su (2016) in which Bai's (2009) interactive fixed effects approach is used to deal with cross-sectional dependence in the errors.

Thus, the model considered in this paper is

$$\begin{aligned} y_{it} &= x_{it}'\beta_i(k_0) + e_{it} = \begin{cases} x_{it}'\beta_{1i} + e_{it}, & t = 1, ..., k_0, \\ x_{it}'\beta_{2i} + e_{it}, & t = k_0 + 1, ..., T, \end{cases} \\ e_{it} &= \gamma_i(k_1)' f_t + \varepsilon_{it},\ x_{it} = \Gamma_i' f_t + v_{it}, \end{aligned} \tag{5}$$

where e_{it} and x_{it} are defined in (2) and (3), and $Cov(\varepsilon_{it}, v_{it}) \neq 0$. Assume there are q instruments z_{it} with $q \geq p$. z_{it} could be affected by f_t.

The model (5) departs from BFK's by allowing ε_{it} to be correlated with v_{it} and break in factor loadings. There are two sources of endogeneity in x_{it} in this model: one is due to common factors f_t, and the other one is due to $Cov(\varepsilon_{it}, v_{it}) \neq 0$. In this way, this model accommodates 4 important empirical features: slope heterogeneity, cross-sectional dependence, structural breaks and endogeneity. Such rich empirical flexibility makes this model more appealing in applied studies. In this model, the parameters of interest are cross-sectional averages of the slopes β_{1i} and β_{2i}, and the common break k_0.

3 Estimation Results

3.1 A simplified case

In this section, to facilitate the discussion, we start with a simple case where there are no unobserved common effects f_t in the errors. For $i = 1, ..., N$,

$$y_{it} = x_{it}'\beta_i(k_0) + \varepsilon_{it} = \begin{cases} x_{it}'\beta_{1i} + \varepsilon_{it}, & t = 1, ..., k_0, \\ x_{it}'\beta_{2i} + \varepsilon_{it}, & t = k_0 + 1, ..., T. \end{cases} \tag{6}$$

x_{it} is correlated with ε_{it}, and z_{it} is the instrumental variable for x_{it}. In this model, we first estimate k_0 and then β_{1i} and β_{2i} in each regime, respectively.

When $x_{it} = 1$, (6) reduces to Bai's (2010) panel mean-shift model. When $N = 1$, (6) is the time series model considered in Boldea, Hall and Han (2012), Hall, Han and Boldea (2012), Perron and Yamamoto (2014, 2015), with one change point. With endogenous regressors, Hall, Han and Boldea (2012) use IV estimation and show that break fractions and slopes can be consistently estimated in a time series setup. However, Perron and Yamamoto (2015) find that the OLS estimator of break fractions is still consistent even in the presence of endogeneity, and that OLS is better than IV in terms of efficiency, and avoids potential weak identification problems due to weak instruments. Conditioning on the OLS estimate of change points, slope parameters can be consistently estimated by IV regression in each regime.

In our panel data setup, following Perron and Yamamoto (2015), we use OLS to estimate k_0, and use IV to estimate the slope parameters. Let $b_i = (\beta_{1i}', \beta_{2i}')', i = 1, ..., N$. For every i, and $k = 1, ..., T-1$, define $X_{1i}(k) = (x_{i1}, \cdots, x_{i,k})'$, $X_{2i}(k) = (x_{i,k+1}, \cdots, x_{iT})'$. Similarly, define $Y_{1i}(k) = (y_{i1}, \cdots, y_{i,k})'$, $Y_{2i}(k) = (y_{i,k+1}, \cdots, y_{iT})'$. Let $Y_i = (y_{i1}, \cdots, y_{iT})'$ and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$ denote the stacked data and errors over time, thus $Y_i = (Y_{1i}(k)', Y_{2i}(k)')'$. Using the notation $\mathbb{X}_i(k) = \begin{pmatrix} X_{1i}(k) & 0 \\ 0 & X_{2i}(k) \end{pmatrix}$, equation (6) can be written in matrix form as

$$Y_i = \mathbb{X}_i(k_0) b_i + \varepsilon_i, \; i = 1, ..., N. \tag{7}$$

Given any $k = 1, ..., T-1$, the least squares estimator of b_i is

$$\hat{b}_i(k) = \begin{pmatrix} \hat{\beta}_{1i}(k) \\ \hat{\beta}_{2i}(k) \end{pmatrix} = [\mathbb{X}_i(k)'\mathbb{X}_i(k)]^{-1}\mathbb{X}_i(k)'Y_i, \; i = 1, ..., N. \tag{8}$$

The corresponding sum of squared residuals is given by

$$SSR_i\,(k) = [Y_i - \mathbb{X}_i(k)\hat{b}_i(k)]'[Y_i - \mathbb{X}_i(k)\hat{b}_i(k)], \; i = 1, ..., N.$$

As in Bai (2010), Kim (2011) and BFK, the least squares estimator of k_0 is defined as

$$\hat{k} = \arg\min_{1 \le k \le T-1} \sum_{i=1}^{N} SSR_i(k). \tag{9}$$

Different from BFK, here ε_{it} is allowed to be correlated with x_{it}. Following Perron and Yamamoto (2015), we can project ε_i on the column space spanned by $\mathbb{X}_i(k_0)$ such that the new error term ε_i^* (defined below) is uncorrelated with $\mathbb{X}_i(k_0)$. Rewrite equation (7) above as:

$$Y_i = \mathbb{X}_i(k_0)\beta_{iT}^*(k_0) + \varepsilon_i^*, \tag{10}$$

where $\varepsilon_i^* = (I - P_{\mathbb{X}})\varepsilon_i = (\varepsilon_{i1}^*, ..., \varepsilon_{iT}^*)'$ and $P_{\mathbb{X}}$ is the projection matrix based on $\mathbb{X}_i(k_0)$, and

$$\beta_{iT}^*(k_0) = \begin{pmatrix} \beta_{1i}^* \\ \beta_{2i}^* \end{pmatrix} = b_i + [\mathbb{X}_i(k_0)'\mathbb{X}_i(k_0)]^{-1} \mathbb{X}_i(k_0)'\varepsilon_i.$$

As argued by Perron and Yamamoto (2015) in a time series model, a structural change in the original parameter $\beta_i(k_0)$ implies a shift in the new parameter $\beta_i^*(k_0)$,

$$\beta_i^*(k_0) = \mathrm{p}\lim_{T\to\infty} \beta_{iT}^*(k_0) = b_i + \mathrm{p}\lim_{T\to\infty} [\mathbb{X}_i(k_0)'\mathbb{X}_i(k_0)]^{-1} \mathbb{X}_i(k_0)'\varepsilon_i,$$

at the same break date k_0, except for a knife-edge case.[1] Since the new errors ε_{it}^* are uncorrelated with x_{it}, (10) becomes BFK's Model 1. Following the same lines of proof as in BFK's Theorem 1, it can be shown that $\hat{k}$ is consistent for k_0, i.e., $\hat{k} - k_0 = o_p(1)$, under appropriate assumptions.

Given the estimated break date $\hat{k}$, regimes 1 and 2 are defined. In each regime, β_{1i} and β_{2i} can be consistently estimated by IV regression using instruments z_{it}, as suggested by Hall, Han and Boldea (2012), Perron and Yamamoto (2015) for a time series setup.

3.2 The general case

Next, we consider the case with common correlated effects (5) in the errors: for $i = 1, ..., N$,

$$y_{it} = x_{it}'\beta_i(k_0) + e_{it} = \begin{cases} x_{it}'\beta_{1i} + e_{it}, & t = 1, ..., k_0, \\ x_{it}'\beta_{2i} + e_{it}, & t = k_0 + 1, ..., T. \end{cases}$$

where $e_{it} = \gamma_i(k_1)' f_t + \varepsilon_{it}$. Besides nonzero $Cov(v_{it}, \varepsilon_{it})$, this model has an additional source of endogeneity due to the unobservable common factors f_t that affect both $x_{it} = \Gamma_i' f_t + v_{it}$ and e_{it}.

Even with endogenous regressors x_{it}, this general model with a multifactor error structure could still be fit into the simplified case discussed in the previous subsection. Hence, we could still use OLS to estimate k_0. However, due to the common f_t, errors e_{it} are no longer cross-sectionally independent. This is a major concern in the general case considered here. As pointed out by Kim (2011), the cross-sectional correlation in the errors could offset the information across the cross-sectional dimension under the common break assumption. Thus, $\hat{k} - k_0 = o_p(1)$ is not necessarily achieved without controlling for f_t. It depends on the magnitude of the cross-sectional correlation governed by the unobservable loadings. This finding is also confirmed in BFK's Figure 7 in the case of exogenous regressors x_{it}.

As in BFK, to control for the cross-sectional dependence due to f_t, we partial out f_t first. We show that the CCE approach is still valid in the presence of endogeneity (and a common break point). Let $F = (f_1', f_2', \cdots, f_T')'$ and $M_f = I_T - F(F'F)^{-1}F'$. (5) can be

[1] This finding also explains why the consistency of the pooled OLS estimator of change point is still achieved in a panel data fixed effects model considered in Feng, Kao and Lazarova (2009), Baltagi, Kao and Liu (2017), Boldea, Gan and Drepper (2016). In these models endogeneity arises when the potential correlation between regressors and fixed effects is ignored in the pooled OLS.

written in matrix form as

$$Y_i = \mathbb{X}_i(k_0)b_i + F\gamma_i(k_1) + \varepsilon_i, i = 1, ..., N. \tag{11}$$

Since f_t are unobservable, we follow Pesaran's (2006) idea of using the cross-sectional averages of y_{it} and x_{it} as proxies for f_t. Combining (3) and (5) yields

$$\underset{(p+1)\times 1}{w_{it}} = \begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \underset{(p+1)\times m}{C_i(k_0, k_1)'} \underset{m\times 1}{f_t} + \underset{(p+1)\times 1}{u_{it}(k_0)}, \tag{12}$$

where

$$\underset{m\times(p+1)}{C_i(k_0, k_1)} = (\gamma_i(k_1), \Gamma_i) \begin{pmatrix} 1 & 0 \\ \beta_i(k_0) & I_p \end{pmatrix} \text{ and } u_{it}(k_0) = \begin{pmatrix} \varepsilon_{it} + v_{it}'\beta_i(k_0) \\ v_{it} \end{pmatrix}.$$

In the case that the instruments z_{it} are affected by f_t, z_{it} can be included in the vector w_{it}. Note that like $\beta_i(k_0)$, the slope $C_i(k_0, k_1)$ in (12) also shifts at k_0, and k_1. Without loss of generality, we assume $k_1 > k_0$.[2] Thus,

$$C_i(k_0, k_1) = \begin{cases} C_{1i} = (\gamma_{1i} + \Gamma_i\beta_{1i},\ \Gamma_i), & t = 1, ..., k_0, \\ C_{2i} = (\gamma_{1i} + \Gamma_i\beta_{2i},\ \Gamma_i), & t = k_0 + 1, ..., k_1, \\ C_{3i} = (\gamma_{2i} + \Gamma_i\beta_{2i},\ \Gamma_i), & t = k_1 + 1, ..., T. \end{cases} \tag{13}$$

Let $\bar{w}_t = \sum_{i=1}^N \theta_i w_{it}$ be the cross-sectional average of w_{it} using weights θ_i, $i = 1, ..., N$. In particular,

$$\bar{w}_t = \bar{C}(k_0, k_1)' f_t + \bar{u}_t(k_0) \tag{14}$$

where

$$\bar{C}(k_0, k_1) = \sum_{i=1}^N \theta_i C_i(k_0, k_1) = \begin{cases} \bar{C}_1 = \sum_{i=1}^N \theta_i C_{1i}, & t = 1, ..., k_0, \\ \bar{C}_2 = \sum_{i=1}^N \theta_i C_{2i}, & t = k_0 + 1, ..., k_1, \\ \bar{C}_3 = \sum_{i=1}^N \theta_i C_{3i}, & t = k_1 + 1, ..., T, \end{cases} \tag{15}$$

and

$$\bar{u}_t(k_0) = \sum_{i=1}^N \theta_i u_{it}(k_0) = \begin{cases} \begin{pmatrix} \bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}'\beta_{1i} \\ \bar{v}_t \end{pmatrix}, & t = 1, ..., k_0, \\ \begin{pmatrix} \bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}'\beta_{2i} \\ \bar{v}_t \end{pmatrix}, & t = k_0 + 1, ..., T, \end{cases} \tag{16}$$

where $\bar{\varepsilon}_t = \sum_{i=1}^N \theta_i \varepsilon_{it}$, $\bar{v}_t = \sum_{i=1}^N \theta_i v_{it}$.

[2] Theorem 1 below still holds when $k_1 \leq k_0$. As shown in Figure A3 in the online Appendix, a different break point in factor loadings could lead to a spurious break in slopes if we ignore the unobserved factors in errors.

For equation (14), when $\bar{C}(k_0, k_1)$ is of full rank, f_t can be written as

$$f_t = \left[\bar{C}(k_0, k_1)\bar{C}(k_0, k_1)'\right]^{-1} \bar{C}(k_0, k_1)(\bar{w}_t - \bar{u}_t(k_0)). \tag{17}$$

For simplicity, we assume that the rank condition is satisfied. If the rank condition is not satisfied, additional variables like z_{it} that are also affected by f_t can be included in w_{it} in (12) to proxy f_t, as in Chudik and Pesaran (2015).[3]

As shown in Lemma 1 of Pesaran (2006), in (16), the cross-sectional averages of the errors $\bar{\varepsilon}_t$, $\bar{v}_t$, $\sum_{i=1}^{N} \theta_i v_{it}' \beta_{1i}$ and $\sum_{i=1}^{N} \theta_i v_{it}' \beta_{2i}$ all vanish as $N \to \infty$, thus

$$\bar{u}_t(k_0) \xrightarrow{p} 0$$

in both regimes as $N \to \infty$, regardless of the correlation between ε_{it} and v_{it}, yielding

$$f_t - \left[\bar{C}(k_0, k_1)\bar{C}(k_0, k_1)'\right]^{-1} \bar{C}(k_0, k_1)\bar{w}_t \xrightarrow{p} 0. \tag{18}$$

This suggests that it is asymptotically valid to use $\bar{w}_t$ as observable proxies for f_t. This finding also shows that the CCE approach proposed by Pesaran (2006) is robust to endogeneity and structural changes in slopes and factor structures.[4]

Let $\bar{W} = (\bar{w}_1', \bar{w}_2', \cdots, \bar{w}_T')'$ denote the $T \times (p+1)$ matrix of cross-sectional averages. Denote the $T \times T$ matrix M_w by $M_w = I_T - \bar{W}(\bar{W}'\bar{W})^{-1}\bar{W}'$. Premultiplying (11) by M_w, we obtain

$$M_w Y_i = M_w \mathbb{X}_i(k_0) b_i + M_w F \gamma_i(k_1) + M_w \varepsilon_i, i = 1, ..., N. \tag{19}$$

By defining $\tilde{Y}_i = M_w Y_i$, $\tilde{\mathbb{X}}_i(k_0) = M_w \mathbb{X}_i(k_0)$ and $\tilde{\varepsilon}_i = M_w \varepsilon_i$, (19) becomes

$$\tilde{Y}_i = \tilde{\mathbb{X}}_i(k_0) b_i + \tilde{\varepsilon}_i^0, i = 1, ..., N, \tag{20}$$

where $\tilde{\varepsilon}_i^0 = M_w F \gamma_i(k_1) + \tilde{\varepsilon}_i$.

[3] From equation (15), under random coefficient assumptions, matrix $\bar{C}(k_0, k_1)$ is not of full rank asymptotically when the means of Γ_i (and /or γ_{1i}, γ_{2i}, β_{1i}, β_{2i}) are zero. Compared with the case of no break considered in Pesaran (2006), the rank condition here needs to hold in each of the 3 regimes split by k_0 and k_1. Thus, it is more restrictive than Pesaran (2006). In addition, when the number of unobserved factors exceeds the number of regressors and dependent variable, i.e., $m > p + 1$, the rank condition is not satisfied.

[4] As shown by Breitung and Eickmeier (2011), ignoring breaks in the factors leads to doubling the number of factors. Hence, as long as the rank condition holds with increased number of factors, the CCE approach is robust to breaks in the factor structure. In the case where the effective number of factors is greater than $p + 1$, the loadings in the errors need to be independent of the loadings in the regressors. We would like to thank one of our referees for pointing this out.

Like Lemma 6 in BFK in the case of exogeneity, the online Appendix shows that each element of $M_w F\gamma_i(k_1)$ is of order $O_p(\frac{1}{\sqrt{N}})$ and vanishes as $(N,T) \to \infty$, implying that $\tilde{\varepsilon}_i^0$ can be treated as $\tilde{\varepsilon}_i$ asymptotically in the case of endogeneity. This implies that the partition regression (19) wipes out the unobserved factors F asymptotically regardless of structural changes in factor loadings $\gamma_i(k_1)$ and that breaks in γ_i can be ignored.[5] The Monte Carlo experiments in the online Appendix also confirm that a common break in loadings $\gamma_i(k_1)$ does not affect the consistency of the break date estimator and the slope parameters estimator asymptotically. Therefore, for simplicity we use the notation γ_i and $\bar{C}(k_0)$ instead of $\gamma_i(k_1)$ and $\bar{C}(k_0, k_1)$ in the rest of paper, where $\bar{C}_2$ includes two values over the span of $k_0, ..., T$.

Hence, the general case (11) considered here can be treated as the simple case (7) using the transformed data $\{\tilde{Y}_i, \tilde{\mathbb{X}}_i(k_0), i = 1, ..., N\}$. Similarly, for any possible change point $k = 1, ..., T-1$, the least squares estimates of b_i are

$$\tilde{b}_i(k) = [\tilde{\mathbb{X}}_i(k)'\tilde{\mathbb{X}}_i(k)]^{-1}\tilde{\mathbb{X}}_i(k)'\tilde{Y}_i$$

for $i = 1, ..., N$, and the corresponding sum of squared residuals is

$$\widetilde{SSR}_i\,(k) = [\tilde{Y}_i - \tilde{\mathbb{X}}_i(k)\tilde{b}_i(k)]'[\tilde{Y}_i - \tilde{\mathbb{X}}_i(k)\tilde{b}_i(k)].$$

The estimator of k_0 is defined similarly as

$$\tilde{k} = \arg \min_{1 \le k \le T-1} \textstyle\sum_i \widetilde{SSR}_i(k). \tag{21}$$

As in BFK, it can be shown that the estimator $\tilde{k}$ is still consistent in the general case with endogenous regressors and structural break in factor loadings, i.e., $\tilde{k} - k_0 = o_p(1)$.

Theorem 1 *Under Assumptions 1-11 in the online Appendix,* $\lim_{(N,T)\to\infty} P(\tilde{k} = k_0) = 1$.

In order to save space, all assumptions and proofs are provided in the online Appendix. Given the consistency of $\tilde{k}$, we can estimate the slope parameters in (11). Define $\mathbb{C} = \left(\left[\bar{C}(k_0)\bar{C}(k_0)'\right]^{-1}\bar{C}(k_0)\right)'$ in (18). Equation (11) can be written as

$$Y_i = \mathbb{X}_i(k_0)b_i + F\gamma_i + \varepsilon_i = \mathbb{X}_i(\tilde{k})b_i + \bar{W}\mathbb{C}\gamma_i + \varepsilon_i^0, \tag{22}$$

[5] It is important to note that the break k_1 in factor loadings cannot be ignored even when the factors are observed and treated as regressors when $k_1 \neq k_0$.

where $\varepsilon_i^0 = \varepsilon_i + [\mathbb{X}_i(k_0) - \mathbb{X}_i(\tilde{k})]b_i + (F - \bar{W}\mathbb{C})\gamma_i$.

Different from Pesaran (2006) and BFK, the CCE estimators of b_i are inconsistent due to the correlation between $\mathbb{X}_i(k_0)$ and ε_i (or between $\mathbb{X}_i(\tilde{k})$ and ε_i^0). This is also the case in equation (20) using transformed data. As in Hall, Han and Boldea (2012), Perron and Yamamoto (2015) in the time series setup, we obtain IV estimators for individual slope parameters and their cross-sectional means.

Similar to the case of exogenous regressors considered by BFK, the nature of super consistency of $\tilde{k}$ allows us to estimate b_i in the regimes split by $\tilde{k}$ as if k_0 were known. The impact of the difference between $\tilde{k}$ and k_0 can be ignored asymptotically, i.e., $[\mathbb{X}_i(k_0) - \mathbb{X}_i(\tilde{k})]b_i = o_p(1)$. As discussed above, the cross-sectional averages of the data $\bar{w}_t$ can be used to proxy the unobservable factors f_t, i.e., $(F - \bar{W}\mathbb{C})\gamma_i = o_p(1)$.

In the presence of endogenous regressors, we run an augmented IV regression with extra regressors $\bar{W}_t$. Given that $\bar{u}_t(k_0) \xrightarrow{p} 0$ in equation (16), the correlation between x_{it} (or v_{it}) and ε_{it} vanishes in $\bar{w}_t$ when N is large. This implies that $\bar{W}$ can be treated as exogenous asymptotically, and can be included as the first-stage regressors along with instruments. Similar to the definition of $\mathbb{X}_i(k)$, we define the instrument matrix $\mathbb{Z}_i(k) = \begin{pmatrix} Z_{1i}(k) & 0 \\ 0 & Z_{2i}(k) \end{pmatrix}$ where $Z_{1i}(k) = (z'_{i1}, \cdots, z'_{ik})'$ and $Z_{2i}(k) = (z'_{ik+1}, \cdots, z'_{iT})'$. Denote $Z_i^+(k) = (\mathbb{Z}_i(k), \bar{W})$. The predicted value of $\mathbb{X}_i(\tilde{k})$ is $\widehat{\mathbb{X}}_i(\tilde{k}) = P_{Z_i^+(\tilde{k})}\mathbb{X}_i(\tilde{k})$. Using $\tilde{k}$, the IV estimator of b_i is given by

$$\tilde{b}_{i,IV} = \tilde{b}_{i,IV}(\tilde{k}) = [\widehat{\mathbb{X}}_i(\tilde{k})' M_w \widehat{\mathbb{X}}_i(\tilde{k})]^{-1} \widehat{\mathbb{X}}_i(\tilde{k})' M_w Y_i, \tag{23}$$

$i = 1, ..., N$.

Compared with equation (23) in BFK in the case of exogeneity, the key difference here in (23) is replacing endogenous regressors $\mathbb{X}_i(\tilde{k})$ with their predicted values $\widehat{\mathbb{X}}_i(\tilde{k})$ using instruments $\mathbb{Z}_i(\tilde{k})$ and $\bar{W}$. As in Pesaran (2006) and BFK, the mean group estimator of b, the cross-sectional mean of b_i, $i = 1, ..., N$, is defined as:

$$\tilde{b}_{MG} = \tilde{b}_{MG}(\tilde{k}) = \frac{1}{N}\sum_{i=1i}^{N} \tilde{b}_{i,IV}(\tilde{k}) = \frac{1}{N}\sum_{i=1}^{N} [\widehat{\mathbb{X}}_i(\tilde{k})' M_w \widehat{\mathbb{X}}_i(\tilde{k})]^{-1} \widehat{\mathbb{X}}_i(\tilde{k})' M_w Y_i. \tag{24}$$

Proposition 1 *Under Assumptions 1-11 in the online Appendix,*

$$\sqrt{N}(\tilde{b}_{MG} - b) \xrightarrow{d} N(0, \Sigma_b).$$

As in Pesaran (2006), one of the advantages of the mean group estimator is that Σ_b can be consistently estimated by

$$\frac{1}{N-1}\sum_{i=1}^{N}(\tilde{b}_{i,IV}-\tilde{b}_{MG})(\tilde{b}_{i,IV}-\tilde{b}_{MG})'.$$

4 Monte Carlo Simulations

In this section we examine the properties of the break point estimator using Monte Carlo simulations. Due to space limitations the design as well as Monte Carlo results are relegated the online appendix which is not intended for publication. Briefly, the data generating process (DGP) is a modified design of BFK's Model 2 and is similar to Pesaran's (2006). The main difference is that x_{it} (or v_{it}) and ε_{it} are correlated in the DGP,

$$\begin{aligned} x_{it} &= a_i + \gamma_{2i} f_t + v_{it} \\ e_{it} &= \gamma_{1i}(k_1) f_t + \rho_{e,i} v_{it} + (1-\rho_e^2)^{1/2}\varepsilon_{it} \end{aligned}$$

We check the impact of endogeneity on the consistency of the break point estimator using various experiments. The results indicate that Theorem 1 is supported in the sense that the empirical distribution of the break point estimator tends to collapse to k_0 as N increases.

Figure 1 reports the case of rank deficiency. By changing the distribution of γ_{2i} from $N(0.5, 0.5)$ considered in Figure A2 in the online appendix to $N(0, 0.5)$, the matrix $\bar{C}(k_0)$ is not of full rank asymptotically. The first panel of Figure 1 shows that the consistency of $\tilde{k}$ remains in the case of rank deficiency. As N increases, the probability of choosing the true value k_0 increases.

In Figure 2, we also compare the efficiency of the proposed OLS and IV estimators of k_0. An IV estimator is used in the first step, instead of OLS, in a simplified case without an error factor structure. The DGP is similar to the one used in Figure 1 except that an instrument is introduced and regressor x_{it} is generated in a slightly different way, similar to Hall et al. (2012). As expected, the IV estimator $\check{k}$ is also consistent, and its probability of choosing the true value k_0 increases with N (and T). However, comparison between the histograms of $\hat{k}$ and $\check{k}$ suggest that OLS yields more accuracy in terms of the probability of finding the true value k_0 than the IV estimator.

5 Conclusion

In empirical studies with long panel data sets using the CCE approach, endogenous regressors and structural change are two main concerns. This paper extends Pesaran (2006) and BFK by allowing for endogenous regressors and unknown common structural changes in slopes and error factor loadings in large heterogeneous panels. This paper can also be considered as an extension of the time series regression models studied by Boldea, Hall and Han (2012), Hall, Han and Boldea (2012), Perron and Yamamoto (2014, 2015) to heterogeneous panels with an error factor structure.

We show that the model considered in this paper can be estimated by combining Pesaran's CCE approach and the least squares method proposed by Bai (1997, 2010). The paper also shows that the CCE approach is still valid to control for cross-section dependence due to error factors even in the presence of endogeneity and structural changes in slopes and error factor loadings. Common break points can be consistently estimated by least squares even in the presence of endogeneity. Monte Carlo experiments here and in the online Appendix are used to verify the consistency of common break estimators in various cases.

References

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

[2] Bai, J. (2009), Panel Data Models with Interactive Fixed Effects, *Econometrica*, 77, 1229-1279.

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

[4] Bai, J., Han, X., and Shi, Y., (2017), Estimation and Inference of Change Points in High-Dimensional Factor Models, Working paper, Columbia University.

[5] Baltagi, B.H., Feng, Q., and Kao, C. (2016), Estimation of Heterogeneous Panels with Structural Breaks, *Journal of Econometrics*, 191, 176-195.

[6] Baltagi, B.H., Kao, C., and Liu, L. (2017), Estimation and Identification of Change Points in Panel Models with Nonstationary or Stationary Regressors and Error Term, *Econometric Reviews*, 36, 85-102.

[7] Boldea, O., Hall, A.R., and Han, S. (2012), Asymptotic Distribution Theory for Break Point Estimators in Models Estimated via 2SLS, *Econometric Reviews* 31, 1-33.

[8] Boldea, O., Gan, Z., and B. Drepper, (2016), Break Point Estimation in Fixed Effects Panel Data, Working Paper, Tilburg University.

[9] Breitung, J., and Eickmeier, S. (2011), Testing for Structural Breaks in Dynamic Factor Models, *Journal of Econometrics*, 163, 71-84.

[10] Chen, L., Dolado, J., and Gonzalo, J. (2014), Detecting Big Structural Breaks in Large Factor Models, *Journal of Econometrics,* 180, 30-48.

[11] Chen, B. (2015), Modeling and Testing Smooth Structural Changes with Endogenous Regressors, *Journal of Econometrics* 185, 196-215.

[12] Cheng, X., Liao, Z., and Schorfheide, F., (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies* 83, 1511-1543.

[13] Chudik, A., and Pesaran, M. H. (2015), Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors, *Journal of Econometrics* 188, 393-420.

[14] Feng, Q., Kao, C., and Lazarova, S. (2009), Estimation of Change Points in Panels, Working Paper, Syracuse University.

[15] Forchini, G., Jiang, B., and Peng, B. (2015), Consistent Estimation in Large Heterogeneous Panels with Multifactor Structure Endogeneity, Working Paper, Monash University.

[16] Hall, A.R., Han, S, and Boldea, O. (2012), Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors, *Journal of Econometrics* 170, 281-302.

[17] Harding, M., and Lamarche, C. (2011), Least Squares Estimation of a Panel Data Model with Multifactor Error Structure and Endogenous Covariates, *Economics Letters*, 111, 197-199.

[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] Ma, S., and Su, L. (2018). Estimation of Large Dimensional Factor Models with an Unknown Number of Breaks, *Journal of Econometrics*, forthcoming.

[20] Li, D., Qian, J., and Su, L. (2016), Panel Data Models with Interactive Fixed Effects and Multiple Structural Breaks, *Journal of the American Statistical Association*, 111, 1804-1819.

[21] Neal, T. (2015), Estimating Heterogeneous Coefficients in Panel Data Models with Endogenous Regressors and Common Factors, working paper, University of New South Wales.

[22] Perron, P., and Yamamoto, Y. (2014), A Note on Estimating and Testing for Multiple Structural Changes in Models with Endogenous Regressors via 2SLS, *Econometric Theory*, 30(2), 491-507.

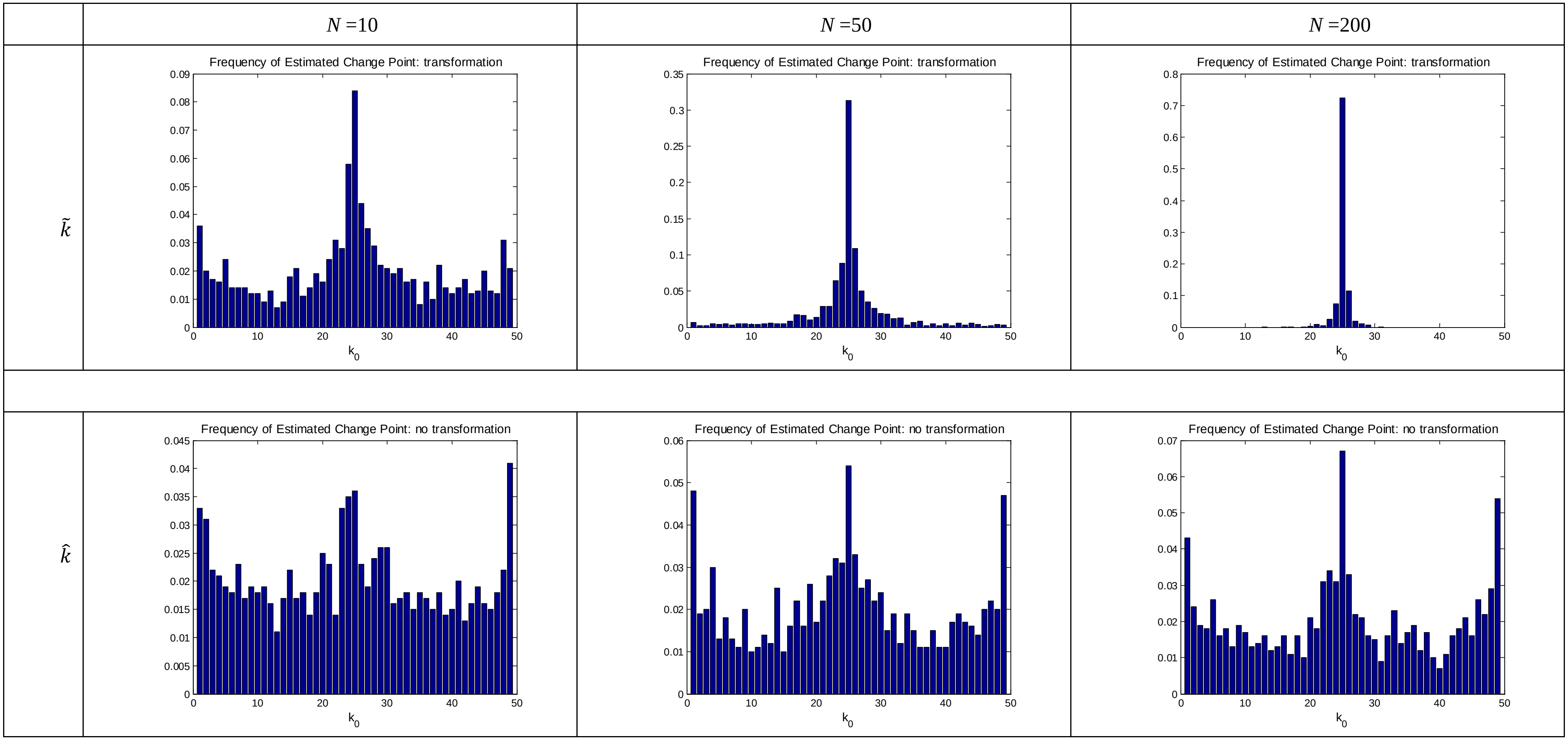
[23] Perron, P., and Yamamoto, Y. (2015), Using OLS to Estimate and Test for Multiple Structural Changes in Models with Endogenous Regressors, *Journal of Applied Econometrics*, 30(1), 119-144.

[24] Pesaran, M. H. (2006), Estimation and Inference in Large Heterogeneous Panels with Multifactor Error Structure, *Econometrica*, 74, 967-1012.

[25] Stock, J., and Watson, M. (2009), Forecasting in Dynamic Factor Models Subject to Structural Instability. In: Shephard N, Castle J, *The Methodology and Practice of Econometrics: A Festschrift in Honor of D.F. Hendry*. Oxford University Press, 173-205.

[26] Yamamoto, Y., and Tanaka, S. (2015), Testing for Factor Loading Structural Change under Common Breaks, *Journal of Econometrics*, 189, 187-206.

Figure 1: Histograms of $\tilde{k}$ and $\hat{k}$ in the general case with rank deficiency (T =50)



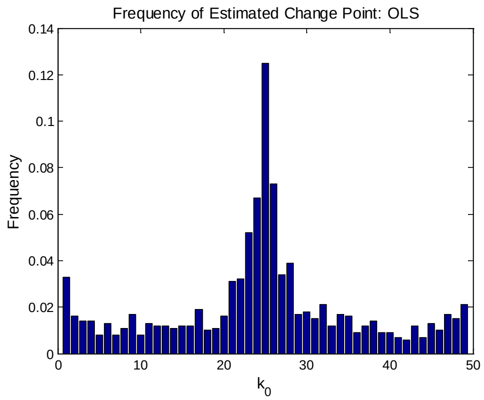
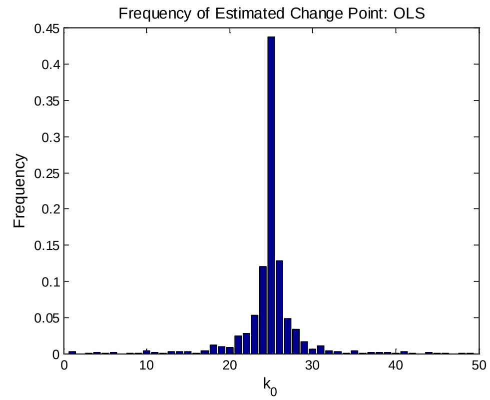
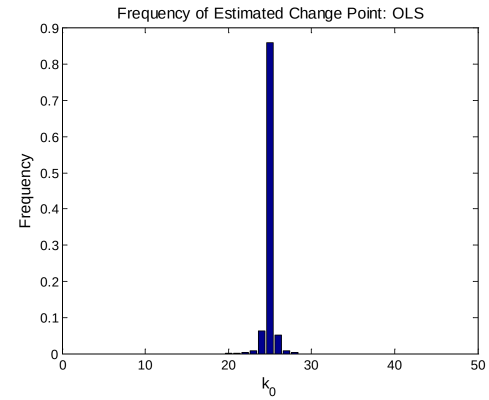
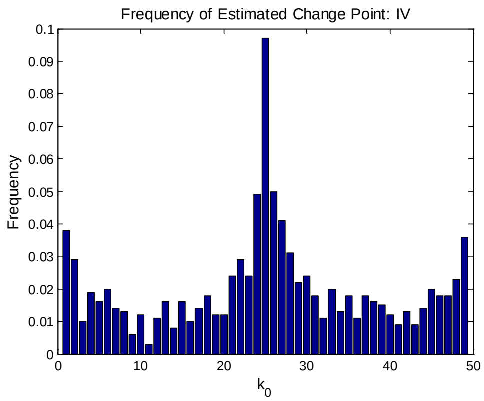
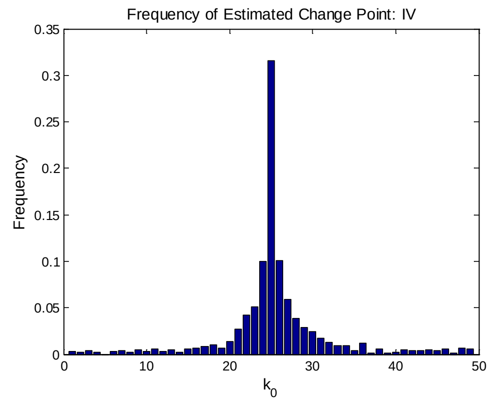
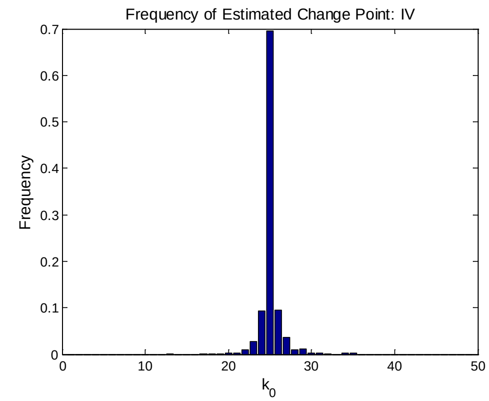
Note: The DGP is a modified design of Model 2 in BFK (2016). The regressors x_{it} are correlated with e_{it}.

$y_{it}=\alpha_i+\beta_i(k_0)x_{i,t}+e_{it}, i=1,\dots,N; t=1,\dots,T.$ $\alpha_i \sim iidN(1,1)$, $\beta_i(k_0)=\begin{cases}\beta_{1i}, & t=1,\dots,k_0,\\ \beta_{2i}=\beta_{1i}+\delta_i, & t=k_0+1,\dots,T.\end{cases}$ $k_0=0.5T$, $\beta_{1i}\sim iidN(1,0.04)$, $\delta_i\sim iidN(0,0.04)$.

$x_{it}=a_i+\gamma_{2i}f_t+v_{it}$; $e_{it}=\gamma_{1i}f_t+\rho_{e,i}v_{it_t}+\left(1-\rho_{e,i}^2\right)^{1/2}\varepsilon_{it}$, $\gamma_{1i}\sim iidN(1,0.2)$, $\rho_{e,i}\sim iidU(-0.5,0.5)$. $f_t=\rho_f f_{t-1}+v_{ft}, t=-49,\dots,0,1,\dots T$, $v_{ft}\sim iidN(0,1-\rho_f^2)$, $\rho_f=0.5$, $f_{-50}=0$. $\varepsilon_{it}\sim iidN(0,\sigma_i^2)$, $\sigma_i^2\sim iidU(0.5,1.5)$, $\gamma_{i1}\sim iidN(1,0.2)$, $v_{it}\sim iidN(0,1-\rho_{vi}^2)$, $\rho_{vi}=0.5$. These variables are mutually independent. The replication number is 1000. $T=50, k_0=25$. Different from the design of Figure A2 in the online appendix, the means of γ_{i2} and a_i change to zero, i.e., $\gamma_{i2}\sim iidN(0,0.5)$, $a_i\sim iidN(0,0.5)$, so the rank condition is not satisfied asymptotically.

$\tilde{k}$: The OLS estimator of the change point after removing the common correlated factors. $\hat{k}$: The OLS estimator of the change point without removing the common correlated factors.

Figure 2: Histograms of the OLS estimator $\hat{k}$ and IV estimator $\check{k}$ in a simplified case without a factor structure in the errors (T =50)

	N =10	N =50	N =200
$\hat{k}$			
$\check{k}$			

Note: In this simplified case, there is no factor structure in the errors. The instrument $z3_{it}$ is introduced and regressor x_{it} is generated in a slightly different way (similar to Hall et al., 2012). $z3_{it} = 2a_i + \gamma_{3i} f_t + v2_{it}$ where $\gamma_{3i} \sim iidN(1, 0.5)$, $v2_{it} \sim iidN(0, 1)$, and $v2_{it}$ is independent of v_{it} and ε_{it}.

$x_{it} = 0.5z3_{it} + v_{it}$; $e_{it} = \rho_{e,i} v_{it} + \left(1 - \rho_{e,i}^2\right)^{1/2} \varepsilon_{it}$, $\rho_{e,i} \sim iidU(-0.5, 0.5)$. $\varepsilon_{it} \sim iidN(0, \sigma_i^2)$, $\sigma_i^2 \sim iidU(0.5, 1.5)$, $\gamma_{1i} \sim iidN(1, 0.2)$, $\gamma_{2i} \sim iidN(0.5, 0.5)$, $a_i \sim iidN(0.5, 0.5)$,

$v_{it} \sim iidN(0, 1 - \rho_{vi}^2)$, $\rho_{vi} = 0.5$. These variables are mutually independent. The replication number is 1000. $T = 50, k_0 = 25$.

$\hat{k}$: The OLS estimator of the change point.

$\check{k}$: The IV estimator of the change point: the IV estimator is used in the first step, instead of OLS.

Online Appendix (not for publication)**:**

Structural Changes in Heterogeneous Panels with Endogenous Regressors

by Badi H. Baltagi, Qu Feng, Chihwa Kao

A1 Online Appendix 1: Mathmetical Proofs

This online Appendix 1 includes the assumptions required in the text and the proofs of Theorem 1 and Proposition 1.

A1.1 Assumptions

To estimate the common change point k_0, we need the following assumptions:

Assumption 1 *$k_0 = [\tau_0 T]$, where $\tau_0 \in (0,1)$ and $[\cdot]$ is the greatest integer function.*

This assumption is standard in the literature, including Bai (1997), Bai (2010), Kim (2011) and BFK. It assumes that there are enough observations to consistently estimate the slopes in each regime.

Define $\Delta\beta_i = \beta_{2i} - \beta_{1i}$ and $\phi_N = \sum_{i=1}^{N} \Delta\beta_i' \Delta\beta_i$. For series i, $\Delta\beta_i'\Delta\beta_i$ measures the magnitude of the structural break, thus ϕ_N is an indicator of the break magnitude for all N series sharing a common break.

Assumption 2 *$\phi_N \to \infty$ and $\phi_N \frac{T}{N} \to \infty$ as $(N,T) \to \infty$.*

Assumption 3 *(i) The disturbances $\varepsilon_{it}, i = 1, ..., N$, are cross-sectionally independent; (ii) ε_{it} is a stationary process with absolute summable autocovariances,*

$$\varepsilon_{it} = \sum_{l=0}^{\infty} a_{il} \zeta_{i,t-l}$$

where $\{\zeta_{it}, t = 1, ..., T\}$ are independent and identically distributed (IID) random variables with finite fourth-order cumulants. Assume $0 < Var(\varepsilon_{it}) = \sum_{l=0}^{\infty} a_{il}^2 = \sigma_i^2 < \infty$. Also, for the $T \times 1$ vector $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iT})'$, $Var(\varepsilon_i) = \Sigma_{\varepsilon,i}$.

Assumption 4 *Common factors* f_t, $t = 1, ..., T$, *are covariance sationary with absolute summable autocovariances, independent of the errors* ε_{is} *and* v_{is} *for all* i, s, t.

Assumption 5 v_{it} *can be correlated with* ε_{it}. v_{it}, $i = 1, ..., N$, *are linear stationary processes with absolute summable autocovariances,* $v_{it} = \sum_{l=0}^{\infty} S_{il} \upsilon_{i,t-l}$, *where* $(\zeta_{it}, \upsilon'_{it})'$ *are* $(p+1) \times 1$ *vectors of IID random variables with variance-covariance matrix* I_{p+1} *and finite fourth-order cumulants, and*

$$Var(v_{it}) = \sum_{l=0}^{\infty} S_{il} S'_{il} = \Sigma_{i,v}, \text{ and } 0 < \|\Sigma_{i,v}\| < \infty.$$

Let $\gamma_i = (\gamma'_{1i}, \gamma'_{2i})'$ for $i = 1, ..., N$, in the case of a break in error factor loading. Otherwise, γ_i is defined as $\gamma_{1i} = \gamma_{2i}$.

Assumption 6 *Factor loadings* g_i *and* Γ_i *are IID across* i, *and independent of* ε_{jt}, v_{jt} *and* f_t *for all* i, j, t. *Assume* $\gamma_i = \gamma + \eta_i$, $\eta_i \sim IID(0, \Omega_\eta)$ *and* $\Gamma_i = \Gamma + \xi_i$, $\xi_i \sim IID(0, \Omega_\xi)$, $i = 1, ...N$, *where the means* γ, Γ *are non-zero and fixed, and the variances* Ω_η, Ω_ξ *are finite.*

Compared with Assumptions 1-3 of Pesaran (2006), Assumptions 3, 4, 5 and 6 allow for correlation between v_{it} (x_{it}) and ε_{it}. In addition, the restrictions $\gamma \neq 0$ and $\Gamma \neq 0$ are required to exclude the case of deficient rank.

Under a random coefficient model considered in Pesaran (2006) and BFK:

Assumption 7 *For* $i = 1, ..., N$,

$$b_i = b + v_{b,i}, v_{b,i} \sim IID(0, \Sigma_b),$$

where $b = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$, $v_{b,i} = \begin{pmatrix} v_{\beta_1,i} \\ v_{\beta_2,i} \end{pmatrix}$ *and* $\Sigma_b \sim \begin{pmatrix} \Sigma_{\beta_1} & 0 \\ 0 & \Sigma_{\beta_2} \end{pmatrix}$, $\|b\|$ *and* $\|\Sigma_b\| < \infty$, *and the random deviations* $v_{b,i}$ *are independent of* γ_j, Γ_j, ε_{jt}, *and* v_{jt} *for all* i, j *and* t.

For any matrix or vector A, the norm of A is defined as $\|A\| = \sqrt{tr(AA')}$. Under Assumption 7, b_i is independent of Γ_j, implying that as $N \to \infty$, $\bar{C}_1 = \sum_{i=1}^{N} \theta_i C_{1i} \xrightarrow{p} E(C_{1i}) = (\gamma_1 + \Gamma\beta_1, \Gamma)$, $\bar{C}_2 \xrightarrow{p} E(C_{2i}) = (\gamma_1 + \Gamma\beta_2, \Gamma)$ and $\bar{C}_3 \xrightarrow{p} E(C_{3i}) = (\gamma_2 + \Gamma\beta_2, \Gamma)$.

Assumption 8 $Rank(\bar{C}_1) = Rank(\bar{C}_2) = Rank(\bar{C}_3) = m \leq p + 1$.

Different from BFK's Theorem 1, to accommodate the correlation between ε_{it} and x_{it}, heterogeneity and also lagged dependent variables, we follow Assumption 4 of Perron and Yamamoto (2015) and require an additional assumption. Let $\tilde{x}'_{it}$ and $\tilde{\varepsilon}_{it}$ be the t^{th} elements of $\tilde{\mathbb{X}}_i(k_0)$ and $\tilde{\varepsilon}_i$, respectively.

Assumption 9 *Let the L^r-norm of a random matrix A be defined by $\|A\|_r = (\sum_i \sum_j E|A_{ij}|^r)^{1/r}$ for any $r \geq 1$. With $\{\mathcal{F}_t : t = 1, 2, ...\}$ a sequence of increasing σ-fields, we assume that $\{x_{it}\varepsilon_{it}, \mathcal{F}_t\}$, $\{\tilde{x}_{it}\tilde{\varepsilon}_{it}, \mathcal{F}_t\}$ form an L^r-mixingale sequence with $r = 2 + \epsilon$ for some $\epsilon > 0$ for each series i.*

Let $\tilde{x}'_{it}$ be the t^{th} element of matrix $\tilde{\mathbb{X}}_i(k_0)$, $i = 1, ..., N$. As in Perron and Yamamoto (2015), define $\tilde{h}_{it} = (\tilde{x}'_{it}, z'_{it})'$.

Assumption 10 *For $i = 1, ..., N$, $(1/j)\sum_{t=1}^{j} \tilde{h}_{it}\tilde{h}'_{it}$, $(1/j)\sum_{t=T-j+1}^{T} \tilde{h}_{it}\tilde{h}'_{it}$, $(1/j)\sum_{t=k_0-j+1}^{k_0} \tilde{h}_{it}\tilde{h}'_{it}$ and $(1/j)\sum_{t=k_0+1}^{k_0+j} \tilde{h}_{it}\tilde{h}'_{it}$ are stochastically bounded and have minimum eigenvalues bounded away from zero in probability for all large j. In addition, for each i, $(1/T)\sum_{t=1}^{T} \tilde{h}_{it}\tilde{h}'_{it}$ converges in probability to a nonrandom and positive definite matrix as $T \to \infty$.*

Assumption 11 *For any positive finite integer s, the matrices $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=k_0-s+1}^{k_0} \tilde{h}_{it}\tilde{h}'_{it}$ and $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=k_0+1}^{k_0+s} \tilde{h}_{it}\tilde{h}'_{it}$, $i = 1, ..., N$, are stochastically bounded, with minimum eigenvalues bounded away from zero in probability for large N. In addition, for each t, $(1/N)\sum_{i=1}^{N} \tilde{h}_{it}\tilde{h}'_{it}$ is stochastically bounded as $N \to \infty$.*

To identify the cross-sectional means of slopes, as in Pesaran (2006), we assume:

Assumption 12 *For $i = 1, ..., N$, matrices $\frac{1}{T}\widehat{\mathbb{X}}_i(k_0)'M_w\widehat{\mathbb{X}}_i(k_0)$ are nonsingular, and their inverses have finite second-order moments.*

A1.2 Proof of Theorem 1

In the general case with a factor structure in the error term,

$$Y_i = \mathbb{X}_i(k_0)b_i + F\gamma_i + \varepsilon_i, i = 1, ..., N,$$

as shown in Section 3.2, the CCE approach proposed by Pesaran (2006) is still valid in the presence of endogeneity and structural changes in slopes and factor structures. The unobserved common factors F can be proxied by the cross-sectional averages of the data. In the transformed model (20)

$$\tilde{Y}_i = \tilde{\mathbb{X}}_i(k_0)b_i + M_w F\gamma_i + \tilde{\varepsilon}_i,$$

$\tilde{k}$ is the least squares estimator of k_0.

To show $\tilde{k} - k_0 = o_p(1)$, we can follow the same way of proving BFK's Theorem 1 in the case of exogeneity. The complication comes from endogenous regressors. We need to show that:

(i) the magnitude of the extra term $M_w F\gamma_i$ is unaffected by endogeneity; and that

(ii) the consistency of the least squares estimator $\tilde{k}$ still holds in the presence of correlation between $\tilde{\mathbb{X}}_i(k_0)$ and $\tilde{\varepsilon}_i$ as in the simplified case (7) in Section 3.1.

The key step of proving part (i) is to show that BFK's Lemma 4 still holds in the presence of endogeneity. If so, BFK's Lemmas 5 and 6 remain unchanged. This implies that each element of $M_w F\gamma_i$ is of order $O_p(\frac{1}{\sqrt{N}})$, vanishing as $N \to \infty$. Thus, the extra term $M_w F\gamma_i$ can be ignored asymptotically, and k_0 can be estimated by least squares using transformed data as in the simplified case.

For part (ii), we follow Perron and Yamamoto (2015) and project $\tilde{\varepsilon}_i$ on the column space spanned by $\tilde{\mathbb{X}}_i(k_0)$ such that the resulting new error term $\tilde{\varepsilon}_i^*$ is uncorrelated with $\tilde{\mathbb{X}}_i(k_0)$. After controlling for the endogeneity, BFK's Lemmas 7, 8, 9 still hold, therefore, $\tilde{k} - k_0 = o_p(1)$ can be shown similarly. In this way, Theorem 1 is proved.

In what follows, we prove that BFK's Lemma 4 still holds under endogeneity. To proceed, we follow BFK's notation. Since $x_{it} = \Gamma_i' f_t + v_{it}$ in (3), we write

$$\underset{T\times p}{X_i} = \underset{T\times m}{F}\,\underset{m\times p}{\Gamma_i} + \underset{T\times p}{V_i},$$

where $V_i = (v_{i1,} \cdots, v_{iT})'$. Denote $F_0 = (0, \cdots, 0, f_{k_0+1}, \cdots, f_T)'$ and $V_{0i} = (0, \cdots, 0, v_{i,k_0+1}, \cdots, v_{i,T})'$. Thus,

$$\begin{aligned} X_{0i} &= (0, \cdots, 0, x_{i,k_0+1}, \cdots, x_{i,T})' = (0, \cdots, 0, \Gamma_i' f_{k_0+1} + v_{i,k_0+1}, \cdots, \Gamma_i' f_T + v_{i,T})' \\ &= F_0 \Gamma_i + V_{0i}. \end{aligned}$$

For the error term (16), denote $\bar{u}_t = \begin{pmatrix} \bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}' \beta_i \\ \bar{v}_t \end{pmatrix}$ and

$$\Delta \bar{u}_t(k_0) = \begin{cases} \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & t = 1, ..., k_0, \\ \begin{pmatrix} \sum_{i=1}^N \theta_i v_{it}' \Delta\beta_i \\ 0 \end{pmatrix}, & t = k_0 + 1, ..., T. \end{cases}$$

Thus, $\bar{u}_t(k_0) = \sum_{i=1}^N \theta_i u_{it}(k_0) = \bar{u}_t + \Delta \bar{u}_t(k_0)$. Denote $\bar{U} = (\bar{u}_1, ..., \bar{u}_T)'$ and

$$\Delta \bar{U}(k_0) = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{i=1}^N \theta_i v_{i,k_0+1}' \Delta\beta_i \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} \sum_{i=1}^N \theta_i v_{i,T}' \Delta\beta_i \\ 0 \end{pmatrix} \right)'.$$

Thus, stacking cross-sectional averages $\bar{w}_t = \bar{C}(k_0)' f_t + \bar{u}_t(k_0)$, we obtain

$$\begin{aligned} \underset{T\times(p+1)}{\bar{W}} &= (\bar{w}_1, ..., \bar{w}_{k_0}, \bar{w}_{k_0+1}, \cdots, \bar{w}_T)' \\ &= (\bar{C}_1' f_1 + \bar{u}_1, ..., \bar{C}_1' f_{k_0} + \bar{u}_{k_0}, \bar{C}_2' f_{k_0+1} + \bar{u}_{k_0+1}, \cdots, \bar{C}_2' f_T + \bar{u}_T)' \\ &= F\bar{C}_1 + F_0(\bar{C}_2 - \bar{C}_1) + \bar{U} + \Delta\bar{U}(k_0). \end{aligned}$$

Denote $\underset{T\times 2m}{\mathbb{F}} = (F, F_0)$ and $\underset{T\times(p+1)}{\bar{\mathbb{U}}} = \bar{U} + \Delta\bar{U}(k_0)$. Therefore,

$$\bar{W} = \mathbb{F}(\bar{C}_1', (\bar{C}_2 - \bar{C}_1)')' + \bar{\mathbb{U}}. \tag{25}$$

Lemma 1 *Under Assumptions 1-11, uniformly on the set*

$K(C_k) = \{k : 1 \le |k - k_0| < C_k,\ aT < k < (1-a)T\}$, *where* C_k *is a finite large number and* $a < \tau_0$ *is an arbitrarily small positive number.*

(i) $\bar{u}_t = O_p(\frac{1}{\sqrt{N}})$, $\Delta\bar{u}_t(k_0) = O_p(\frac{1}{\sqrt{N}})$;

(ii) $\frac{1}{T}\bar{\mathbb{U}}'\bar{\mathbb{U}} = O_p(\frac{1}{N})$; $\frac{1}{T}\mathbb{F}'\bar{\mathbb{U}} = O_p(\frac{1}{\sqrt{NT}})$, $\frac{1}{T}V_i'\mathbb{F} = O_p(\frac{1}{\sqrt{T}})$;

(iii) $\frac{1}{T}V_i'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, $\frac{1}{T}\varepsilon_i'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, $\frac{1}{T}V_{0i}'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$;

(iv) $\frac{1}{T}X_i'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$; $\frac{1}{T}X_{0i}'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$.

Now we verify that this lemma holds in the case of endogeneity, i.e., $Var(\varepsilon_{it}, v_{it}) \neq 0$ (but $Var(\varepsilon_{it}, v_{jt}) = 0$ for $i \neq j$). We need to check the terms affected by $Var(\varepsilon_{it}, v_{it}) \neq 0$, but not those under $Var(\varepsilon_{it}, v_{jt}) = 0$ for $i \neq j$.

Consider (i):

$$\bar{u}_t = \begin{pmatrix} \bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}'\beta_i \\ \bar{v}_t \end{pmatrix}$$

$Var(\varepsilon_{it}, v_{it}) \neq 0$ only affects the 1st component $\bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}'\beta_i$. But if we look at these two terms separately, their orders of magnitude are not affected by $Var(\varepsilon_{it}, v_{it}) \neq 0$. Thus, $\bar{u}_t = O_p(\frac{1}{\sqrt{N}})$ remains unchanged. ($\Delta\bar{u}_t(k_0)$ is dominated by $\bar{u}_t$.)

(ii) Consider the first part:

$$\frac{1}{T}\bar{\mathbb{U}}'\bar{\mathbb{U}} = \frac{1}{T}\bar{U}'\bar{U} + \frac{2}{T}\bar{U}'\Delta\bar{U}(k_0) + \frac{1}{T}\Delta\bar{U}(k_0)'\Delta\bar{U}(k_0).$$

Since the 2nd and 3rd terms are dominated by the 1st one, we only need to consider

$$\begin{aligned} \frac{1}{T}\bar{U}'\bar{U} &= \frac{1}{T}\sum_{t=1}^T \bar{u}_t'\bar{u}_t = \frac{1}{T}\sum_{t=1}^T \left[(\bar{\varepsilon}_t + \textstyle\sum_{i=1}^N \theta_i v_{it}'\beta_i)(\bar{\varepsilon}_t + \sum_{i=1}^N \theta_i v_{it}'\beta_i) + \bar{v}_t'\bar{v}_t\right] \\ &= \frac{1}{T}\sum_{t=1}^T \left[\bar{\varepsilon}_t^2 + 2\bar{\varepsilon}_t \textstyle\sum_{i=1}^N \theta_i v_{it}'\beta_i + (\sum_{i=1}^N \theta_i v_{it}'\beta_i)^2 + \bar{v}_t'\bar{v}_t\right]. \end{aligned}$$

Notice that $Var(\varepsilon_{it}, v_{it}) \neq 0$ only affects the 2nd term $\frac{1}{T}\sum_{t=1}^T \bar{\varepsilon}_t \sum_{i=1}^N \theta_i v_{it}'\beta_i$. Now we check whether endogeneity changes the result $O_p(\frac{1}{N})$. For simplicity, assume $\theta_i = 1/N$ and

$p = 1$.

$$\frac{1}{T}\sum_{t=1}^{T}\bar{\varepsilon}_t \sum_{i=1}^{N}\theta_i v_{it}'\beta_i = \frac{1}{TN^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{it}v_{jt}\beta_j$$

$$= \frac{1}{TN^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{it}v_{it}\beta_i + \frac{1}{TN^2}\sum_{t=1}^{T}\sum_{i\neq j}^{N}\sum_{j=1}^{N}\varepsilon_{it}v_{jt}\beta_j$$

The 2nd term above $\frac{1}{TN^2}\sum_{t=1}^{T}\sum_{i\neq j}^{N}\sum_{j=1}^{N}\varepsilon_{it}v_{jt}\beta_j$ is not affected by endogeneity. We only consider the first term $\frac{1}{TN^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{it}v_{it}\beta_i$. It is easy to verify that it is bounded by $O_p(\frac{1}{N})$. (Note: $\frac{1}{TN^2}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{it}v_{it}\beta_i = O_p(\frac{1}{\sqrt{T}N})$ under exogeneity.) Thus, we have verified that $\frac{1}{T}\bar{U}'\bar{U} = O_p(\frac{1}{N})$. Thus, $\frac{1}{T}\bar{\mathbb{U}}'\bar{\mathbb{U}} = O_p(\frac{1}{N})$.

(ii) 2nd part $\frac{1}{T}\mathbb{F}'\bar{\mathbb{U}}$ and the third part $\frac{1}{T}V_i'\mathbb{F}$ are not affected by endogeneity since no product of v_{it} and ε_{it} appears in these two parts. Thus, the results still hold, and ii) is satisfied under endogeneity.

(iii) Consider the first part $\frac{1}{T}V_i'\bar{\mathbb{U}} = \frac{1}{T}V_i'\bar{U} + \frac{1}{T}V_i'\Delta\bar{U}(k_0)$. The 2nd term above $\frac{1}{T}V_i'\Delta\bar{U}(k_0)$ is dominated by the first term $\frac{1}{T}V_i'\bar{U}$, so we only consider

$$\frac{1}{T}V_i'\bar{U} = \frac{1}{T}\sum_{t=1}^{T}v_{it}\bar{u}_t' = \frac{1}{T}\sum_{t=1}^{T}v_{it}(\bar{\varepsilon}_t + \sum_{i=1}^{N}\theta_i v_{it}'\beta_i, \bar{v}_t')$$

$$= (\frac{1}{T}\sum_{t=1}^{T}v_{it}\bar{\varepsilon}_t + \frac{1}{T}\sum_{t=1}^{T}v_{it}\sum_{i=1}^{N}\theta_i v_{it}'\beta_i, \frac{1}{T}\sum_{t=1}^{T}v_{it}\bar{v}_t').$$

Since the endogeneity does not affect terms $\frac{1}{T}\sum_{t=1}^{T}v_{it}\sum_{i=1}^{N}\theta_i v_{it}'\beta_i$ and $\frac{1}{T}\sum_{t=1}^{T}v_{it}\bar{v}_t'$, we need to verify the first term

$$\frac{1}{T}\sum_{t=1}^{T}v_{it}\bar{\varepsilon}_t = \frac{1}{TN}\sum_{t=1}^{T}\sum_{j=1}^{N}v_{it}\varepsilon_{jt}$$

$$= \frac{1}{TN}\sum_{t=1}^{T}v_{it}\varepsilon_{it} + \frac{1}{TN}\sum_{t=1}^{T}\sum_{j\neq i}v_{it}\varepsilon_{jt}.$$

Since the 2nd term above $\frac{1}{TN}\sum_{t=1}^{T}\sum_{j\neq i}^{N}v_{it}\varepsilon_{jt}$ is not affected by $Var(\varepsilon_{it}, v_{it}) \neq 0$. $\frac{1}{TN}\sum_{t=1}^{T}v_{it}\varepsilon_{it} = O_p(\frac{1}{N})$ (instead of $O_p(\frac{1}{\sqrt{T}N})$ under exogeneity). Therefore, the first part $\frac{1}{T}V_i'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$ holds.

(iii) 2nd part: $\frac{1}{T}\varepsilon_i'\bar{\mathbb{U}} = \frac{1}{T}\varepsilon_i'(\bar{U}+\Delta\bar{U}(k_0)) = \frac{1}{T}\varepsilon_i'\bar{U}+\frac{1}{T}\varepsilon_i'\Delta\bar{U}(k_0)$. The 2nd term $\frac{1}{T}\varepsilon_i'\Delta\bar{U}(k_0)$ is dominated by the first term

$$\begin{aligned}\frac{1}{T}\varepsilon_i'\bar{U} &= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\bar{u}_t' = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}(\bar{\varepsilon}_t + \sum_{j=1}^{N}\theta_j v_{jt}'\beta_j, \bar{v}_t') \\ &= (\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\bar{\varepsilon}_t + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\sum_{j=1}^{N}\theta_j v_{jt}'\beta_j, \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\bar{v}_t').\end{aligned}$$

Among these three terms, $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\sum_{j=1}^{N}\theta_j v_{jt}'\beta_j$ and $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\bar{v}_t'$ are affected by endogeneity.

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\textstyle\sum_{j=1}^{N}\theta_j v_{jt}'\beta_j &= \frac{1}{TN}\sum_{t=1}^{T}\sum_{j=1}^{N}\varepsilon_{it}v_{jt}'\beta_j \\ &= \frac{1}{TN}\sum_{t=1}^{T}\varepsilon_{it}v_{it}'\beta_i + \frac{1}{TN}\sum_{t=1}^{T}\sum_{j\neq i}^{N}\varepsilon_{it}v_{jt}'\beta_j.\end{aligned}$$

The term $\frac{1}{TN}\sum_{t=1}^{T}\sum_{j\neq i}^{N}\varepsilon_{it}v_{jt}'\beta_j = O_p(\frac{1}{\sqrt{NT}})$ is not affected by endogeneity, and $\frac{1}{TN}\sum_{t=1}^{T}\varepsilon_{it}v_{it}'\beta_i = O_p(\frac{1}{N})$ (the bound under exogeneity is $O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, thus having endogeneity does not change the bound.)

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\bar{v}_t' = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\varepsilon_{it}v_{jt}' = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}).$$

Combining these three terms, we obtain $\frac{1}{T}\varepsilon_i'\bar{U} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, which is the same under exogeneity.

(iii) third part $\frac{1}{T}V_{0i}'\bar{\mathbb{U}}$ is dominated by the first part $\frac{1}{T}V_i'\bar{\mathbb{U}}$ by the definition of V_{0i} (setting 0 for the first k_0 rows of V_i). Thus,

$\frac{1}{T}V_{0i}'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$ unchanged under endogeneity.

(iv) The 2nd part $\frac{1}{T}X_{0i}'\bar{\mathbb{U}}$ is dominated by the first part $\frac{1}{T}X_i'\bar{\mathbb{U}}$, so we only consider

$$\frac{1}{T}X_i'\bar{\mathbb{U}} = \frac{1}{T}(F\Gamma_i + V_i)'\bar{\mathbb{U}} = \frac{1}{T}\Gamma_i'F'\bar{\mathbb{U}} + \frac{1}{T}V_i'\bar{\mathbb{U}}.$$

By (ii) and (iii), $\frac{1}{T}\Gamma_i'F'\bar{\mathbb{U}} = O_p(\frac{1}{\sqrt{NT}})$ and $\frac{1}{T}V_i'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, therefore, $\frac{1}{T}X_i'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, and $\frac{1}{T}X_{0i}'\bar{\mathbb{U}} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$ remain unchanged under endogeneity.

A1.3 Proof of Proposition 1

Under the random coefficient Assumption 7, by plugging in equations (22) and (23), the mean-group estimator (24)

$$\begin{aligned}
\tilde{b}_{MG} &= \frac{1}{N}\sum_{i=1}^{N}\tilde{b}_{i,IV}(\tilde{k}) = \frac{1}{N}\sum_{i=1}^{N}[\widehat{\mathbb{X}}_i(\tilde{k})'M_w\widehat{\mathbb{X}}_i(\tilde{k})]^{-1}\widehat{\mathbb{X}}_i(\tilde{k})'M_wY_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\left([\widehat{\mathbb{X}}_i(\tilde{k})'M_w\widehat{\mathbb{X}}_i(\tilde{k})]^{-1}\widehat{\mathbb{X}}_i(\tilde{k})'M_w\mathbb{X}_i(\tilde{k})b_i + [\widehat{\mathbb{X}}_i(\tilde{k})'M_w\widehat{\mathbb{X}}_i(\tilde{k})]^{-1}\widehat{\mathbb{X}}_i(\tilde{k})'M_w\varepsilon_i^0\right) \\
&= b + \frac{1}{N}\sum_{i=1}^{N}v_{b,i} + \frac{1}{N}\sum_{i=1}^{N}[\widehat{\mathbb{X}}_i(\tilde{k})'M_w\widehat{\mathbb{X}}_i(\tilde{k})]^{-1}\widehat{\mathbb{X}}_i(\tilde{k})'M_w\varepsilon_i^0,
\end{aligned}$$

where $\varepsilon_i^0 = \varepsilon_i + [\mathbb{X}_i(k_0) - \mathbb{X}_i(\tilde{k})]b_i + (F - \bar{W}\mathbb{C})\gamma_i$. We obtain

$$\begin{aligned}
&\sqrt{N}(\tilde{b}_{MG} - b) \\
=\; &\frac{1}{\sqrt{N}}\sum_{i=1}^{N}v_{b,i} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\widehat{\mathbb{X}}_i(\tilde{k})'M_w\widehat{\mathbb{X}}_i(\tilde{k})]^{-1}\widehat{\mathbb{X}}_i(\tilde{k})'M_w\varepsilon_i \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\widehat{\mathbb{X}}_i(\tilde{k})'M_w\widehat{\mathbb{X}}_i(\tilde{k})]^{-1}\widehat{\mathbb{X}}_i(\tilde{k})'M_w[\mathbb{X}_i(k_0) - \mathbb{X}_i(\tilde{k})]b_i \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\widehat{\mathbb{X}}_i(\tilde{k})'M_w\widehat{\mathbb{X}}_i(\tilde{k})]^{-1}\widehat{\mathbb{X}}_i(\tilde{k})'M_wF\gamma_i.
\end{aligned}$$

By Assumption 7, the limiting distribution of the first term is $N(0,\Sigma_b)$. The second and fourth term above are $O_p(\frac{1}{\sqrt{T}})$ and $O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}})$, respectively, as in the proof of Proposition 2 of BFK's Supplementary Appendix. In addition, by Theorem 1, $P(|\tilde{k}-k_0| \geq 1) \to 0$, it can be shown that the third term

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\widehat{\mathbb{X}}_i(\tilde{k})'M_w\widehat{\mathbb{X}}_i(\tilde{k})]^{-1}\widehat{\mathbb{X}}_i(\tilde{k})'M_w[\mathbb{X}_i(k_0) - \mathbb{X}_i(\tilde{k})]b_i = o_p(1).$$

Therefore, as $(N,T) \to \infty$,

$$\sqrt{N}(\tilde{b}_{MG} - b) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}v_{b,i} + o_p(1) \xrightarrow{d} N(0,\Sigma_b).$$

In the simplified case:

$$Y_i = \mathbb{X}_i(k_0)b_i + \varepsilon_i = \mathbb{X}_i(\hat{k})b_i + \varepsilon_i^0,$$

where $\varepsilon_i^0 = [\mathbb{X}_i(k_0) - \mathbb{X}_i(\hat{k})]b_i + \varepsilon_i$. As in the exogenous case in BFK, the super-consistency of $\hat{k}$ above allows us to treat k_0 as known. Due to the correlation between x_{it} and ε_{it}, the OLS estimator $\hat{b}_i = \hat{b}_i(\hat{k}) = [\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'Y_i$ is inconsistent.

Instead, b_i can be consistently estimated by the IV estimator:

$$\hat{b}_{i,IV} = \hat{b}_{i,IV}(\hat{k}) = [\mathbb{X}_i(\hat{k})'P_{\mathbb{Z}_i(\hat{k})}\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'P_{\mathbb{Z}_i(\hat{k})}Y_i$$

where the projection matrix $P_{\mathbb{Z}_i(\hat{k})} = \mathbb{Z}_i(\hat{k})[\mathbb{Z}_i(\hat{k})'\mathbb{Z}_i(\hat{k})]^{-1}\mathbb{Z}_i(\hat{k})'$. And b, the cross-sectional mean b_i under the random coefficients Assumption 7, can be consistently estimated by a mean group (called MG-IV) estimator:

$$\hat{b}_{MG-IV} = \hat{b}_{MG-IV}(\hat{k}) = \frac{1}{N}\sum_{i=1}^{N}\hat{b}_{i,IV}(\hat{k}) \tag{26}$$

For the purpose of comparison in the empirical example in Section 5, we define the mean group (MG) estimator using the OLS estimator $\hat{b}_i$ as follows:

$$\hat{b}_{MG} = \hat{b}_{MG}(\hat{k}) = \frac{1}{N}\sum_{i=1}^{N}\hat{b}_i(\hat{k}) = \frac{1}{N}\sum_{i=1}^{N}[\mathbb{X}_i(\hat{k})'\mathbb{X}_i(\hat{k})]^{-1}\mathbb{X}_i(\hat{k})'Y_i.$$

A2. Online Appendix 2: Monte Carlo Simulations

In this Appendix, we examine the properties of the break point estimator and check whether there is support for Theorem 1 using Monte Carlo simulations. The DGP used here is a modified design of BFK's Model 2 and is similar to Pesaran's (2006). The main difference is the correlation between x_{it} (or v_{it}) and ε_{it}. We check the impact of endogeneity on the consistency of the break point estimator using various experiments.

A2.1 Designs

The DGP:

$$y_{it} = \alpha_i + \beta_i(k_0)x_{i,t} + e_{it}, i = 1, ..., N; t = 1, ..., T$$

There is a common break $k_0 = 0.5T$ in the slopes β_i:

$$\beta_i(k_0) = \begin{cases} \beta_{1i}, & t = 1, ..., k_0, \\ \beta_{2i} = \beta_{1i} + \Delta\beta_i, & t = k_0 + 1, ..., T. \end{cases}$$

where $\Delta\beta_i$ is the jump in the slope for each series. In addition, $x_{it} = a_i + \gamma_{2i}f_t + v_{it}$ is correlated with e_{it}:

$$e_{it} = \gamma_{1i}(k_1)f_t + \rho_{e,i}v_{it} + (1 - \rho_e^2)^{1/2}\varepsilon_{it} \quad (27)$$

where $\rho_{e,i}$ denotes the correlation between x_{it} and e_{it}. We also allow a break in the factor loading $\gamma_{1i}(k_1)$ at a different time point $k_1 = [0.7T]$:

$$\gamma_{1i}(k_1) = \begin{cases} \gamma_{1i}, & t = 1, ..., k_1, \\ \gamma_{1i} + \Delta\gamma_{1i}, & t = k_1 + 1, ..., T. \end{cases}$$

We set the intercepts $\alpha_i \sim iidN(1, 1)$, the slopes $\beta_{1i} = 1 + \eta_i, \eta_i \sim iidN(0, 0.04)$ and $\Delta\beta_i \sim iidN(0, 0.04)$. In the process generating x_{it}, $a_i \sim iidN(0.5, 0.5)$, $\gamma_{2i} \sim iidN(0.5, 0.5)$ and $v_{it} \sim iidN(0, 1 - \rho_{vi}^2)$, with $\rho_{vi} = 0.5$. The factor f_t is generated by the stationary process:

$$\begin{aligned} f_t &= \rho_f f_{t-1} + v_{ft}, t = -49, ..., 0, 1, ..., T; \\ \rho_f &= 0.5, v_{ft} \sim iidN(0, 1 - \rho_f^2), f_{-50} = 0. \end{aligned}$$

In the process generating e_{it}, the loadings $\gamma_{1i} \sim iidN(1, 0.2)$, $\Delta\gamma_{1i} \sim iidN(0.5, 0.5)$, $\rho_{e,i} \sim iidU(-0.5, 0.5)$ and $\varepsilon_{it} \sim iidN(0, \sigma_i^2)$ with $\sigma_i^2 \sim iidU(0.5, 1.5)$.

Several experiments are run by modifying the values of the parameters above in Figures A1-A5.

A2.2 Findings

In the error structure (27), there are two sources of endogeneity due to the unobserved factor f_t and the random component v_{it}. To mute the correlation between x_{it} and e_{it} due to f_t, we first consider the simplified case without the factor structure in the errors, or factor loadings $\gamma_{1i} = 0$ in Figure A1. The first panel of Figure A1 reports the histograms of $\hat{k}$ for $T = 20$ and $N = 1, 10, 50, 200$. As in BFK, the distribution of $\hat{k}$ tightens with N in the presence of endogeneity. The frequency of choosing the true value k_0 increases from 6% to 58% when N increases from 1 to 200. The second panel of Figure A1 reports the case of $T = 50$ and shows that the frequency of choosing the true value k_0 improves to almost 80% for $N = 200$. This finding confirms the message delivered in Section 3.1 that the consistency of the OLS estimator of the break point $\hat{k}$ is not affected by endogeneity as N increases.

In Figures A2-A5, we consider the general case (11). For simplicity, we assume that there is no break in factor loading γ_{1i}, i.e., $\Delta\gamma_{1i} = 0$, in Figures A2, A4, A5. As pointed out by Perron and Yamamoto (2015), the break fraction $\tau_0 = k_0/T$ can be consistently estimated by OLS even in the presence of correlation between x_{it} and e_{it} in a time series regression. However, in a panel data setup, the cross-sectional correlation in the errors due to the common f_t could fail to improve the accuracy of the OLS estimator of k_0, as pointed out by Theorem 1A(iii) of Kim (2011) and Figure 7 in BFK. Thus, the transformation (19) using cross-sectional averages of y_{it} and x_{it} is needed to remove f_t before conducting least squares.

The first panel of Figure A2 presents the histograms of the estimated change point $\tilde{k}$

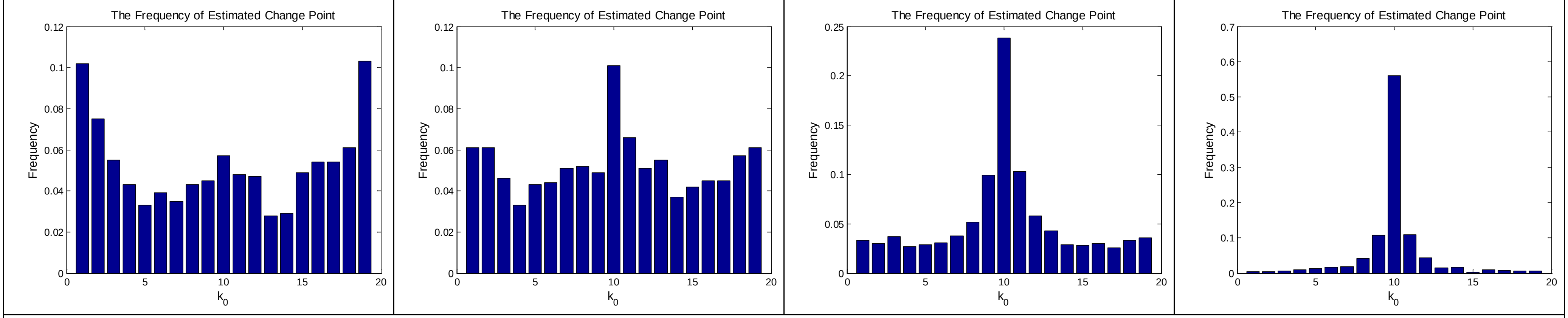
for $T = 50$. It replicates the pattern in Figure A1, showing that after the transformation, the frequency of choosing the true value k_0 increases significantly with N. It confirms Theorem 1 that the distribution of $\tilde{k}$ collapses to k_0 as $N \to \infty$. The second panel of Figure A2 also reports the histograms of the estimated change point $\hat{k}$ without conducting the CCE transformation (19). It indicates that in the presence of common correlated effects, cross-sectional information using multiple series fails to improve the accuracy of the estimated change point.

Figure A3 presents the case when there is a common break in the factor loading $\gamma_{1i}(k_1)$, with $k_1 = [0.7T] > k_0$. Consistent with our theory, $\tilde{k}$, our estimator of the break point in the slope parameters is robust to a break in the error factor loadings γ_{1i}. This holds since f_t is asymptotically removed by the CCE transformation (19). However, as shown in the second panel of Figure A3, the break point in factor loadings could lead to a spurious break in the slope parameters if we ignore the unobserved factors in the errors.

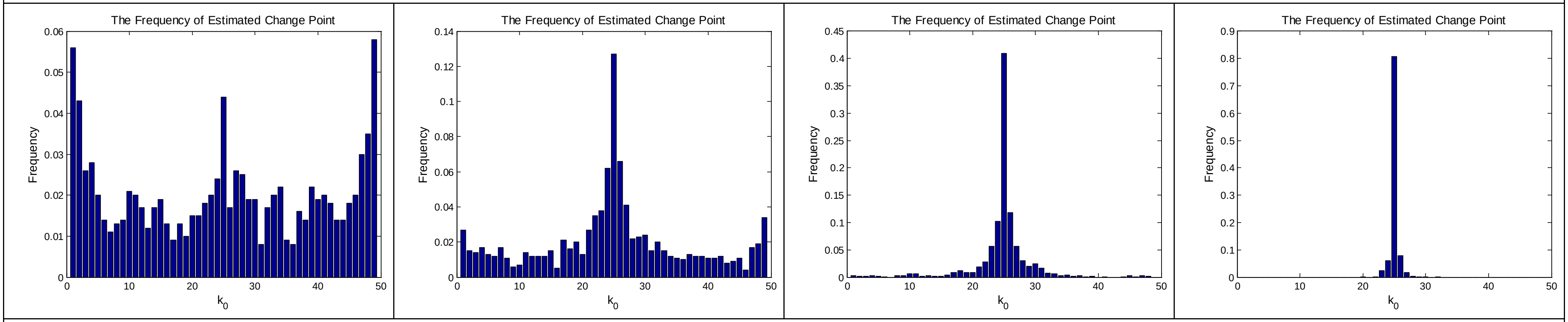
In Figure A4, the cross-sectional dependence is reduced by changing the distribution of γ_{1i} from $iidN(1, 0.2)$ in Figure 2 to $iidN(0.5, 0.2)$. Different from Figure A2, the histograms of $\hat{k}$ in the second panel of Figure A4 show that the OLS estimator of the break point becomes more accurate as N increases even without conducting the transformation (19). However, this is not the case in Figure A5 when we reduce the correlation between x_{it} and e_{it} by changing the distribution of the loading γ_{2i} from $N(0.5, 0.5)$ to $N(0.1, 0.1)$.

Figure A1: Histograms of the OLS estimator $\hat{k}$ in a simplified case without a factor structure in the errors

$N=1$	$N=10$	$N=50$	$N=200$



$T = 20,\ k_0 = 10$



$T = 50, k_0 = 25$

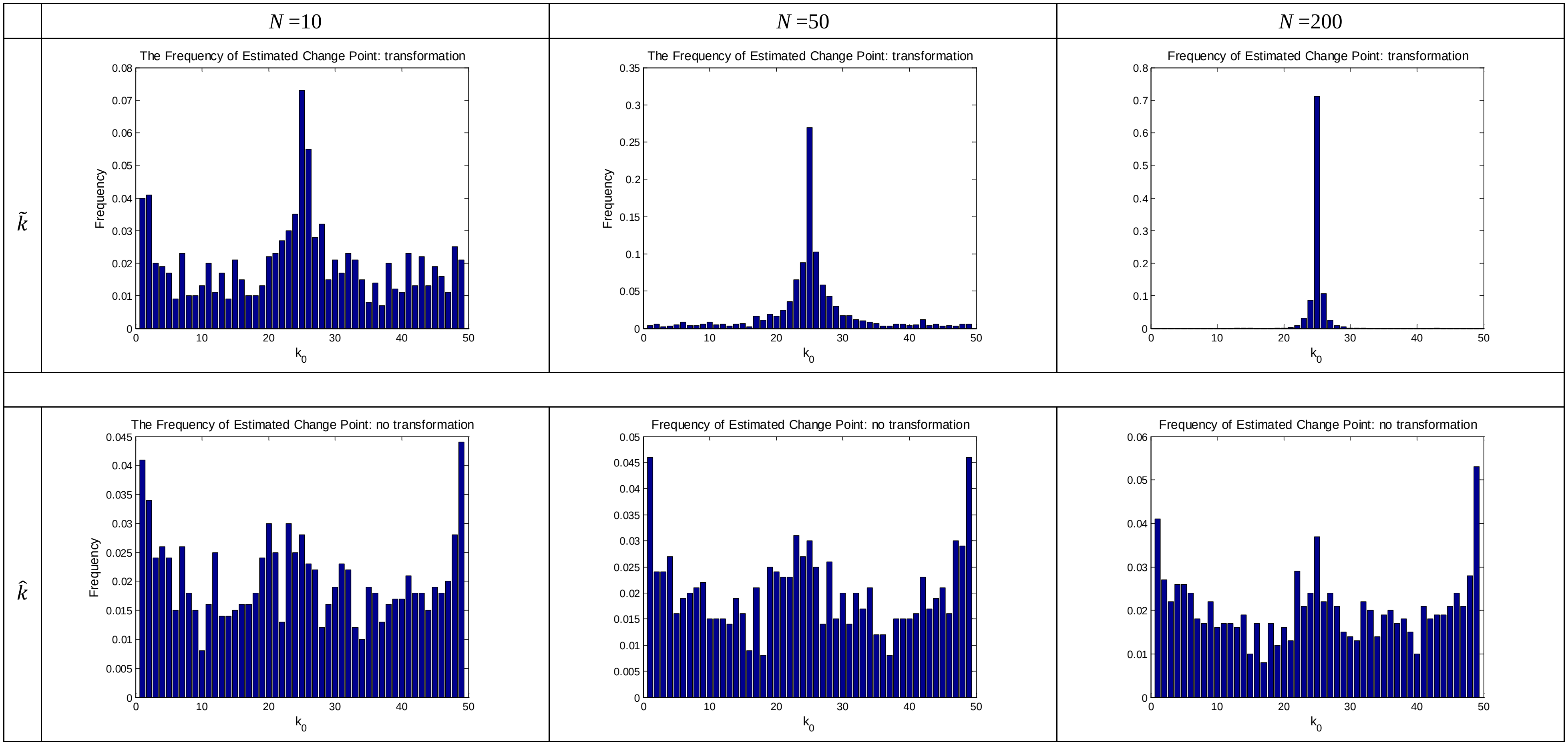
Note: The DGP is a modified design of Model 2 in BFK (2016). In this simplified case, there is no factor structure in the errors. The regressors x_{it} are correlated with e_{it}.

$y_{it} = \alpha_i + \beta_i(k_0)x_{i,t} + e_{it}, i = 1, \dots, N; t = 1, \dots, T$. $\alpha_i \sim iidN(1,1)$, $\beta_i(k_0) = \begin{cases} \beta_{1i}, & t = 1, \dots, k_0, \\ \beta_{2i} = \beta_{1i} + \delta_i, & t = k_0 + 1, \dots, T. \end{cases}$ $k_0 = 0.5T$, $\beta_{1i} \sim iidN(1, 0.04)$, $\delta_i \sim iidN(0, 0.04)$.

$x_{it} = a_i + \gamma_{2i} f_t + v_{it}; e_{it} = \rho_{e,i} v_{it} + \left(1 - \rho_{e,i}^2\right)^{1/2} \varepsilon_{it}$, $\rho_{e,i} \sim iidU(-0.5, 0.5)$.

$f_t = \rho_f f_{t-1} + v_{ft}, t = -49, \dots, 0, 1, \dots T$, $v_{ft} \sim iidN(0, 1 - \rho_f^2)$, $\rho_f = 0.5$, $f_{-50} = 0$. $\varepsilon_{it} \sim iidN(0, \sigma_i^2)$, $\sigma_i^2 \sim iidU(0.5, 1.5)$, $\gamma_{1i} \sim iidN(1, 0.2)$, $\gamma_{2i} \sim iidN(0.5, 0.5)$, $a_i \sim iidN(0.5, 0.5)$, $v_{it} \sim iidN(0, 1 - \rho_{vi}^2)$, $\rho_{vi} = 0.5$. These variables are mutually independent. The replication number is 1000.

Figure A2: Histograms of $\tilde{k}$ and $\hat{k}$ in the general case (T =50)



Note: The DGP is a modified design of Model 2 in BFK (2016). The regressors x_{it} are correlated with e_{it}.

$$y_{it} = \alpha_i + \beta_i(k_0)x_{i,t} + e_{it}, i = 1, \dots, N; t = 1, \dots, T. \ \ \alpha_i \sim iidN(1, 1), \ \ \beta_i(k_0) = \begin{cases} \beta_{1i}, & t = 1, \dots, k_0, \\ \beta_{2i} = \beta_{1i} + \delta_i, & t = k_0 + 1, \dots, T. \end{cases} \ \ k_0 = 0.5T, \ \ \beta_{1i} \sim iidN(1, 0.04), \ \ \delta_i \sim iidN(0, 0.04).$$

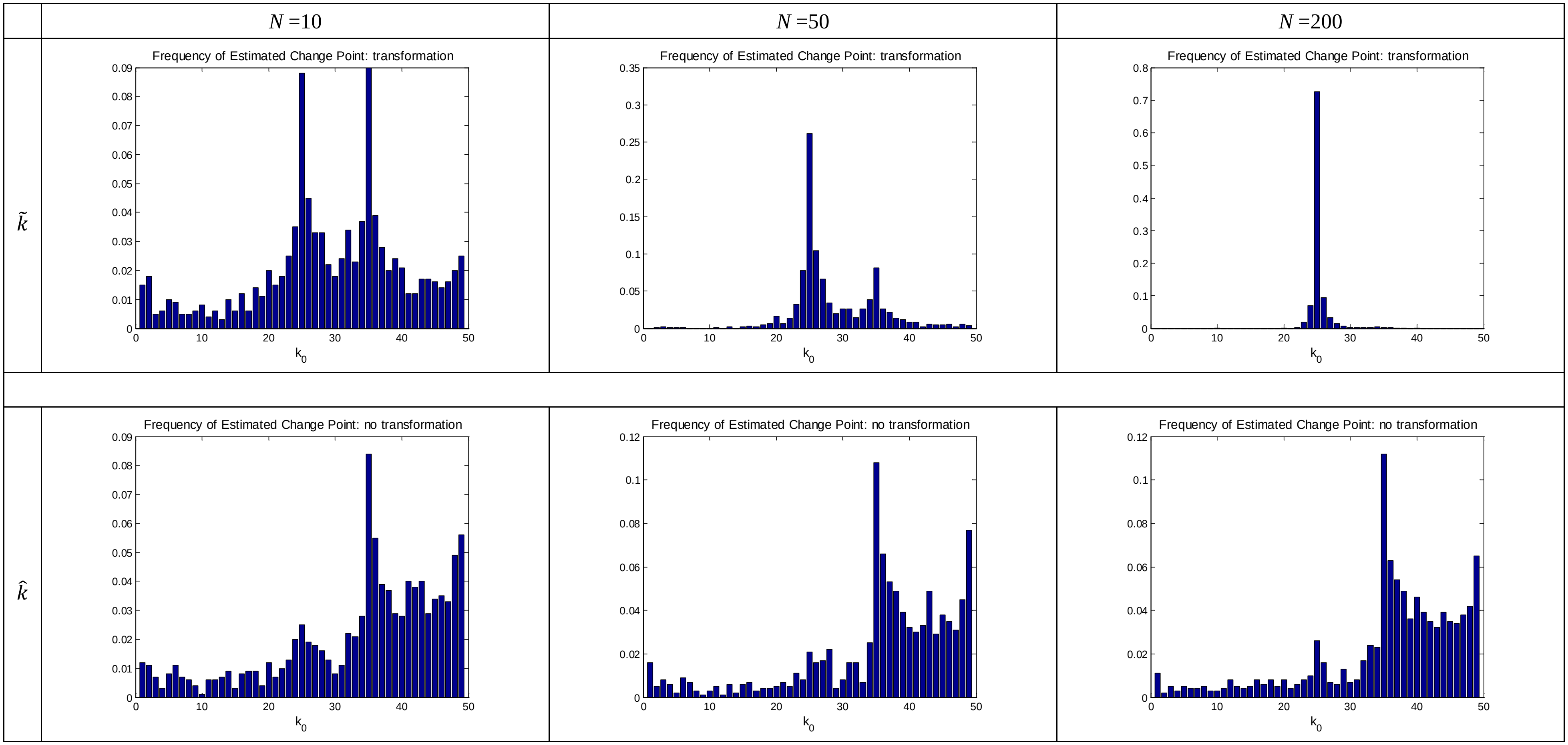
$$x_{it} = a_i + \gamma_{2i} f_t + v_{it}; \ e_{it} = \gamma_{1i} f_t + \rho_{e,i} v_{it_t} + \left(1 - \rho_{e,i}^2\right)^{1/2} \varepsilon_{it}, \ \ \gamma_{1i} \sim iidN(1, 0.2), \ \ \rho_{e,i} \sim iidU(-0.5, 0.5).$$

$f_t = \rho_f f_{t-1} + v_{ft}, t = -49, \dots, 0, 1, \dots T, \ \ v_{ft} \sim iidN(0, 1 - \rho_f^2), \ \ \rho_f = 0.5, \ \ f_{-50} = 0. \ \ \varepsilon_{it} \sim iidN(0, \sigma_i^2), \ \ \sigma_i^2 \sim iidU(0.5, 1.5), \ \ \gamma_{i1} \sim iidN(1, 0.2), \ \ \gamma_{i2} \sim iidN(0.5, 0.5), \ \ a_i \sim iidN(0.5, 0.5),$
$v_{it} \sim iidN(0, 1 - \rho_{vi}^2), \ \ \rho_{vi} = 0.5$. These variables are mutually independent. The replication number is 1000. $T = 50, k_0 = 25$.

$\tilde{k}$: OLS estimator of change point after removing common correlated factors

$\hat{k}$: OLS estimator of change point without removing common correlated factors

Figure A3: Histograms of $\tilde{k}$ and $\hat{k}$ in the general case with a structural change in the error factor loading (T =50)



Note: The DGP is the same as the one in Figure A2 above except that there is a common break in the error factor loadings γ_{1i}

$$e_{it} = \gamma_{1i}(k_1)f_t + \rho_{e,i}v_{it} + (1-\rho_{e,i}^2)^{1/2}\varepsilon_{it}, \qquad \gamma_{1i}(k_1) = \begin{cases} \gamma_{1i}, & t = 1, \dots, k_1, \\ \gamma_{1i} + \Delta\gamma_{1i}, & t = k_1 + 1, \dots, T. \end{cases} \quad k_1 = [0.7T]$$

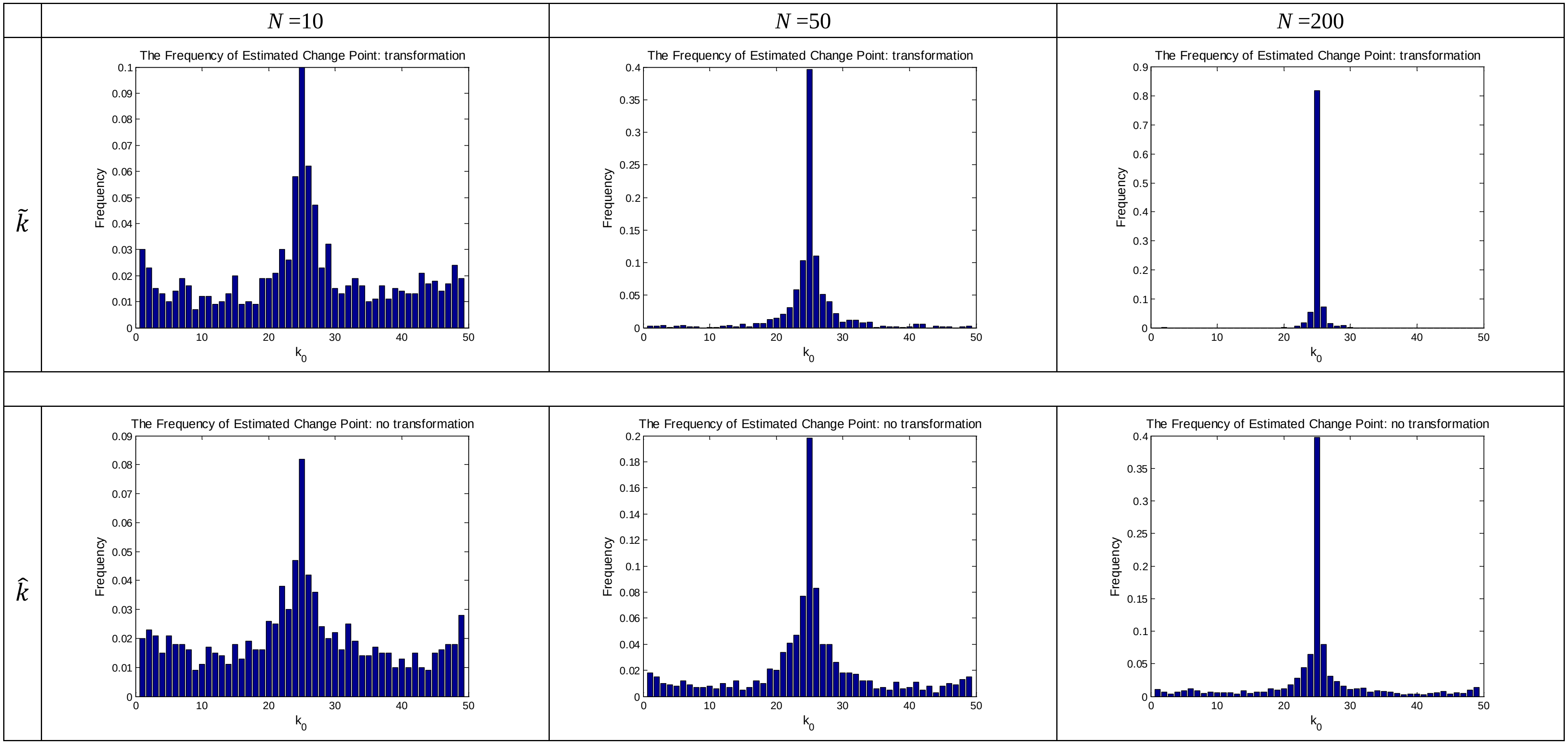
$\gamma_{1i} \sim iidN(1,0.2)$ and $\Delta\gamma_{1i} \sim iidN(0.5,0.5)$

$T = 50, k_0 = 25$

$\tilde{k}$: OLS estimator of change point after removing common correlated factors

$\hat{k}$: OLS estimator of change point without removing common correlated factors

Figure A4: Histograms of $\tilde{k}$ and $\hat{k}$ in the general case with reduced cross-sectional correlation (T =50)



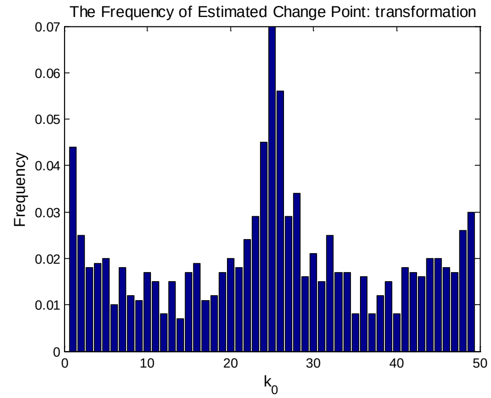
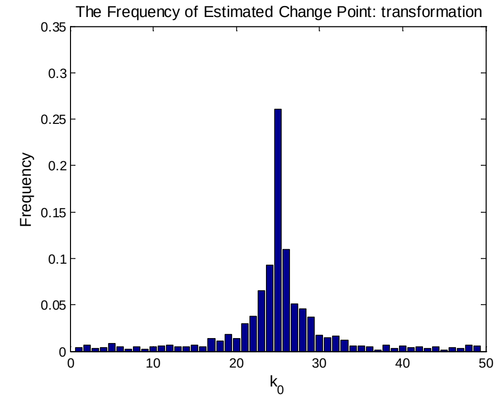
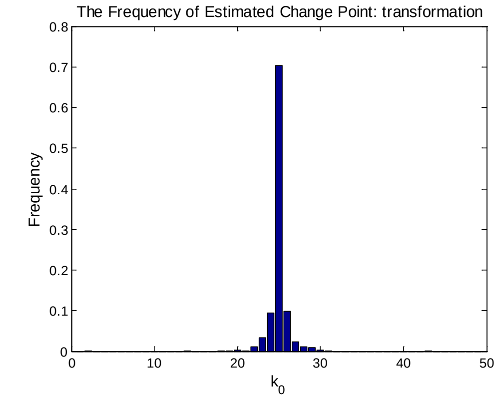
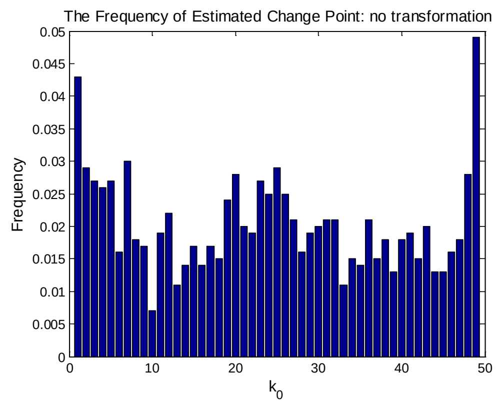
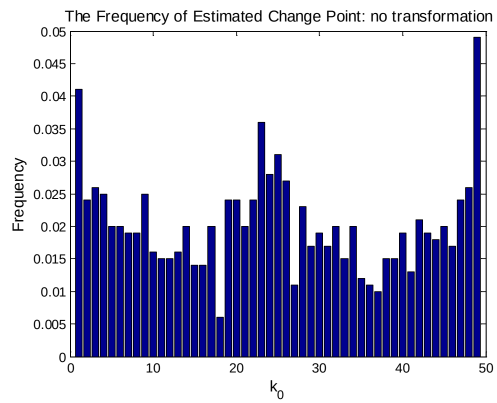
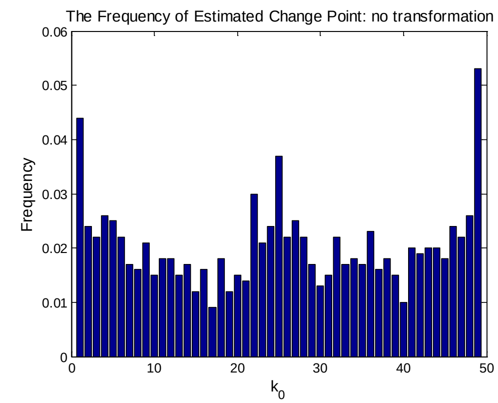
Note: The DGP is the same as the one in Figure A2 above, except for reducing the cross-sectional correlation by changing the distribution of the error factor loading γ_{1i} from $iidN(1, 0.2)$ to $iidN(0.5, 0.2)$

$T = 50, k_0 = 25$

$\tilde{k}$: OLS estimator of change point after removing common correlated factors

$\hat{k}$: OLS estimator of change point without removing common correlated factors

Figure A5: Histograms of $\tilde{k}$ and $\hat{k}$ in the general case with reduced endogeneity (T =50)

	N =10	N =50	N =200
$\tilde{k}$			
$\hat{k}$			

Note: The DGP is the same as the one in Figure A2 above, except for reducing the correlation between $x_{i,t}$ and $e_{i,t}$ by changing the distribution of the loading γ_{2i} from $iidN(0.5, 0.5)$ to $iidN(0.1, 0.1)$

$T = 50, k_0 = 25$

$\tilde{k}$: OLS estimator of change point after removing common correlated factors

$\hat{k}$: OLS estimator of change point without removing common correlated factors