Stein-like Common Correlated Effects Estimation Under Structural Breaks

Shahnaz Parsaeian[†]

Department of Economics, University of Kansas, USA
E-mail: sh.parsaeian@ku.edu

April 12, 2024

Abstract

This paper develops a Stein-like combined estimator for large heterogeneous panel data models under common structural breaks. The model allows for cross-sectional dependence through a general multifactor error structure. By utilizing the common correlated effects (CCE) estimation technique, we propose a Stein-like combined estimator of the CCE full-sample estimator (i.e., estimation using both the pre-break and post-break observations) and the CCE post-break estimator (i.e., estimation using only the post-break sample observations). The proposed Stein-like combined estimator benefits from exploiting the pre-break sample observations. We derive the optimal combination weight by minimizing the asymptotic risk. We show the superiority of the CCE Stein-like combined estimator over the CCE post-break estimator in terms of the asymptotic risk. Further, we establish the asymptotic properties of the CCE mean group Stein-like combined estimator. The finite sample performance of our proposed estimator is investigated using Monte Carlo experiments and an empirical application of predicting the output growth of industrialized countries.

Keywords: *Common correlated effects, Cross-sectional dependence, Heterogeneous panels, Structural breaks.*

JEL Classification: C13, C23, C33

[†]Correspondence to: Shahnaz Parsaeian, Department of Economics, University of Kansas, Lawrence, KS 66045. Email: sh.parsaeian@ku.edu.

1 Introduction

Panel data sets have been increasingly used in economics and statistics, as they provide a flexible way to model variations over both cross-section units and time. Considering structural breaks in panel data is of great importance in empirical economics questions. This is because a structural break is considered as an exogenous shock (such as financial crises or technological progress) which may influence the relationship among economic variables. It is likely that this shock to have impacts on economic variables simultaneously. Recently there has been a growing literature on the detection of changes and common structural breaks, and its associated asymptotic properties in panel data models. The importance of common structural breaks is evident when the global financial or technological shocks affect all markets or firms at the same time.

Estimation of multiple break points in the linear regression model is analyzed in Bai and Perron (1998, 2003) in which they also propose tests for detecting the number of breaks. Bai (2010) studies the asymptotic properties of the break point estimator for the cross sectionally independence panel data model but allowing serially correlation within each individual unit. For break points detection in panel data models, see Kao et al. (2012), Kim (2011, 2014), Qian and Su (2016), Li et al. (2016), Baltagi et al. (2016, 2019), Baltagi et al. (2017), Perron et al. (2020), Okui and Wang (2021), and Lumsdaine et al. (2023), among others. While the issues related to the detection of break points have drawn a lot of attention in both econometrics and statistics, relatively small attention has been paid to improving estimation of unknown slope coefficients under structural breaks. It is known that model averaging and combined estimation techniques can improve estimations and consequently forecasts under model uncertainty. Since the structural break models can be viewed as the model uncertainty, one can benefit from combined estimation techniques, see for example Stock and Watson (2004), Hansen (2009), Lee et al. (2022, 2022a), and Parsaeian (2023) who considers a weighted average estimator in a seemingly unrelated regression model in which the cross-section dimension is small while the time series dimension is large.

Pesaran (2006) develops a common correlated effects (CCE) estimator that filters out the unobserved common factors by means of the cross-sectional averages of the dependent variable and the individual-specific regressors as the cross-section dimension tends to infinity. In this paper, we develop a Stein-like combined estimator for a large heterogeneous panel data model under structural breaks with a general multifactor error structure which is due to unobservable common factors. The break point is common for all individual units. By utilizing the CCE estimation method, we introduce a CCE Stein-like combined estimator which can improve estimation of the slope coefficients in the sense of asymptotic risk. Our proposed estimator is a combination of two estimators: the CCE full sample estimator and the CCE post-break estimator. The CCE full-sample estimator ignores the break point and uses the full-sample of observations to estimate the slope coefficients. Thus, it is the most efficient estimator while it is biased. The CCE post-break estimator is the consistent estimator but less efficient since it only uses the observations after the break point. Therefore, combining these two estimators balances the trade-off between the bias and variance efficiency. The combination weight is inversely related to a weighted quadratic loss function, and takes the form of the James-Stein weight, cf. James and Stein (1961).[1] We establish the asymptotic risk of the proposed Stein-like combined estimator, and show that its asymptotic risk is less than that of the CCE post-break estimator, which is the common estimation method of the slope coefficients under structural breaks. Furthermore, we develop the CCE mean group (referred to as CCEMG) Stein-like combined estimator, which is a simple average of the individual CCE estimators. We obtain the asymptotic distribution, and the asymptotic risk of the CCEMG Stein-like combined estimator, and show that it has a lower asymptotic risk than that of the CCEMG post-break estimator.

It is a well-established idea in the panel data models to use the averaging estimates of the individual cross-section units to estimate the common mean, see for example chapter 6 of Hsiao and Pesaran (2008), and chapter 28 of Pesaran (2015). Since the CCEMG estimators estimate the common mean in the panel, it is fruitful to compare these estimators. Specifically when the number of individual units (N) in panel is large, it is hard and not quite informative to compare the individual CCE estimators. We therefore undertake Monte Carlo simulation studies to evaluate the finite sample forecasting performance of the proposed CCEMG Stein-like combined estimator, the CCEMG post-break estimator, and the CCEMG full-sample estimator. We further compare the performance of the CCEMG estimators in an empirical study of forecasting the output growth rate of industrialized countries.

[1]See also Hansen(2016, 2017) and Mehrabani and Ullah (2020) for the use of the Stein-like estimator in different contexts.

The rest of the paper is organized as follows. Section 2 introduces a heterogenous panel data model which allows for structural changes and multifactor error structure, and develops the CCE Stein-like combined estimator. For simplicity, we discuss the model under a single break, which simplifies the idea. Although, the generalization of the method to multiple break points is straightforward. Section 3 establishes the asymptotic distribution and asymptotic risk of the proposed CCE Stein-like combined estimator. Section 4 develops the asymptotic risk of the CCEMG Stein-like combined estimator. Section 5 reports Monte Carlo simulation. Section 6 presents empirical analysis. Section 7 concludes. All the proofs are given in the Appendix.

Notation: For an $m \times n$ real matrix A, we denote its transpose as A'. When $m = n$, $\lambda_{max}(A)$ denotes the maximum eigenvalues of A. Let $tr(A)$ be the trace of a square matrix A. I_m and $\mathbf{0}_{m\times 1}$ denote $m \times m$ identity matrix and $m \times 1$ vector of zeros. The operators $\xrightarrow{d}$, and $\xrightarrow{p}$ denote convergence in distribution, and probability, respectively. $a_n = O(b_n)$ states that the deterministic sequence a_n is at most of order b_n, $x_n = O_p(y_n)$ states that the vector of random variables, x_n, is at most of order y_n in probability. Joint convergence of N and T will be denoted by $(N, T) \to \infty$. Restrictions on the relative rates of convergence of N and T will be specified separately.

2 The model

Consider the following heterogeneous panel data model with a multifactor error structure, and a common structural break at time T_1 as

$$y_{it} = x_{it}'\beta_i(T_1) + e_{it}, \tag{2.1}$$

$$e_{it} = \gamma_i' f_t + \epsilon_{it}, \tag{2.2}$$

where $i = 1, \ldots, N, t = 1, \ldots, T$, x_{it} is a $k \times 1$ vector of regressors, e_{it} is the error terms which are cross-sectionally correlated and modelled by a multifactor structure in (2.2), f_t is an $m \times 1$ vector of unobserved factors, γ_i is the corresponding loading vector, and ϵ_{it} are the idiosyncratic errors independent of x_{it}. The unobserved factors f_t could be correlated with x_{it} as

$$x_{it} = \Gamma_i' f_t + v_{it}, \tag{2.3}$$

where Γ_i is an $m \times k$ factor loading matrix, and v_{it} is a $k \times 1$ vector of zero mean disturbances assumed to follow a general covariance stationary process. The vector of coefficients, $\beta_i(T_1)$, are subject to the structural break across individuals at time T_1 such that

$$\beta_i(T_1) = \begin{cases} \beta_{i(1)} & \text{for} \quad t = 1, \ldots, T_1, \\ \beta_{i(2)} & \text{for} \quad t = T_1 + 1, \ldots, T. \end{cases} \tag{2.4}$$

Let $Y_i = \left(Y'_{i(1)}, Y'_{i(2)}\right)'$ be a vector of $T \times 1$ dependent variable with $Y_{i(1)} = \left(y_{i1}, \ldots, y_{iT_1}\right)'$ and $Y_{i(2)} = \left(y_{iT_1+1}, \ldots, y_{iT}\right)'$, $X_i = \left(X'_{i(1)}, X'_{i(2)}\right)'$ be a $T \times k$ matrix of regressors with $X_{i(1)} = \left(x_{i1}, \ldots, x_{iT_1}\right)'$ and $X_{i(2)} = \left(x_{iT_1+1}, \ldots, x_{iT}\right)'$, and $e_i = \left(e'_{i(1)}, e'_{i(2)}\right)'$ be a $T \times 1$ vector of error terms with $e_{i(1)} = \left(e_{i1}, \ldots, e_{iT_1}\right)'$ and $e_{i(2)} = \left(e_{iT_1+1}, \ldots, e_{iT}\right)'$ denote the stacked data and errors for individuals, $i = 1, \ldots, N$, over the time period observed. Let $b_1 \equiv T_1/T \in (0, 1)$. Thus, the model in (2.1) can be written as

$$\begin{cases} Y_{i(1)} = X_{i(1)}\beta_{i(1)} + e_{i(1)}, \\ Y_{i(2)} = X_{i(2)}\beta_{i(2)} + e_{i(2)}. \end{cases} \tag{2.5}$$

Remark 1: *We note that one can also allow a common structural break in the error factor loadings at the same or different break point, T_1. As shown in Section 3, the CCE method filters out the unobserved common factors by means of the cross-sectional averages of the dependent variable and the individual-specific regressors. Thus, a common break in loadings does not affect the consistency of the break point estimator and the slope parameters estimator asymptotically, see Baltagi et al. (2019).*

Remark 2: *We note that the method of break point estimation is based on the least-squares method. That is, for the given break point T_1, the associated least-squares estimates of the slope coefficients are obtained by minimizing the sum of squared residuals.By substituting these estimated slope coefficients in the objective function, the estimated break point is obtained. Theorems 1-2 in Baltagi et al. (2016) show that, in large panels, $(N, T) \to \infty$, the break point T_1 can be consistently estimated, i.e., $\widehat{T}_1 - T_1 = o_p(1)$, which implies that compared to a time series setting,* *the cross-sectional observations improve the accuracy of the estimated break point.*[2]

2.1 CCE Stein-like combined estimator

We propose the CCE Stein-like combined estimator, which is a combination of the CCE full-sample estimator and the CCE post-break estimator, under common correlated effect models.

Since the ultimate interest is on forecasting, the parameters of interest are $\beta_{i(2)}$. We propose the CCE Stein-like combined estimator for $\beta_{i(2)}$ as

$$\widetilde{\beta}_i = \alpha_{NT}\ \widetilde{\beta}_{i,Full} + (1-\alpha_{NT})\ \widetilde{\beta}_{i(2)}, \tag{2.6}$$

where $\widetilde{\beta}_i$ is the CCE Stein-like combined estimator, $\widetilde{\beta}_{i,Full}$ is the CCE full-sample estimator which uses all observations in the full sample and therefore ignores the break across individuals, and $\widetilde{\beta}_{i(2)}$ is the CCE post-break estimator which uses the observations in the post-break sample, $t > T_1$, for each individual. The combination weight, α_{NT}, depends on the sample. For notational simplicity, we use $\alpha \equiv \alpha_{NT}$. Specifically, the combination weight depends on a weighted squared loss function, and is defined as

$$\alpha = \begin{cases} \frac{\tau}{\mathcal{D}_T} & \text{if } \mathcal{D}_T \geq \tau \\ 1 & \text{if } \mathcal{D}_T < \tau, \end{cases} \tag{2.7}$$

where τ is a positive parameter that controls the degree of shrinkage, and $\mathcal{D}_T$ is the weighted quadratic loss function which measure the distance between the CCE full-sample and the CCE post-break estimators and is equal to

$$\mathcal{D}_T = T\big(\widetilde{\beta}_{i(2)} - \widetilde{\beta}_{i,Full}\big)' W \big(\widetilde{\beta}_{i(2)} - \widetilde{\beta}_{i,Full}\big), \tag{2.8}$$

in which W is a positive definite weight matrix. For example, when $W = \big(\widetilde{V}_{i(2)} - \widetilde{V}_{i,Full}\big)^{-1}$, where $\widetilde{V}_{i,Full}$ and $\widetilde{V}_{i(2)}$ are the consistent estimators of the asymptotic variances of the CCE full-sample and the CCE post-break estimators, then $\mathcal{D}_T$ becomes the Hausman statistics. This is a suitable choice for W since the Hausman statistics is the ratio of the bias over variance efficiency. Thus, using this weight matrix in the combination weight helps to give suitable weight to each of the

[2] In a time series model, only the estimated break fraction, $\widehat{b}_1$, can be consistently estimated and not the estimated break point. In other words, in time series models $\widehat{T}_1 - T_1 = O_p(1)$ for large T, see Bai and Perron (1998).

CCE full-sample estimator and the CCE post break estimator. Alternatively, when $W = I_k$, then $\mathcal{D}_T$ is unweighted quadratic loss, which is a suitable choice for W when the slope parameters are of equal importance.

We note that the combination weight, α, is inversely proportional to the weighted quadratic loss function, and the degree of shrinkage depends on the ratio of $\tau/\mathcal{D}_T$. Large values of $\mathcal{D}_T$ (or when $\mathcal{D}_T > \tau$) indicate large break sizes. In this case, the combined estimator assigns a large weight to the CCE post-break estimator and a small weight to the CCE full-sample estimator which is largely biased under large break sizes. However, small values of $\mathcal{D}_T$ indicate small break sizes. In this case, the combined estimator assigns a large weight to the CCE full-sample estimator to gain from its efficiency, and a small weight to the CCE post-break estimator. Therefore, the proposed CCE Stein-like combined estimator balances the trade-off between the bias and variance efficiency, and assigns appropriate weights to each of the CCE full-sample and the CCE post-break estimators based on the break sizes.

3 Common correlated effects model

In this section, we derive the asymptotic distributions of the full-sample, the post-break and the Stein-like combined estimators, for each $i = \{1, \ldots, N\}$, by considering the common correlated effects f_t in the errors and regressors, as defined in (2.2) and (2.3). We call them the CCE full-sample estimator, the CCE post-break estimator, and the CCE Stein-like combined estimator, respectively. The asymptotic distribution theory is developed under a local asymptotic framework under which the CCE Stein-like combined estimator has a nondegenerate asymptotic distribution. Further, we derive the asymptotic risk for the CCE Stein-like combined estimator.

Assumption 1: *The break size in the coefficients is local to zero, i.e., for $i = 1, \ldots, N$,*

$$\beta_{i(2)} - \beta_{i(1)} = \frac{\delta_i}{\sqrt{T}}, \tag{3.1}$$

where $\delta_i \in \mathbb{R}$. Assumption 1 indicates that for any fixed δ_i, the break size shrinks as the sample size T increases. We note that the break size in the slope coefficients (δ_i) can be different across individuals.

Assumption 2: *The disturbances* ϵ_{it}*, for* $i = 1, ..., N$*, are cross-sectionally independent, with a well-defined variance,* $Var(\epsilon_{it}) = \sigma_i^2 < \infty$*. Besides, for each series* i*,* ϵ_{it} *is serially uncorrelated, and independent of* x_{it} *for all* i *and* t*.*

Assumption 3: *Common factors* f_t *are covariance stationary with absolute summable autocovariances, distributed independently of errors* ϵ_{is} *and* v_{is} *for all* i, s, t*.*

Assumption 4: ϵ_{is} *and* v_{jt} *are independent for all* i, j, s, t*, and* $Var(v_{it}) = \Sigma_{i,v} < \infty$*.*

Assumption 5: *Factor loadings* γ_i *and* Γ_i *are i.i.d. across* i*, and independent of* ϵ_{jt}*,* v_{jt} *and* f_t *for all* i, j, t*. Also, assume that* $\gamma_i = \gamma + \varrho_i$*,* $\varrho_i \sim i.i.d.\ (0, \Omega_\varrho)$*, and* $\Gamma_i = \Gamma + \xi_i$*,* $\xi_i \sim i.i.d.\ (0, \Omega_\xi)$*, where the means,* γ *and* Γ*, are nonzero and the variances,* Ω_ϱ *and* Ω_ξ*, are well-defined.*

Assumptions 3-5, are the same as Assumptions 8-10 of Baltagi et al. (2016), or similar to Assumptions 1-3 of Pesaran (2006).

Because of the unobserved common factor effect f_t in the error terms and its correlation with x_{it}, the usual ordinary least squares (OLS) estimation method is inconsistent. Pesaran (2006) proposes to use the cross-sectional averages of y_{it} and x_{it} as proxies for the unobserved f_t. We define the $(k+1) \times 1$ vector of w_{it} as

$$w_{it} = \begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = C_i'(T_1) f_t + u_{it}(T_1), \tag{3.2}$$

where $\underbrace{C_i(T_1)}_{m \times (k+1)} = \left(\gamma_i, \Gamma_i\right) \begin{pmatrix} 1 & 0_{1 \times k} \\ \beta_i(T_1) & I_k \end{pmatrix} = \begin{cases} C_{i(1)} = \left(\gamma_i + \Gamma_i \beta_{i(1)}, \Gamma_i\right), & \text{for} \quad t = 1, \dots, T_1, \\ C_{i(2)} = \left(\gamma_i + \Gamma_i \beta_{i(2)}, \Gamma_i\right), & \text{for} \quad t = T_1 + 1, \dots, T, \end{cases}$

and $\underbrace{u_{it}(T_1)}_{(k+1) \times 1} = \begin{pmatrix} \epsilon_{it} + v_{it}' \beta_i(T_1) \\ v_{it} \end{pmatrix}$.

Let $\bar{w}_t = \sum_{i=1}^N \theta_i w_{it}$ be the cross-sectional averages of w_{it} using weights θ_i, for $i = 1, \dots, N$, where the weights satisfy $\theta_i = O(\frac{1}{N}), \sum_{i=1}^N \theta_i = 1$ and $\sum_{i=1}^N |\theta_i| < \infty$. Therefore,

$$\bar{w}_t = \bar{C}'(T_1) f_t + \bar{u}_t(T_1), \tag{3.3}$$

where $\underbrace{\bar{C}(T_1)}_{m\times(k+1)} = \sum_{i=1}^{N}\theta_i C_i(T_1) = \begin{cases} \bar{C}_{(1)} = \sum_{i=1}^{N}\theta_i C_{i(1)} & \text{for} \quad t = 1,\dots,T_1, \\ \bar{C}_{(2)} = \sum_{i=1}^{N}\theta_i C_{i(2)} & \text{for} \quad t = T_1+1,\dots,T, \end{cases}$

and $\underbrace{\bar{u}_t(T_1)}_{(k+1)\times 1} = \sum_{i=1}^{N}\theta_i u_{it}(T_1) = \begin{cases} \begin{pmatrix} \bar{\epsilon}_t + \sum_{i=1}^{N}\theta_i v_{it}'\beta_{i(1)} \\ \bar{v}_t \end{pmatrix}, & \text{for} \quad t = 1,\dots,T_1, \\ \begin{pmatrix} \bar{\epsilon}_t + \sum_{i=1}^{N}\theta_i v_{it}'\beta_{i(2)} \\ \bar{v}_t \end{pmatrix}, & \text{for} \quad t = T_1+1,\dots,T. \end{cases}$

Assumption 6: $Rank(\bar{C}_{(1)}) = Rank(\bar{C}_{(2)}) = m \leq k+1.$

Assumption 6 indicates that the number of common factors cannot be larger than the number of observable used in estimation. Under Assumption 6, $\bar{C}(T_1)$ is of full rank. Thus, f_t can be written as

$$f_t = \left(\bar{C}(T_1)\bar{C}(T_1)'\right)^{-1} \bar{C}(T_1)\left(\bar{w}_t - \bar{u}_t(T_1)\right). \tag{3.4}$$

As shown in Lemma 1 in Pesaran (2006), as $N \to \infty$, the cross-sectional averages of the errors, $\bar{\epsilon}_t$ and $\bar{v}_t$, disappear in both regimes. Therefore,

$$f_t - \left(\bar{C}(T_1)\bar{C}(T_1)'\right)^{-1} \bar{C}(T_1)\bar{w}_t \xrightarrow{p} 0. \tag{3.5}$$

This suggests using $\bar{w}_t$ as observable proxies for f_t. We note that the model, presented in (2.1)-(2.2), for each $i = \{1,\dots,N\}$, is given by

$$\begin{cases} Y_{i(1)} = X_{i(1)}\beta_{i(1)} + F_{(1)}\gamma_i + \epsilon_{i(1)} \\ Y_{i(2)} = X_{i(2)}\beta_{i(2)} + F_{(2)}\gamma_i + \epsilon_{i(2)}, \end{cases} \tag{3.6}$$

where $F_{(1)} = \left(f_1,\dots,f_{T_1}\right)'$ is the $T_1 \times m$ matrix of the unobserved factors in the first regime, and $F_{(2)} = \left(f_{T_1+1},\dots,f_T\right)'$ is the $(T-T_1)\times m$ matrix of the unobserved factors in the second regime. Let $\bar{W}_{(1)} = \left(\bar{w}_1,\dots,\bar{w}_{T_1}\right)'$ be a matrix of $T_1 \times (k+1)$, and $\bar{W}_{(2)} = \left(\bar{w}_{T_1+1},\dots,\bar{w}_T\right)'$ be a matrix of $(T-T_1)\times(k+1)$. Thus, their corresponding orthogonal projection matrices is defined as

$M_{w_{(1)}} = I_{T_1} - \bar{W}_{(1)}\big(\bar{W}'_{(1)}\bar{W}_{(1)}\big)^{-1}\bar{W}'_{(1)}$ and $M_{w_{(2)}} = I_{T-T_1} - \bar{W}_{(2)}\big(\bar{W}'_{(2)}\bar{W}_{(2)}\big)^{-1}\bar{W}'_{(2)}$. Pre-multiplying each regime in (3.6) by $M_{w_{(1)}}$ and $M_{w_{(2)}}$, respectively, we get

$$\begin{cases} \widetilde{Y}_{i(1)} = \widetilde{X}_{i(1)}\beta_{i(1)} + M_{w_{(1)}}F_{(1)}\gamma_i + \widetilde{\epsilon}_{i(1)} = \widetilde{X}_{i(1)}\beta_{i(1)} + \widetilde{\epsilon}^*_{i(1)} \\ \widetilde{Y}_{i(2)} = \widetilde{X}_{i(2)}\beta_{i(2)} + M_{w_{(2)}}F_{(2)}\gamma_i + \widetilde{\epsilon}_{i(2)} = \widetilde{X}_{i(2)}\beta_{i(2)} + \widetilde{\epsilon}^*_{i(2)}, \end{cases} \tag{3.7}$$

where $\widetilde{Y}_{i(1)} = M_{w_{(1)}}Y_{i(1)}$, $\widetilde{X}_{i(1)} = M_{w_{(1)}}X_{i(1)}$, and $\widetilde{\epsilon}_{i(1)} = M_{w_{(1)}}\epsilon_{i(1)}$. Similarly, we can define the transformed data in the second regime as $\widetilde{Y}_{i(2)} = M_{w_{(2)}}Y_{i(2)}$, $\widetilde{X}_{i(2)} = M_{w_{(2)}}X_{i(2)}$, and $\widetilde{\epsilon}_{i(2)} = M_{w_{(2)}}\epsilon_{i(2)}$. Also, $\widetilde{\epsilon}^*_{i(1)} = M_{w_{(1)}}F_{(1)}\gamma_i + \widetilde{\epsilon}_{i(1)}$, and $\widetilde{\epsilon}^*_{i(2)} = M_{w_{(2)}}F_{(2)}\gamma_i + \widetilde{\epsilon}_{i(2)}$. As shown in Appendix A.1, the order of each elements of $M_{w_{(1)}}F_{(1)}\gamma_i = O_p(\frac{1}{\sqrt{N}})$, and $M_{w_{(2)}}F_{(2)}\gamma_i = O_p(\frac{1}{\sqrt{N}})$. Thus, the order vanishes as $(N,T) \to \infty$. This implies that asymptotically $\widetilde{\epsilon}^*_{i(1)}$ and $\widetilde{\epsilon}^*_{i(2)}$ can be treated as $\widetilde{\epsilon}_{i(1)}$ and $\widetilde{\epsilon}_{i(2)}$, respectively. In Theorem 3.1, we derive the asymptotic distribution and the asymptotic risk for the proposed CCE Stein-like combined estimator.

Assumption 7: *For $i = 1, \ldots, N$, the matrices $\widetilde{X}'_i\widetilde{X}_i/T$, $\widetilde{X}'_{i(1)}\widetilde{X}_{i(1)}/T_1$, and $\widetilde{X}'_{i(2)}\widetilde{X}_{i(2)}/(T-T_1)$ are non-singular and converge in probability to some non-random positive definite matrices.*

Theorem 3.1: *Under Assumptions 1-7, when $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$, the joint asymptotic distribution of the CCE full-sample estimator and the CCE post-break estimator, for each $i = \{1, \ldots, N\}$, is*

$$\sqrt{T}\begin{bmatrix} \widetilde{\beta}_{i,Full} - \beta_{i(2)} \\ \widetilde{\beta}_{i(2)} - \beta_{i(2)} \end{bmatrix} \xrightarrow{d} V_i^{1/2}Z_i, \tag{3.8}$$

where $Z_i \sim N\big(\eta_i, I_{2k}\big)$, $\eta_i = V_i^{-1/2}\begin{bmatrix} b_1\Sigma_i^{-1}\Sigma_{i(1)}\delta_i \\ 0_{k\times 1} \end{bmatrix}$, and $V_i = \begin{bmatrix} V_{i,Full} & V_{i,Full} \\ V_{i,Full} & V_{i(2)} \end{bmatrix}$,

with $V_{i,Full} \equiv \underset{T\to\infty}{plim}\ \sigma_i^2\big(\frac{\widetilde{X}'_i\widetilde{X}_i}{T}\big)^{-1}$, $V_{i(2)} \equiv \underset{T\to\infty}{plim}\ \frac{1}{1-b_1}\sigma_i^2\big(\frac{\widetilde{X}'_{i(2)}\widetilde{X}_{i(2)}}{T-T_1}\big)^{-1}$, $\Sigma_i^{-1} \equiv \underset{T\to\infty}{plim}\ \big(\frac{\widetilde{X}'_i\widetilde{X}_i}{T}\big)^{-1}$, and $\Sigma_{i(1)} \equiv \underset{T\to\infty}{plim}\big(\frac{\widetilde{X}_i(1)'\widetilde{X}_i(1)}{T_1}\big)$. Besides, the asymptotic distribution of the Hausman statistic is

$$
\begin{aligned}
\mathcal{D}_T &= T\big(\widetilde{\beta}_{i(2)} - \widetilde{\beta}_{i,Full}\big)'\big(\widetilde{V}_{i(2)} - \widetilde{V}_{i,Full}\big)^{-1}\big(\widetilde{\beta}_{i(2)} - \widetilde{\beta}_{i,Full}\big) \\
&\xrightarrow{d} Z_i' V_i^{1/2}\ G\big(V_{i(2)} - V_{i,Full}\big)^{-1} G'\ V_i^{1/2} Z_i \\
&\equiv Z_i' M_i Z_i,
\end{aligned} \tag{3.9}
$$

where $G = \big(-I_k \quad I_k\big)'$ *and* $M_i \equiv V_i^{1/2}\ G\big(V_{i(2)} - V_{i,Full}\big)^{-1} G'\ V_i^{1/2}$ *is an idempotent matrix with rank* k*. Finally, the asymptotic distribution of the CCE Stein-like combined estimator is*

$$
\begin{aligned}
\sqrt{T}\big(\widetilde{\beta}_i - \beta_{i(2)}\big) &= \sqrt{T}\big(\widetilde{\beta}_{i(2)} - \beta_{i(2)}\big) - \alpha\sqrt{T}\big(\widetilde{\beta}_{i(2)} - \widetilde{\beta}_{i,Full}\big) \\
&\xrightarrow{d} G_2'\ V_i^{1/2} Z_i - \left(\frac{\tau}{Z_i' M_i Z_i}\right)_1 G'\ V_i^{1/2} Z_i,
\end{aligned} \tag{3.10}
$$

where $G_2 = \big(0 \quad I_k\big)'$ *and* $(a)_1 = min[1, a]$.

Proof: See Appendix A.1. ■

Theorem 3.1 shows that the joint asymptotic distribution of the CCE full-sample estimator and the CCE post-break estimator is normally distributed. The Hausman statistic has an asymptotic non-central chi-square distribution. Furthermore, the asymptotic distribution of the CCE Stein-like combined estimator is a nonlinear function of normal random vector Z_i.

3.1 Asymptotic risk for the CCE estimator

In this section, we derive the asymptotic risk of the proposed estimator by using the results of Theorem 3.1. When an estimator has an asymptotic distribution, $\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} \xi$, we define the asymptotic risk of the estimator as $\rho(\widehat{\beta}, \mathbb{W}) = \mathbb{E}(\xi' \mathbb{W} \xi)$, where $\mathbb{W}$ is a positive definite weight matrix, see Lehmann and Casella (1998). Utilizing the asymptotic distribution of the CCE Stein-like combined estimator in (3.10), we obtain the asymptotic risk for this estimator for any positive definite choice of weight matrix $\mathbb{W}$. We minimize the asymptotic risk to derive the optimal combination weight, α. Theorem 3.2 shows the asymptotic risk of the CCE Stein-like combined estimator when $\mathbb{W} = \big(V_{i(2)} - V_{i,Full}\big)^{-1}$, which is the inverse of the difference of the asymptotic variances of the CCE post-break and the CCE full-sample estimators. This choice of weight greatly simplifies the calculations. The asymptotic risk of the CCE Stein-like combined estimator for any user-specific positive definite choice of $\mathbb{W}$ is available in the Appendix A.2.

Theorem 3.2: *Under Assumptions 1-7, when $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$, the asymptotic risk of the CCE Stein-like combined estimator, for each $i = \{1,\ldots,N\}$, is*

$$\rho\big(\widetilde{\beta}_i, \mathbb{W}\big) = \rho\big(\widetilde{\beta}_{i(2)}, \mathbb{W}\big) - \tau\big(2 - \frac{\tau}{k-2}\big)\Big[e^{-\mu_i}{}_1F_1\big(\frac{k}{2}-1; \frac{k}{2}; \mu_i\big)\Big], \tag{3.11}$$

provided $k > 2$, where ${}_1F_1(.;.;.)$ is the confluent hypergeometric function defined as ${}_1F_1(a;b;\mu_i) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n}\frac{\mu_i^n}{n!}$, where $(a)_n = a(a+1)\ldots(a+n-1)$, $(a)_0 = 1$, and $\mu_i = \eta_i' M_i \eta_i/2$ is the non-centrality parameter. *Also,* $\rho\big(\widetilde{\beta}_{i(2)}, \mathbb{W}\big) = tr(\mathbb{W}\, V_{i(2)})$.

Proof: See Appendix A.2. ■

Theorem 3.2 shows that the asymptotic risk of the CCE Stein-like combined estimator is less than the asymptotic risk of the CCE post-break estimator if $k > 2$, meaning that as long as the number of regressors are greater than two, the CCE Stein-like combined estimator outperforms the CCE post-break estimator in term of the asymptotic risk.

Using the results presented in Theorem 3.2, we derive the optimal value for the shrinkage parameter, denoted by τ^*, as $\tau^* = k-2$, which is positive so long as $k > 2$. By substituting the optimal value of the shrinkage parameter into the asymptotic risk formula in (3.11), we obtain the asymptotic risk of the CCE Stein-like combined estimator. Corollary 3.2.1 summarizes the results.

Corollary 3.2.1: *Under Assumptions 1-7, when $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$, if $0 \le \tau \le 2(k-2)$, the asymptotic risk for the CCE Stein-like combined estimator, for each $i = \{1,\ldots,N\}$, is*

$$\rho\big(\widetilde{\beta}_i, \mathbb{W}\big) = \rho\big(\widetilde{\beta}_{i(2)}, \mathbb{W}\big) - (k-2)\Big[e^{-\mu_i}\ {}_1F_1\big(\frac{k}{2}-1; \frac{k}{2}; \mu_i\big)\Big]. \tag{3.12}$$

■

Corollary 3.2.1 shows that the asymptotic risk of the CCE Stein-like combined estimator is strictly less than the asymptotic risk of the CCE post-break estimator, for all parameter values and each $i = \{1,\ldots,N\}$, so long as the number of regressors (k) exceeds two. We note that (3.12) holds for all values of localizing parameter δ_i, even very large values of break sizes. Thus, the CCE

Stein-like combined estimator dominates the CCE post-break estimator.

Remark 3: *We note that the extension of the proposed estimator to the multiple break points is straightforward. Under multiple breaks, the proposed CCE Stein-like combined estimator is the combination of the CCE full-sample estimator (the efficient estimator) and the CCE post-break estimator after the most recent break point (the consistent estimator). For a similar discussion in a time-series model with no common correlated factor structures see Lee et al. (2022).*

4 Mean group Stein-like combined estimator

It is known in the panel data model literature to use the averaging estimates of the individual cross-section units to estimate the common mean. Pesaran (2006) introduces the common correlated effect mean group estimator (CCEMG), which is a simple average of the individual-specific CCE estimators. In this section, we develop the mean group estimator based on the individual CCE Stein-like combined estimators introduced in the previous Section. We define the CCEMG Stein-like combined estimator as $\widetilde{\beta}_{MG} = \alpha\ \widetilde{\beta}_{MG,Full} + (1-\alpha)\ \widetilde{\beta}_{MG(2)}$, where $\widetilde{\beta}_{MG,Full}$ is the CCEMG full-sample estimator, $\widetilde{\beta}_{MG(2)}$ is the CCEMG post-break estimator, and α is similar to (2.7) except that the weighted squared loss function defined in (2.8) is now $\mathcal{D}_N = N\big(\widetilde{\beta}_{MG(2)} - \widetilde{\beta}_{MG,Full}\big)' \mathbb{W}\left(\widetilde{\beta}_{MG(2)} - \widetilde{\beta}_{MG,Full}\right)$. In this section, we first develop the asymptotic distribution of the CCEMG Stein-like combined estimator, and then we establish its asymptotic risk.

Assumption 8: *For $i = 1, \dots, N$, $\beta_{i(1)} = \beta_{(1)} + \nu_{i,\beta_{(1)}}$ with $\nu_{i,\beta_{(1)}} \sim i.i.d.(0, \Sigma_{\beta_{(1)}})$, and $\beta_{i(2)} = \beta_{(2)} + \nu_{i,\beta_{(2)}}$ with $\nu_{i,\beta_{(2)}} \sim i.i.d.(0, \Sigma_{\beta_{(2)}})$. Besides, the random deviations $\nu_{i,\beta_{(1)}}$ and $\nu_{i,\beta_{(2)}}$ are independent of $\gamma_j, \Gamma_j, \epsilon_{jt}$ and v_{jt} for all i, j and t.*

Assumption 8 states that $\beta_{i(1)}$ and $\beta_{i(2)}$ are independent of Γ_j. This implies that as $N \to \infty$, $\bar{C}_{(1)} = \sum_{i=1}^{N} \theta_i C_{i(1)} \xrightarrow{p} \mathbb{E}(C_{i(1)}) = (\gamma + \Gamma\beta_{(1)}, \Gamma)$ and $\bar{C}_{(2)} = \sum_{i=1}^{N} \theta_i C_{i(2)} \xrightarrow{p} \mathbb{E}(C_{i(2)}) = (\gamma + \Gamma\beta_{(2)}, \Gamma)$. Thus, the rank condition in Assumption 6 requires non-zero means for γ and Γ which is satisfied based on Assumption 5.

Theorem 4.1: *Under Assumptions 1-8, and $\sqrt{N/T} \to c$ as $(N,T) \to \infty$ in which c is fixed, the joint asymptotic distribution of the CCEMG full-sample estimator and the CCEMG post-break* *estimator is*

$$\sqrt{N}\begin{bmatrix}\widetilde{\beta}_{MG,Full}-\beta_{(2)}\\ \widetilde{\beta}_{MG(2)}-\beta_{(2)}\end{bmatrix}\xrightarrow{d}\dot{\eta}+\dot{Z}, \tag{4.1}$$

where $\dot{\eta}=\begin{bmatrix}b_1 c\ \mathcal{Q}^{-1}\mathcal{Q}_1\delta_1\\ 0_{k\times 1}\end{bmatrix}$, $\dot{Z}\sim N\big(0,\dot{V}\big)$, $\frac{1}{N}\sum_{i=1}^{N}\big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\big)^{-1}\xrightarrow{p}\mathcal{Q}^{-1}$, $\frac{1}{N}\sum_{i=1}^{N}\big(\frac{\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}}{T_1}\big)\xrightarrow{p}\mathcal{Q}_1$, *and* $\dot{V}=\begin{bmatrix}V_{MG,Full} & cov\\ cov & V_{MG(2)}\end{bmatrix}$ *is the variance-covariance matrices of* $\widetilde{\beta}_{MG,Full}$ *and* $\widetilde{\beta}_{MG(2)}$ *with cov being their asymptotic covariance matrix.*

Besides, the asymptotic distribution of the weighted squared loss function is

$$\begin{aligned}\mathcal{D}_N &= N\big(\widetilde{\beta}_{MG(2)}-\widetilde{\beta}_{MG,Full}\big)'\,\mathbb{W}\,\big(\widetilde{\beta}_{MG(2)}-\widetilde{\beta}_{MG,Full}\big)\\ &\xrightarrow{d}\big(\dot{Z}+\dot{\eta}\big)'P\big(\dot{Z}+\dot{\eta}\big),\end{aligned} \tag{4.2}$$

where $P\equiv G\,\mathbb{W}\,G'$. *Finally, the asymptotic distribution of the CCEMG Stein-like combined estimator, denoted by* $\widetilde{\beta}_{MG}$, *is*

$$\begin{aligned}\sqrt{N}\big(\widetilde{\beta}_{MG}-\beta_{(2)}\big) &= \sqrt{N}\big(\widetilde{\beta}_{MG(2)}-\beta_{(2)}\big)-\alpha\sqrt{N}\big(\widetilde{\beta}_{MG(2)}-\widetilde{\beta}_{MG,Full}\big)\\ &\xrightarrow{d} G_2'\dot{Z}-\Big(\frac{\tau}{\big(\dot{Z}+\dot{\eta}\big)'P\big(\dot{Z}+\dot{\eta}\big)}\Big)_1\ G'\ \big(\dot{Z}+\dot{\eta}\big).\end{aligned} \tag{4.3}$$

Proof: See Appendix A.3. ■

Theorem 4.1 shows that the joint asymptotic distribution of the CCEMG full-sample estimator and the CCEMG post-break estimator is normally distributed. Besides, the asymptotic distribution of the CCEMG Stein-like combined estimator is a nonlinear function of $\dot{Z}$.

Using the results of Theorem 4.1, we derive the asymptotic risk for the CCEMG Stein-like combined estimator. The results are summarized in Theorem 4.2 below.

Theorem 4.2: *Under Assumptions 1-8, and* $\sqrt{N/T}\to c$ *as* $(N,T)\to\infty$ *in which c is fixed, for* $0\le\tau\le 2\big(tr(\mathcal{A})-2\lambda_{max}(\mathcal{A})\big)$, *the asymptotic risk of the CCEMG Stein-like combined estimator* *for any user specific positive definite choice of* $\mathbb{W}$ *is*

$$\rho\big(\widetilde{\beta}_{MG}, \mathbb{W}\big) \leq \rho\big(\widetilde{\beta}_{MG(2)}, \mathbb{W}\big) - \tau\Big[\frac{2\big(tr(\mathcal{A}) - 2\lambda_{max}(\mathcal{A})\big) - \tau}{\dot{\eta}' P \dot{\eta} + tr(P\dot{V})}\Big], \tag{4.4}$$

where $\mathcal{A} \equiv \mathbb{W} G_2' \dot{V} G$, *and* $\lambda_{max}(\mathcal{A})$ *denotes the maximum eigenvalues of* $\mathcal{A}$. *Thus, the asymptotic risk of the CCEMG Stein-like combined estimator is less than that of the CCEMG post-break estimator.*

Proof: See Appendix A.4. ■

The optimal value of τ, denoted by τ^*_{MG}, can be obtained by minimizing the asymptotic risk in (4.4). Corollary 4.2.1 shows the optimal value of the shrinkage parameter and its corresponding asymptotic risk.

Corollary 4.2.1: *The optimal value of the shrinkage parameter which minimizes the asymptotic risk of the CCEMG Stein-like combined estimator is*

$$\tau^*_{MG} = tr(\mathcal{A}) - 2\lambda_{max}(\mathcal{A}), \tag{4.5}$$

which is positive if $tr(\mathcal{A}) > 2\lambda_{max}(\mathcal{A})$. *Further, the asymptotic risk of the CCEMG Stein-like combined estimator after substituting* τ^*_{MG} *is*

$$\rho\big(\widetilde{\beta}_{MG}, \mathbb{W}\big) \leq \rho\big(\widetilde{\beta}_{MG(2)}, \mathbb{W}\big) - \frac{\big(tr(\mathcal{A}) - 2\lambda_{max}(\mathcal{A})\big)^2}{\dot{\eta}' P \dot{\eta} + tr(P\dot{V})}. \tag{4.6}$$

■

Corollary 4.2.1 shows that the asymptotic risk of the CCEMG Stein-like combined estimator is less than that of the CCEMG post-break estimator.

5 Monte Carlo simulations

This section employs Monte Carlo simulations to examine the performance of the CCEMG Stein-like combined estimator developed in the previous section. To do this, we consider the following data generating process, which is similar to the one considered in Pesaran (2006) but allows for a structural break,

$$y_{it} = \alpha_i + \sum_{j=1}^{3} x_{it,j}\beta_{ij}(T_1) + \sum_{j=1}^{4} \gamma_{ij} f_{jt} + \epsilon_{it}, \tag{5.1}$$

where $\alpha_i \sim i.i.d.$ N$(1,1)$, $\gamma_{ij} \sim i.i.d.$ N$(1, 0.2)$, and the idiosyncratic errors are generated as $\epsilon_{it} \sim i.i.d.$ N$(0, \sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.$ U$(0.5, 1.5)$. The break in the slope coefficients is

$$\beta_{ij}(T_1) = \begin{cases} \beta_{ij(1)} & \text{for} \quad t = 1, \ldots, T_1, \\ \beta_{ij(2)} & \text{for} \quad t = T_1 + 1, \ldots, T, \end{cases} \tag{5.2}$$

where $\beta_{ij(2)} = 1 + \eta_{ij}$, and $\eta_{ij} \sim i.i.d.$ N$(0, 0.04)$, and the break size in the individual slope coefficients, δ, is $\{0.1, 0.3, 0.7, 1\}$. We consider different break points in the individual slopes as $\{0.2T, 0.5T, 0.8T\}$ where $T = 100$.

The regressor $x_{it,j}$ contain the common correlated effect f_t,

$$x_{it,j} = a_{ij} + \gamma_{2ij} f_{jt} + v_{it,j}, \tag{5.3}$$

where $a_{ij} \sim i.i.d.$ N$(0.5, 0.5)$, $\gamma_{2ij} \sim i.i.d.$ N$(0.5, 0.5)$ and $v_{it,j} \sim i.i.d.$ N$(0, 1 - \rho_{vij}^2)$ with $\rho_{vij} \sim i.i.d.$ U$[0.05, 0.95]$. The factor f_{jt} is generated by the stationary process

$$f_{jt} = \rho_{fj} f_{jt-1} + v_{fjt}, \quad j = 1, 2, 3, 4, \; t = -49, \ldots, 0, 1, \ldots, T;$$

$v_{fjt} \sim i.i.d.$ N$(0, 1 - \rho_{fj}^2), \rho_{fj} = 0.5$, and $f_{j,-50} = 0$.

We consider different values for individuals, $N = \{100, 150, 200\}$. As discussed in Section 3, the estimated value for the unobserved factors are the cross-sectional averages of the dependent variable and regressors, i.e, $\widehat{f}_t = \bar{w}_t$, which by inserting Equation (3.3) $\widehat{f}_t = \bar{w}_t = \bar{C}'(T_1) f_t + \bar{u}_t(T_1)$. As $n \to \infty, \bar{u}_t(T_1) \xrightarrow{p} 0$. This means that $\widehat{f}_t$ is consistent for the space spanned by f_t, which is sufficient to control their effects. [3]

We compare the forecasting performance of the CCEMG Stein-like combined estimator, the CCEMG post-break estimator and the CCEMG full-sample estimator. Specifically, we report the

[3]See for example Westerlund (2018), and Karabiyik and Westerlund (2021).

relative mean squared forecast error (RMSFE) considering the CCEMG post-break estimator as the benchmark method.

Tables 1–2 report the simulation results. Based on the results, the CCEMG Stein-like combined estimator outperforms the CCEMG post-break estimator under any break points and break size. When the break happens toward the end of the sample (i.e., $b_1 = 0.8$), the out-performance of the CCEMG Stein-like combined estimator is larger since there are fewer observations in the post-break sample. This is expected because when the break point happens toward the end of the sample, the gain obtained from using the CCEMG Stein-like combine estimator relative to the CCEMG post-break estimator increases. This shows that one can have a better estimation by exploiting the observations in both regimes (full-sample) instead of only using the post-break sample observations. As the break size in the slope coefficients increases, the performance of the CCEMG Stein-like combined estimator becomes closer to the CCEMG post-break estimator. The CCEMG full-sample estimator performs better than the CCEMG post-break estimator only if the break size in the slope coefficients is small, and in the other cases, it under-performs it. This is because under a large break size, the CCEMG full-sample estimator has a large bias and its efficiency can not offset the large bias.

We have also compared the performance of the proposed combined estimator with $k = 6$ regressors in Table 2. The results show that when the number of regressors increases, there is a larger reduction in the RMSFE for the CCEMG Stein-like combined estimator. Overall the simulation results show that the CCEMG Stein-like combined estimator always out-performs or performs equivalent to the CCEMG post-break estimator. Thus, there is no cost in using the CCEMG combined estimator instead of the CCEMG post-break estimator.

Table 1: Simulation results for CCEMG estimator, with $T = 100$ and $k = 3$

	b_1 :	0.2		0.5		0.8	
	δ	$RMSFE_{Stein}$	$RMSFE_{Full}$	$RMSFE_{Stein}$	$RMSFE_{Full}$	$RMSFE_{Stein}$	$RMSFE_{Full}$
$N = 100$	0.1	0.9450	0.9579	0.7826	0.7417	0.6419	0.3147
	0.3	0.9487	1.4013	0.9481	2.8345	0.9442	2.2970
	0.7	0.9917	4.2168	0.9775	14.369	0.9917	12.158
	1.0	0.9895	7.8498	0.9842	29.299	0.9953	24.770
$N = 150$	0.1	0.8595	0.8693	0.8122	0.8480	0.6810	0.3969
	0.3	0.9548	1.7928	0.9464	3.8780	0.9414	3.3605
	0.7	0.9900	6.7270	0.9756	20.497	0.9735	17.71
	1.0	0.9894	13.076	0.9827	42.017	0.9864	35.99
$N = 200$	0.1	0.9664	1.1066	0.8746	1.0783	0.7200	0.5649
	0.3	0.9728	2.2019	0.9671	5.6798	0.9771	4.1756
	0.7	0.9967	8.7846	0.9852	29.944	0.9955	22.404
	1.0	0.9950	17.192	0.9896	61.104	0.9977	45.715

Note: This table reports the RMSFE where the benchmark model is the CCEMG post-break estimator. The first column shows different values of N, and the second column, δ, is the break size in the slope coefficients. In the heading of the table, $b_1 = T_1/T$, $RMSFE_{Stein}$ denotes the relative MSFE of the CCEMG Stein-like estimator over the CCEMG post-break estimator, and $RMSFE_{Full}$ denotes the relative MSFE of the CCEMG full-samplle estimator over the CCEMG post-break estimator.

Table 2: Simulation results for CCEMG estimator, with $T = 100$ and $k = 6$

	b_1 :	0.2		0.5		0.8	
	δ	$RMSFE_{Stein}$	$RMSFE_{Full}$	$RMSFE_{Stein}$	$RMSFE_{Full}$	$RMSFE_{Stein}$	$RMSFE_{Full}$
$N = 100$	0.1	0.8471	0.8476	0.6056	0.6171	0.3151	0.1301
	0.3	0.9427	1.2536	0.9379	2.1680	0.6025	0.6994
	0.7	0.9839	3.4538	0.9817	9.9828	0.8911	3.5679
	1.0	0.9956	6.2303	0.9887	19.949	0.9429	7.2355
$N = 150$	0.1	0.8732	0.8737	0.6906	0.7327	0.3363	0.1695
	0.3	0.9881	1.5031	0.9873	2.9272	0.6719	0.9848
	0.7	0.9929	4.5610	0.9989	13.782	0.9335	5.0456
	1.0	0.9924	8.3743	1.0000	27.556	0.9690	10.223
$N = 200$	0.1	0.9937	0.9966	0.8457	0.9630	0.3688	0.2436
	0.3	0.9851	1.9375	0.9632	3.9758	0.7764	1.3880
	0.7	0.9948	5.9770	0.9950	18.087	0.9685	6.8912
	1.0	0.9960	10.835	0.9956	35.727	0.9898	13.840

Note: See the notes to Table 1.

6 Empirical analysis

In this section, we evaluate the performance of the proposed CCEMG Stein-like combined estimator in forecasting the growth rate of real output. We use a quarterly data set of 18 industrialized countries from 1979:Q1 to 2016:Q4.[4] The predictors are: the real equity prices (eq_{it}), real short term interest rate (r_{it}), term spread ($l_{it} - r_{it}$) where l_{it} is real long term interest rate, and the corresponding country-specific foreign variables for each of the predictors. The foreign variables are generated using moving averages of the annual trade weights over three year period. The trade weight are computed as shares of exports and imports for each country. Therefore, the h-step ahead linear forecasting model is

$$y_{it+h} = x'_{it}\beta_{it} + \gamma'_i f_t + \epsilon_{it+h}, \tag{6.1}$$

[4]The countries are: Australia, Austria, Belgium, Canada, Finland, France, Germany, Italy, Japan, Netherlands, Norway, New Zealand, Spain, Sweden, Switzerland, United Kingdom, USA, and China. The data is available from Mohaddes and Raissi (2018).

where $x'_{it} = (eq_{it}, r_{it}, l_{it} - r_{it}, eq^*_{it}, r^*_{it}, l^*_{it} - r^*_{it})$, in which a "star" indicates the foreign variables. We compute h-step ahead forecasts ($h = 1, 4$) for different estimation methods (i.e., the CCEMG Stein-like combined estimator, the CCEMG post-break estimator, and the CCEMG full-sample estimator), using both rolling and expanding window forecasts. The estimated value for the unobserved factors is $\widehat{f}_t = \bar{w}_t$, which is the cross-sectional averages of the dependent and independent variables. Each time that we expand the window, we first estimate the break point and then obtain forecasts based on the CCEMG Stein-like combined estimator, the CCEMG post-break estimator, and the CCEMG full-sample estimator. The method of estimating the break point is the standard least-squares, i.e., minimizing the overall sum of squared residuals, see Baltagi et al. (2016). The rolling window forecasts is based on the most recent 10 years (40 quarters) of observations.

In order to evaluate the performance of our proposed CCEMG Stein-like combined estimator, we compute its mean squared forecast errors (MSFE) and compare it with those of the CCEMG post-break estimator and the CCEMG full-sample estimator. For this purpose, we divide the sample of observations into two parts. The first T observations are used as the initial in-sample estimation period, and the remaining observations are the pseudo out-of-sample evaluation period. We consider two different out-of-sample evaluation periods, 1995:Q1-2016:Q4 and 2005:Q1-2016:Q4, to see the forecasting performance of estimators with this choice.

Table 3 reports the results based on both rolling window and expanding window forecasts. In the heading of the table, "CCEMG Stein" reports the MSFE results for the CCEMG Stein-like combined estimator, "CCEMG Postbk" reports the MSFE results for the CCEMG post-break estimator, and "CCEMG Full" shows the MSFE results for the CCEMG full-sample estimator. Panel A shows the MSFEs with the out-of-sample evaluation periods of 1995:Q1-2016:Q4, while the results with the out-of-sample evaluation periods of 2005:Q1-2016:Q4 are presented in Panel B. To see whether the CCEMG Stein-like combined estimator significantly outperforms the CCEMG post-break estimator across all individuals, we report a panel version of the Diebold and Mariano test introduced by Pesaran et al. (2009) in Table 3, indicated by asterisks. The 1% significance level is denoted by ***. The break point estimation is stable throughout the estimation procedure. For example, for one-step-ahead forecast and the out-of-sample evaluation periods of 2005:Q1-2016:Q4, for the first 18 expanding windows a break point around dot-com bubble of early 2000 is estimated

Table 3: Empirical MSFE results for forecasting output growth

	Rolling Window			Expanding Window		
h	CCEMG Stein	CCEMG Postbk	CCEMG Full	CCEMG Stein	CCEMG Postbk	CCEMG Full
Panel A: 1995:Q1-2016:Q4						
1	21.979***	46.763	22.028	21.800***	23.477	21.858
4	22.802***	27.522	23.252	22.489***	24.079	22.381
Panel B: 2005:Q1-2016:Q4						
1	23.404***	69.608	23.908	23.722***	25.895	23.880
4	24.323***	33.138	25.218	24.426***	26.204	24.136

Note: This table reports the empirical MSFE results for the CCEMG Stein-like combined estimator, CCEMG post-break estimator, and CCEMG full-sample estimator. The first column shows different forecast horizons, h. Panel A reports the results with the out-of-sample evaluation periods of 1995:Q1-2016:Q4, while the results with the out-of-sample evaluation periods of 2005:Q1-2016:Q4 are presented in Panel B. An asterisk, ***, denotes forecast that is significantly better than that of obtained from the CCEMG post-break forecasts based on the panel Diebold–Mariano test statistic at 1% significance level.

with the estimated CCEMG Stein-like combination weight, $\hat{\alpha}$, roughly between 0.1 and 0.3. For the remaining expanding windows, a break point around financial crises of 2008 is detected with the range of $\hat{\alpha}$ approximately between 0.5-1. We see a similar pattern for other specifications.

Based on the results, the CCEMG Stein-like combined estimator has a lower MSFE than that of the CCEMG post-break estimator and the CCEMG full-sample estimator for various out-of-sample evaluation periods and forecast horizons. This out-performance is statistically significant at 1% significance level. In Panel A and for $h = 1$, the out-performance of the CCEMG Stein-like combined estimator relative to the CCEMG post-break estimator is 52.9% (7.1%) for the rolling window (expanding window) forecasts. When $h = 4$, the out-performance is 17.2% (6.6%) for the rolling window (expanding window) forecasts. When $h = 1$ and with the out-of-sample evaluation periods of 2005:Q1-2016:Q4, the out-performance of the CCEMG Stein-like combined estimator relative to the CCEMG post-break estimator is 66.3% (8.4%) for the rolling window (expanding window) forecasts. The out-performance becomes 26.6% (6.7%) for the rolling window (expanding window) forecasts when $h = 4$.

7 Conclusion

In this paper, we introduce a Stein-like combined estimator for estimating the slope coefficients of large heterogenous panel models with a general multifactor error structure which is due to unobservable common factors. The proposed CCE Stein-like combined estimator is the weighted averages of the CCE post-break estimator (i.e., using observations in the most recent regime) and the CCE full-sample estimator (i.e., using full-sample of observations). The combination weight is inversely proportional to the difference between the CCE post-break and the CCE full-sample estimators, and therefore measures the magnitude of the structural break. Thus, for a large break size, it assigns a small weight to the CCE full-sample estimator (which is biased), and a large weight to the CCE post-break estimator. The opposite is true for a small break size. We establish the asymptotic distribution and the asymptotic risk of the proposed CCE Stein-like combined estimator, and find the conditions under which the combined estimator uniformly out-performs the CCE post-break estimator, for any break sizes and break points. Furthermore, we establish the asymptotic distribution and the asymptotic risk for the CCE mean group (CCEMG) Stein-like combined estimator, and show that its asymptotic risk is smaller than that of the CCEMG post-break estimator. Monte Carlo simulations, and the empirical application of forecasting output growth rates of 18 industrialized countries show the significant superiority of using the proposed CCEMG Stein-like combined estimator over the alternative methods.

Even though the proposed combined estimators reduce the asymptotic risk, it is an open question whether this reduction can be used to improve inference. Besides, the considered regression model in the paper allows for dynamic structures through the general dynamics of the common effects in the error term. Alternatively, lagged dependent variables can be included as the regressors in the model. Furthermore, the extension of the proposed CCE Stein-like combined estimator to panel vector autoregressive models and nonlinear models under structural breaks have not been explored yet. These are beyond the scope of the present paper and we leave them for future work.

References

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

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

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

Baltagi, B. H., Feng, Q. and Kao, C. (2016). “Estimation of heterogeneous panels with structural breaks.” *Journal of Econometrics* 191, 176–195.

Baltagi, B. H., Feng, Q. and Kao, C. (2019). “Structural changes in heterogeneous panels with endogenous regressors.” *Journal of Applied Econometrics* 34, 883-892.

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.

Hansen, B. (2009). “Averaging estimators for regressions with a possible structural break.” *Econometric Theory* 25, 1498-1514.

Hansen, B. (2016). “Efficient shrinkage in parametric models.” *Journal of Econometrics* 190, 115-132.

Hansen, B. (2017). “Stein-like 2SLS estimator.” *Econometrics Reviews* 36, 840-852.

Hsiao, C., and Pesaran, M., (2008). “Random coefficient models.” In: The Econometrics of Panel Data. Chapter 6, Springer, pp. 185–213.

James, W. and Stein, C. M. (1961). “Estimation with quadratic loss.” *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 361-380.

Kao, C., Trapani, L. and Urga, G. (2012). “Asymptotics for panel models with common shocks.” *Econometric Reviews* 31, 390–439.

Karabiyik, H., and Westerlund, J. (2021). “Forecasting using cross-section average–augmented time series regressions.” *The Econometrics Journal* 24, 315-333.

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

Kim, D. (2014). “Common breaks in time trends for large panel data with a factor structure.” *The Econometrics Journal* 17, 301–337.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation.* 2nd ed., New York, Springer.

Lebedev, N. N. (1972). *Special Functions and their Applications.* Dover Publications, New York.

Lee, T., Parsaeian, S. and Ullah, A. (2022). "Optimal forecast under structural breaks." *Journal of Applied Econometrics* 37, 965-987.

Lee, T., Parsaeian, S. and Ullah, A. (2022a). "Forecasting under structural breaks using improved weighted estimation." *Oxford Bulletin of Economics and Statistics* 84, 1485-1501.

Lee, T., Parsaeian, S. and Ullah, A. (2022). "Efficient combined estimation under structural breaks." *Advances in Econometrics*, Vol. 43, ed. by Chudik, A., Hsiao C., and A. Timmermann.

Li, D., Qian, J., and Su, L., (2016). "Panel data models with interactive fixed effects and multiple structural breaks." *Journal of American Statistical Association* 111, 1804-1819.

Mehrabani, A. and Ullah, A. (2020). "Improved average estimation in seemingly unrelated regressions." *Econometrics* 8, 1-22.

Lumsdaine, R. L., Okui, R., and Wand, W. (2023). "Estimation of panel group structure models with structural breaks in group memberships and coefficients." *Journal of Econometrics* 233, 45-65.

Mohaddes, K., and Raissi M. (2018). "Global VAR (GVAR) database, 1979Q2-2016Q4". https://doi.org/10.17863/CAM.27337.

Okui, R. and Wang W. (2021). "Heterogeneous structural breaks in panel data models." *Journal of Econometrics* 220, 447-473.

Parsaeian, S. (2023). "Structural breaks in seemingly unrelated regression models." *Macroeconomic Dynamics*, 1-24.

Perron, P., Yamamoto, Y., and Zhou, J. (2020). "Testing jointly for structural changes in the error variance andcoefficients of a linear regression model." *Quantitative Economics* 11, 1019-1057.

Pesaran, M.H. (2006). "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica* 74, 967-1012.

Pesaran, M., (2015). *Time Series and Panel Data Econometrics.* Oxford University Press, ISBN: 9780198736912.

Pesaran, M.H., Schuermann, T., Smith, L.Vanessa, (2009). "Forecasting economic and financial variables with global VARs." *International Journal of Forecasting* 25, 642–675.

Qian, J., Su, L., (2016). "Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso." *Journal of Econometrics* 191, 86-109.

Stock, J. H., and Watson, M. W. (2004). "Combination forecasts of output growth in a seven-country data set." *Journal of Forecasting* 23, 405–430.

Ullah, A. (1974). "On the sampling distribution of improved estimators for coefficients in linear regression." *Journal of Econometrics* 2, 143-150.

Westerlund, J. (2018). "CCE in panels with general unknown factors." *The Econometrics Journal* 21, 264-276.

A Appendix: Mathematical details

A.1 Proof of Theorem 3.1:

First, we note that $\beta_i(T_1) = \beta_{i(1)} + \frac{\delta_i}{\sqrt{T}}\mathbb{1}(t > T_1) = \begin{cases} \beta_{i(1)} & \text{for} \quad t = 1, \dots, T_1, \\ \beta_{i(2)} = \beta_{i(1)} + \frac{\delta_i}{\sqrt{T}} & \text{for} \quad t = T_1 + 1, \dots, T, \end{cases}$

where $\mathbb{1}(\cdot)$ is an indicator function. Denote $\bar{u}_t = \begin{pmatrix} \bar{\epsilon}_t + \sum_{i=1}^N \theta_i v_{it}' \beta_{i(1)} \\ \bar{v}_t \end{pmatrix}$ and

$$\Delta \bar{u}_t(T_1) = \begin{cases} \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & \text{for} \quad t = 1, \dots, T_1, \\ \begin{pmatrix} \sum_{i=1}^N \theta_i v_{it}' \frac{\delta_i}{\sqrt{T}} \\ 0 \end{pmatrix}, & \text{for} \quad t = T_1 + 1, \dots, T. \end{cases}$$

Thus, $\bar{u}_t(T_1)$ in (3.3) is equal to $\bar{u}_t(T_1) = \bar{u}_t + \Delta\bar{u}_t(T_1)$.

Let $\underbrace{\bar{U}}_{T\times(k+1)} \equiv (\bar{u}_{(1)}', \bar{u}_{(2)}')' = (\bar{u}_1, \dots, \bar{u}_{T_1}, \bar{u}_{T_1+1} \dots, \bar{u}_T)'$, and $\underbrace{\Delta\bar{U}(T_1)}_{T\times(k+1)} \equiv \left(\Delta\bar{u}_{(1)}', \Delta\bar{u}_{(2)}'\right)' =$ $\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{i=1}^N \theta_i v_{i,T_1+1}' \frac{\delta_i}{\sqrt{T}} \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} \sum_{i=1}^N \theta_i v_{i,T}' \frac{\delta_i}{\sqrt{T}} \\ 0 \end{pmatrix} \right)$. Therefore, by stacking the cross-sectional averages within each regime in (3.3), $\bar{w}_t = \bar{C}'(T_1) f_t + \bar{u}_t(T_1)$, we get

$$\begin{cases} \bar{W}_{(1)} = F_{(1)}\bar{C}_{(1)} + \bar{U}_{(1)} & \text{for} \quad t = 1, \dots, T_1, \\ \bar{W}_{(2)} = F_{(2)}\bar{C}_{(2)} + \bar{U}_{(2)} & \text{for} \quad t = T_1 + 1, \dots, T, \end{cases} \tag{A.1}$$

where $\bar{U}_{(1)} = \bar{u}_{(1)} + \Delta\bar{u}_{(1)}$ and $\bar{U}_{(2)} = \bar{u}_{(2)} + \Delta\bar{u}_{(2)}$. With this notation, we obtain the following Lemma, which can be proved similarly to Lemmas 1-3 in Pesaran (2006) and Lemma 5 in Baltagi et al. (2016).

Lemma A.1: *Under Assumptions 1-7,*

(i) $\frac{F'_{(1)}F_{(1)}}{T_1} = O_p(1)$; $\frac{F'_{(2)}F_{(2)}}{T-T_1} = O_p(1)$;

(ii) $\frac{\bar{U}'_{(1)}\bar{U}_{(1)}}{T_1} = O_p\big(\frac{1}{N}\big)$; $\frac{\bar{U}'_{(2)}\bar{U}_{(2)}}{T-T_1} = O_p\big(\frac{1}{N}\big)$;

(iii) $\frac{\bar{F}'_{(1)}\bar{U}_{(1)}}{T_1} = O_p\big(\frac{1}{T\sqrt{N}}\big)$; $\frac{\bar{F}'_{(2)}\bar{U}_{(2)}}{T-T_1} = O_p\big(\frac{1}{T\sqrt{N}}\big)$;

(iv) $\frac{\bar{X}'_{i(1)}F_{(1)}}{T_1} = O_p(1)$; $\frac{\bar{X}'_{i(2)}F_{(2)}}{T-T_1} = O_p(1)$;

(v) $\frac{\bar{X}'_{i(1)}\bar{U}_{(1)}}{T_1} = O_p\big(\frac{1}{N}\big) + O_p\big(\frac{1}{\sqrt{NT}}\big)$; $\frac{\bar{X}'_{i(2)}\bar{U}_{(2)}}{T-T_1} = O_p\big(\frac{1}{N}\big) + O_p\big(\frac{1}{\sqrt{NT}}\big)$;

Using (A.1), we obtain

$$\begin{aligned}\frac{1}{T_1}\bar{W}'_{(1)}\bar{W}_{(1)} &= \frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)} + \frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}\bar{U}_{(1)} + \frac{1}{T_1}\bar{U}'_{(1)}F_{(1)}\bar{C}_{(1)} + \frac{1}{T_1}\bar{U}'_{(1)}\bar{U}_{(1)} \\ &= \frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)} + \mathbb{E}_{(1)} = O_p(1),\end{aligned} \tag{A.2}$$

where $\mathbb{E}_{(1)} \equiv \frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}\bar{U}_{(1)} + \frac{1}{T_1}\bar{U}'_{(1)}F_{(1)}\bar{C}_{(1)} + \frac{1}{T_1}\bar{U}'_{(1)}\bar{U}_{(1)} = O_p\big(\frac{1}{N}\big) + O_p\big(\frac{1}{T\sqrt{N}}\big)$, and $\frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)} = O_p(1)$ by Lemma A.1. Thus, using (A.2),

$$\begin{aligned}\Big(\frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)}\Big)^{-1} - \Big(\frac{1}{T_1}\bar{W}'_{(1)}\bar{W}_{(1)}\Big)^{-1} &= \Big(\frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)}\Big)^{-1} \\ &\quad - \Big(I + (\frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)})^{-1}\ \mathbb{E}_{(1)}\Big)^{-1}\Big(\frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)}\Big)^{-1} \\ &= \Big[I - \Big(I + (\frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)})^{-1}\ \mathbb{E}_{(1)}\Big)^{-1}\Big]\Big(\frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)}\Big)^{-1} \\ &= \sum_{l=1}^{\infty}(-1)^{l+1}\Big((\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1}\bar{C}_{(1)})^{-1}\mathbb{E}_{(1)}\Big)^l\Big(\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1}\bar{C}_{(1)}\Big)^{-1} \equiv f(\mathbb{E}_{(1)}).\end{aligned} \tag{A.3}$$

Therefore, $\left(\frac{1}{T_1}\bar{W}'_{(1)}\bar{W}_{(1)}\right)^{-1} = \left(\frac{1}{T_1}\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)}\right)^{-1} + f(\mathbb{E}_{(1)})$. It follows that

$$
\begin{aligned}
M_{w(1)}F_{(1)} &= \Big(I_{T_1} - \bar{W}_{(1)}\big(\frac{1}{T_1}\bar{W}'_{(1)}\bar{W}_{(1)}\big)^{-1}\ \frac{1}{T_1}\bar{W}'_{(1)}\Big)F_{(1)} \\
&= \Bigg(I_{T_1} - \big(F_{(1)}\bar{C}_{(1)} + \bar{U}_{(1)}\big)\Big(\big(\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1}\bar{C}_{(1)}\big)^{-1} + f(\mathbb{E}_{(1)})\Big)\frac{1}{T_1}\big(F_{(1)}\bar{C}_{(1)} + \bar{U}_{(1)}\big)'\Bigg)F_{(1)} \\
&= \Bigg(I_{T_1} - \big(F_{(1)}\bar{C}_{(1)}\big)\big(\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)}\big)^{-1}\big(F_{(1)}\bar{C}_{(1)}\big)'\Bigg)F_{(1)} \\
&- \big(F_{(1)}\bar{C}_{(1)}\big)\Bigg(f(\mathbb{E}_{(1)})\big(\frac{1}{T_1}F_{(1)}\bar{C}_{(1)}\big)' + \Big(\big(\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1}\bar{C}_{(1)}\big)^{-1} + f(\mathbb{E}_{(1)})\Big)\frac{1}{T_1}\bar{U}'_{(1)}\Bigg)F_{(1)} \\
&- \bar{U}_{(1)}\Big(\big(\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1}\bar{C}_{(1)}\big)^{-1} + f(\mathbb{E}_{(1)})\Big)\big(\frac{F_{(1)}\bar{C}_{(1)}}{T_1} + \frac{\bar{U}_{(1)}}{T_1}\big)'F_{(1)} \\
&= F_{(1)}\mathbb{D}_{1(1)} + \bar{U}_{(1)}\mathbb{D}_{2(1)},
\end{aligned}
\tag{A.4}
$$

where $\Big(I_{T_1} - \big(F_{(1)}\bar{C}_{(1)}\big)\big(\bar{C}'_{(1)}F'_{(1)}F_{(1)}\bar{C}_{(1)}\big)^{-1}\big(F_{(1)}\bar{C}_{(1)}\big)'\Big)F_{(1)} = \Big(I_{T_1} - F_{(1)}\big(F'_{(1)}F_{(1)}\big)^{-1}F'_{(1)}\Big)F_{(1)} = 0$, see Pesaran (2006). Also,

$$
\begin{aligned}
\mathbb{D}_{1(1)} &\equiv -\bar{C}_{(1)}f(\mathbb{E}_{(1)})\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1} - \bar{C}_{(1)}\Bigg(\Big(\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1}\bar{C}_{(1)}\Big)^{-1} + f(\mathbb{E}_{(1)})\Bigg)\frac{\bar{U}'_{(1)}F_{(1)}}{T_1} \\
&= O_p\big(\frac{1}{N}\big) + O_p\big(\frac{1}{T_1\sqrt{N}}\big) + O_p(1)\times O_p\big(\frac{1}{T_1\sqrt{N}}\big) = O_p\big(\frac{1}{N}\big) + O_p\big(\frac{1}{T_1\sqrt{N}}\big),
\end{aligned}
\tag{A.5}
$$

and

$$
\mathbb{D}_{2(1)} \equiv -\Big(\big(\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1}\bar{C}_{(1)}\big)^{-1} + f(\mathbb{E}_{(1)})\Big)\Big(\bar{C}'_{(1)}\frac{F'_{(1)}F_{(1)}}{T_1} + \frac{\bar{U}'_{(1)}F_{(1)}}{T_1}\Big) = O_p(1). \tag{A.6}
$$

Thus, using (A.4), we obtain

$$
\begin{aligned}
M_{w(1)}F_{(1)}\gamma_i &= F_{(1)}\mathbb{D}_{1(1)}\gamma_i + \bar{U}_{(1)}\mathbb{D}_{2(1)}\gamma_i \\
&= O_p\big(\frac{1}{N}\big) + O_p\big(\frac{1}{T_1\sqrt{N}}\big) + O_p\big(\frac{1}{\sqrt{N}}\big) = O_p\big(\frac{1}{\sqrt{N}}\big).
\end{aligned}
\tag{A.7}
$$

Similarly, for the second regime $M_{w(2)}F_{(2)}\gamma_i = F_{(2)}\mathbb{D}_{1(2)}\gamma_i + \bar{U}_{(2)}\mathbb{D}_{2(2)}\gamma_i = O_p\big(\frac{1}{\sqrt{N}}\big)$, where $\mathbb{D}_{1(2)} \equiv -\bar{C}_{(2)}f(\mathbb{E}_{(2)})\bar{C}'_{(2)}\frac{F'_{(2)}F_{(2)}}{T-T_1} - \bar{C}_{(2)}\Big(\Big(\bar{C}'_{(2)}\frac{F'_{(2)}F_{(2)}}{T-T_1}\bar{C}_{(2)}\Big)^{-1} + f(\mathbb{E}_{(2)})\Big)\frac{\bar{U}'_{(2)}F_{(2)}}{T-T_1} = O_p\big(\frac{1}{N}\big) + O_p\big(\frac{1}{T_1\sqrt{N}}\big)$, and $\mathbb{D}_{2(2)} \equiv -\Big(\big(\bar{C}'_{(2)}\frac{F'_{(2)}F_{(2)}}{T-T_1}\bar{C}_{(2)}\big)^{-1} + f(\mathbb{E}_{(2)})\Big)\Big(\bar{C}'_{(2)}\frac{F'_{(2)}F_{(2)}}{T-T_1} + \frac{\bar{U}'_{(2)}F_{(2)}}{T-T_1}\Big) = O_p(1)$.

Using the above derivations, we establish the asymptotic distributions for the CCE estimators. The CCE full-sample estimator, for each i, is written as

$$\begin{aligned}\widetilde{\beta}_{i,Full} &= \left(\widetilde{X}_i'\widetilde{X}_i\right)^{-1}\widetilde{X}_i'\widetilde{Y}_i \\ &= \left(\widetilde{X}_i'\widetilde{X}_i\right)^{-1}\Big(\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}\big(\beta_{i(1)}-\beta_{i(2)}\big)+\widetilde{X}_{i(1)}'\widetilde{\epsilon}^*_{i(1)}+\widetilde{X}_{i(2)}'\widetilde{\epsilon}^*_{i(2)}+\widetilde{X}_i'\widetilde{X}_i\beta_{i(2)}\Big) \\ &= \beta_{i(2)}+\left(\widetilde{X}_i'\widetilde{X}_i\right)^{-1}\Big(\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}\big(\beta_{i(1)}-\beta_{i(2)}\big)+\widetilde{X}_{i(1)}'M_{w_{(1)}}F_{(1)}\gamma_i+\widetilde{X}_{i(1)}'\widetilde{\epsilon}_{i(1)} \\ &\quad+\widetilde{X}_{i(2)}'M_{w_{(2)}}F_{(2)}\gamma_i+\widetilde{X}_{i(2)}'\widetilde{\epsilon}_{i(2)}\Big)\end{aligned} \tag{A.8}$$

where $\widetilde{Y}_i = \big(\widetilde{Y}_{i(1)}', \widetilde{Y}_{i(2)}'\big)'$ is a $T\times 1$ vector of the transformed dependent variable, and $\widetilde{X}_i = \big(\widetilde{X}_{i(1)}', \widetilde{X}_{i(2)}'\big)'$ is a $T\times k$ matrix of the transformed regressors. Therefore, the asymptotic distribution of the CCE full-sample estimator, for each i, is

$$\begin{aligned}\sqrt{T}\big(\widetilde{\beta}_{i,Full}-\beta_{i(2)}\big) &= \Big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\Big)^{-1}\Big(\frac{\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}}{T_1}\delta_i b_1+\frac{\widetilde{X}_i'\widetilde{\epsilon}_i}{\sqrt{T}}+\sqrt{\frac{b_1}{T_1}}\widetilde{X}_{i(1)}'M_{w_{(1)}}F_{(1)}\gamma_i \\ &\quad+\sqrt{\frac{(1-b_1)}{T-T_1}}\widetilde{X}_{i(2)}'M_{w_{(2)}}F_{(2)}\gamma_i\Big) \\ &= \Big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\Big)^{-1}\Big[\frac{\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}}{T_1}\delta_i b_1+\frac{\widetilde{X}_i'\widetilde{\epsilon}_i}{\sqrt{T}}+O_p\big(\frac{\sqrt{T}}{N}\big)+O_p\big(\frac{1}{\sqrt{N}}\big)\Big] \\ &\xrightarrow{d} N\big(b_1\Sigma_i^{-1}\Sigma_{i(1)}\delta_i, \sigma_i^2\Sigma_i^{-1}\big).\end{aligned} \tag{A.9}$$

where $\widetilde{\epsilon}_i = (\widetilde{\epsilon}_{i(1)}', \widetilde{\epsilon}_{i(2)}')'$, $\sigma_i^2\Sigma_i^{-1} \equiv \operatorname*{plim}_{T\to\infty} \sigma_i^2\big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\big)^{-1} = V_{i,Full}$ is the variance of the CCE full-sample estimator, $\Sigma_{i(1)} \equiv \operatorname*{plim}_{T\to\infty}\big(\frac{\widetilde{X}_i(1)'\widetilde{X}_i(1)}{T_1}\big)$, and by using Lemma A.1

$$\begin{aligned}\frac{1}{\sqrt{T_1}}\widetilde{X}_{i(1)}'M_{w_{(1)}}F_{(1)}\gamma_i &= \frac{1}{\sqrt{T_1}}X_{i(1)}'M_{w_{(1)}}F_{(1)}\gamma_i \\ &= \frac{1}{\sqrt{T_1}}X_{i(1)}'F_{(1)}\mathbb{D}_{1(1)}\gamma_i+\frac{1}{\sqrt{T_1}}X_{i(1)}'\bar{U}_{(1)}\mathbb{D}_{2(1)}\gamma_i \\ &= \sqrt{T}\Big(O_p\big(\frac{1}{N}\big)+O_p\big(\frac{1}{T\sqrt{N}}\big)+O_p\big(\frac{1}{N}\big)+O_p\big(\frac{1}{\sqrt{NT}}\big)\Big) \\ &= O_p\big(\frac{\sqrt{T}}{N}\big)+O_p\big(\frac{1}{\sqrt{N}}\big).\end{aligned} \tag{A.10}$$

Similarly, $\frac{1}{\sqrt{T-T_1}}\widetilde{X}_{i(2)}'M_{w_{(2)}}F_{(2)}\gamma_i = \frac{1}{\sqrt{T-T_1}}X_{i(2)}'M_{w_{(2)}}F_{(2)}\gamma_i = O_p\big(\frac{\sqrt{T}}{N}\big)+O_p\big(\frac{1}{\sqrt{N}}\big)$. This order is asymptotically negligible when $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$. Therefore, the additional condition of $\sqrt{T}/N \to 0$ as $(N,T) \to \infty$ is needed to prove this Theorem.

Besides, the asymptotic distribution of the CCE post-break estimator is

$$\begin{aligned}
\sqrt{T}\big(\widetilde{\beta}_{i(2)} - \beta_{i(2)}\big) &= \Big(\frac{1}{\sqrt{1-b_1}}\Big)\Big(\frac{\widetilde{X}'_{i(2)}\widetilde{X}_{i(2)}}{T-T_1}\Big)^{-1}\left(\frac{\widetilde{X}'_{i(2)}\widetilde{\epsilon}_{i(2)}}{\sqrt{T-T_1}} + \frac{1}{\sqrt{T-T_1}}\widetilde{X}'_{i(2)}M_{w_{(2)}}F_{(2)}\gamma_i\right) \\
&= \Big(\frac{1}{\sqrt{1-b_1}}\Big)\Big(\frac{\widetilde{X}'_{i(2)}\widetilde{X}_{i(2)}}{T-T_1}\Big)^{-1}\Big[\Big(\frac{\widetilde{X}'_{i(2)}\widetilde{\epsilon}_{i(2)}}{\sqrt{T-T_1}}\Big) + O_p\big(\frac{\sqrt{T}}{N}\big) + O_p\big(\frac{1}{\sqrt{N}}\big)\Big] \qquad \text{(A.11)} \\
&\xrightarrow{d} N\big(0, \frac{1}{1-b_1}\sigma_i^2\Sigma_{i(2)}^{-1}\big),
\end{aligned}$$

where $\frac{1}{1-b_1}\sigma_i^2\Sigma_{i(2)}^{-1} \equiv \underset{T\to\infty}{\text{plim}}\ \frac{1}{1-b_1}\sigma_i^2\big(\frac{\widetilde{X}'_{i(2)}\widetilde{X}_{i(2)}}{T-T_1}\big)^{-1} = V_{i(2)}$ is the asymptotic variance of the CCE post-break estimator. Using (A.9) and (A.11), the joint asymptotic distribution of the CCE full-sample estimator and the CCE post-break estimator is derived. This completes the proof of Theorem 3.1. ■

A.2 Proof of Theorem 3.2:

The asymptotic risk for the CCE Stein-like combined estimator, for any user specific positive definite choice matrix $\mathbb{W}$, and for each $i = 1, \dots, N$, is

$$\begin{aligned}
\rho\big(\widetilde{\beta}_i, \mathbb{W}\big) &= \mathbb{E}\Big[T\big(\widetilde{\beta}_i - \beta_{i(2)}\big)'\,\mathbb{W}\ \big(\widetilde{\beta}_i - \beta_{i(2)}\big)\Big] \\
&= T\,\mathbb{E}\Big[\big(\widetilde{\beta}_{i(2)} - \beta_{i(2)}\big) - \alpha\big(\widetilde{\beta}_{i(2)} - \widetilde{\beta}_{i,Full}\big)\Big]'\,\mathbb{W}\Big[\big(\widetilde{\beta}_{i(2)} - \beta_{i(2)}\big) - \alpha\big(\widetilde{\beta}_{i(2)} - \widetilde{\beta}_{i,Full}\big)\Big] \\
&= \rho\big(\widetilde{\beta}_{i(2)}, \mathbb{W}\big) + \tau^2\,\mathbb{E}\Big[(Z_i'M_iZ_i)^{-2}Z_i'A_iZ_i\Big] - 2\tau\,\mathbb{E}\Big[(Z_i'M_iZ_i)^{-1}Z_i'B_iZ_i\Big],
\end{aligned} \qquad \text{(A.12)}$$

where $M_i \equiv V_i^{1/2}\ G\ \big(V_{i(2)} - V_{i,Full}\big)^{-1}\ G'\ V_i^{1/2}$, $A_i \equiv V_i^{1/2}G\,\mathbb{W}\,G'V_i^{1/2}$ and $B_i \equiv V_i^{1/2}G\,\mathbb{W}\,G_2'V_i^{1/2}$.

Lemma A.2: *Let $\chi_p^2(\mu_i)$ denote a noncentral chi-square random variable with the noncentral parameter μ_i, for each $i = 1, \dots, N$, and the degrees of freedom p. Besides, let p denote a positive integer such that $p > 2r$. Then*

$$\mathbb{E}\Big[\Big(\chi_p^2(\mu_i)\Big)^{-r}\Big] = 2^{-r}e^{-\mu_i}\frac{\Gamma(\frac{p}{2}-r)}{\Gamma(\frac{p}{2})}\ {}_1F_1\Big(\frac{p}{2}-r; \frac{p}{2}; \mu_i\Big),$$

where ${}_1F_1(.;.;.)$ *is the confluent hypergeometric function which is defined as* ${}_1F_1(a;b;\mu_i) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n} \frac{\mu_i^n}{n!}$*, where* $(a)_n = a(a+1)\dots(a+n-1)$ *and* $(a)_0 = 1$*. See Ullah (1974).*

Lemma A.3: *The definition of the confluent hypergeometric function implies the following relations:*

1. ${}_1F_1(a;b;\mu_i) = {}_1F_1(a+1;b;\mu_i) - \frac{\mu_i}{b}\ {}_1F_1(a+1;b+1;\mu_i)$,
2. ${}_1F_1(a;b;\mu_i) = \frac{b-a}{b}\ {}_1F_1(a;b+1;\mu_i) + \frac{a}{b}\ {}_1F_1(a+1;b+1;\mu_i)$, *and*
3. $(b-a-1)\ {}_1F_1(a;b;\mu_i) = (b-1)\ {}_1F_1(a;b-1;\mu_i) - a\ {}_1F_1(a+1;b+1;\mu_i)$.

See Lebedev (1972), pp. 262.

Lemma A.4: *Let the* $T\times 1$ *vector* Z_i *be normally distributed with mean vector* θ_i *and covariance matrix* I_T*,* M_i *be any* $T\times T$ *idempotent matrix with rank* r*, and* A_i *be any* $T\times T$ *matrix, for each* $i = 1,\dots,N$*. We assume* $\phi(\cdot)$ *is a Borel measurable function. Then:*

$$\begin{aligned}\mathbb{E}\left[\phi\big(Z_i'M_iZ_i\big)Z_i'A_iZ_i\right] &= \mathbb{E}\left[\phi\Big(\chi^2_{r+2}(\mu_i)\Big)\right]tr(A_iM_i) + \mathbb{E}\left[\phi\Big(\chi^2_{r+4}(\mu_i)\Big)\right]\theta_i'M_iA_iM_i\theta_i \\ &+ \mathbb{E}\left[\phi\Big(\chi^2_{r}(\mu_i)\Big)\right]tr\big(A_i - A_iM_i\big) + \mathbb{E}\left[\phi\Big(\chi^2_{r}(\mu_i)\Big)\right]\theta_i'\big(I_T - M_i\big)A_i\big(I_T - M_i\big)\theta_i \\ &+ \mathbb{E}\left[\phi\Big(\chi^2_{r+2}(\mu_i)\Big)\right]\Big(\theta_i'A_iM_i\theta_i + \theta_i'M_iA_i\theta_i - 2\theta_i'M_iA_iM_i\theta_i\Big),\end{aligned}$$

where $\mu_i \equiv \frac{\theta_i'M_i\theta_i}{2}$ *is the non-centrality parameter. See Lee et al. (2022) for the proof.*

By using Lemmas A.2-A.4, we calculate the asymptotic risk for the CCE Stein-like combined estimator in (A.12) as

$$\begin{aligned}\rho\big(\widetilde{\beta}_i,\mathbb{W}\big) &= \rho\big(\widetilde{\beta}_{i(2)},\mathbb{W}\big) + \tau^2\,\mathbb{E}\left[(Z_i'M_iZ_i)^{-2}Z_i'A_iZ_i\right] - 2\tau\,\mathbb{E}\left[(Z_i'M_iZ_i)^{-1}Z_i'B_iZ_i\right] \\ &= \rho\big(\widetilde{\beta}_{i(2)},\mathbb{W}\big) + \tau^2\Big\{\left[\chi^2_{k+2}(\mu_i)\right]^{-2}\mathrm{tr}\big(A_iM_i\big) + \mathbb{E}\left[\chi^2_{k+4}(\mu_i)\right]^{-2}\eta_i'M_iA_iM_i\eta_i\Big\} \\ &- 2\tau\Big\{\mathbb{E}\left[\chi^2_{k+2}(\mu_i)\right]^{-1}\mathrm{tr}\big(B_iM_i\big) + \mathbb{E}\left[\chi^2_{k+4}(\mu_i)\right]^{-1}\eta_i'M_iB_iM_i\eta_i \\ &+ \mathbb{E}\left[\chi^2_{k+2}(\mu_i)\right]^{-1}\Big(\eta_i'B_iM_i\eta_i + \eta_i'M_iB_i\eta_i - 2\eta_i'M_iB_iM_i\eta_i\Big)\Big\} \\ &= \rho\big(\widetilde{\beta}_{i(2)},\mathbb{W}\big) + \tau^2\Bigg\{\Big[\frac{1}{4}\,e^{-\mu_i}\,\frac{\Gamma(\frac{k}{2}-1)}{\Gamma(\frac{k}{2}+1)}\,{}_1F_1\big(\frac{k}{2}-1;\frac{k}{2}+1;\mu_i\big)\Big]\mathrm{tr}(A_i)\end{aligned}$$

$$
\begin{aligned}
&+ \Big[\frac{1}{4}\ e^{-\mu_i}\ \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+2)}\ {}_1F_1\Big(\frac{k}{2};\frac{k}{2}+2;\mu_i\Big)\Big]\big(\eta_i' A_i \eta_i\big)\Bigg\} \\
&- 2\tau\Bigg\{\Big[\frac{1}{2}\ e^{-\mu_i}\ \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k}{2}+1)}\ {}_1F_1\Big(\frac{k}{2};\frac{k}{2}+1;\mu_i\Big)\Big]\big(\mathrm{tr}(A_i) - \eta_i' A_i \eta_i\big) \\
&+ \Big[\frac{1}{2}\ e^{-\mu_i}\ \frac{\Gamma(\frac{k}{2}+1)}{\Gamma(\frac{k}{2}+2)}\ {}_1F_1\Big(\frac{k}{2}+1;\frac{k}{2}+2;\mu_i\Big)\Big]\big(\eta_i' A_i \eta_i\big)\Bigg\} \\
&= \rho\big(\widetilde{\beta}_{i(2)},\mathbb{W}\big) + \tau^2\Bigg\{\Big[\frac{\eta_i' A_i \eta_i}{(k-2)\eta_i' M_i \eta_i}\Big]\Big[e^{-\mu_i}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu_i\Big)\Big] \\
&- \Big[\frac{\eta_i' A_i \eta_i}{(k-2)\eta_i' M_i \eta_i} - \frac{\mathrm{tr}(A_i)}{k(k-2)}\Big]\Big[e^{-\mu_i}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu_i\Big)\Big]\Bigg\} \\
&- \frac{2\ \tau}{k-2}\Bigg\{\Big[\mathrm{tr}(B_i) - \frac{2\ \eta_i' A_i \eta_i}{\eta_i' M_i \eta_i}\Big]\Bigg[e^{-\mu_i}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2};\mu_i\Big)\Bigg] \\
&- \Big[2\frac{\mathrm{tr}(B_i)}{k} - 2\frac{\eta_i' A_i \eta_i}{\eta_i' M_i \eta_i}\Big]\Big[e^{-\mu_i}\ {}_1F_1\Big(\frac{k}{2}-1;\frac{k}{2}+1;\mu_i\Big)\Big]\Bigg\} \\
&= \rho\big(\widetilde{\beta}_{i(2)},\mathbb{W}\big) - \frac{\tau\ \eta_i' A_i \eta_i}{k(k+2)}\Big[2\big(\frac{\mathrm{tr}(A_i)\ \eta_i' M_i \eta_i}{\eta_i' A_i \eta_i} - 2\big) - \tau\Big]\Big[\ e^{-\mu_i}{}_1F_1\big(\frac{k}{2};\frac{k}{2}+2;\mu_i\big)\Big] \\
&- \frac{\tau\ \mathrm{tr}(A_i)}{k(k-2)}\Big[2(k-2) - \tau\Big]\Big[e^{-\mu_i}{}_1F_1\big(\frac{k}{2}-1;\frac{k}{2}+1;\mu_i\big)\Big],
\end{aligned}
\tag{A.13}
$$

where M_i is an idempotent matrix with rank k, $A_iM_i = M_iA_i = A_i$, $M_iA_iM_i = A_i$, $B_iM_i = A_i$, $M_iB_i = B_i$, $B_i\eta_i = 0$, $M_iB_iM_i = A_i$, and ${}_1F_1\big(\frac{k}{2}-1;\frac{k}{2};\mu_i\big) - {}_1F_1\big(\frac{k}{2}-1;\frac{k}{2}+1;\mu_i\big) = \frac{2\mu_i\ (k-2)}{k(k+1)}\Big[{}_1F_1\big(\frac{k}{2};\frac{k}{2}+2;\mu_i\big)\Big]$. Thus, the asymptotic risk of the CCE Stein-like combined estimator is less than that of the CCE post-break estimator for any positive definite choice of $\mathbb{W}$ under the following two conditions:

$$
0 \le \tau \le 2\big(\frac{\mathrm{tr}(A_i)\ \eta_i' M_i \eta_i}{\eta_i' A_i \eta_i} - 2\big), \tag{A.14}
$$

and

$$
0 \le \tau \le 2(k-2). \tag{A.15}
$$

The upper bound in (A.14) is positive if

$$\begin{aligned} \mathrm{tr}(A_i) &> \operatorname*{Sup}_{\Phi} \frac{2\eta_i' A_i \eta_i}{\eta_i' M_i \eta_i}, \\ \mathrm{tr}\big(\,\mathbb{W}(V_{i(2)} - V_{i,Full})\big) &> \operatorname*{Sup}_{\Phi} \frac{2\ \Phi'\ (V_{i(2)} - V_{i,Full})^{1/2}\,\mathbb{W}(V_{i(2)} - V_{i,Full})^{1/2}\ \Phi}{\Phi'\Phi} \\ \mathrm{tr}\big(\,\mathbb{W}(V_{i(2)} - V_{i,Full})\big) &> 2\lambda_{\max}\big((V_{i(2)} - V_{i,Full})^{1/2}\,\mathbb{W}(V_{i(2)} - V_{i,Full})^{1/2}\big), \end{aligned} \tag{A.16}$$

where $\Phi \equiv (V_{i(2)} - V_{i,Full})^{-1/2} G' V_i^{1/2} \eta_i$. Also, the upper bound in (A.15) is positive if $k > 2$.

Using the results in (A.13), the optimal value for the shrinkage parameter, denoted by τ_i^*, is $\tau_i^* = \frac{\mathrm{tr}(A_i)\ \eta_i' M_i \eta_i}{\eta_i' A_i \eta_i} - 2$, which is positive so long as the condition in (A.16) is satisfied. Substituting the optimal value of the shrinkage parameter into the asymptotic risk function in (A.13), if $0 \leq \tau \leq 2\big(\frac{\mathrm{tr}(A_i)\ \eta_i' M_i \eta_i}{\eta_i' A_i \eta_i} - 2\big)$ and the condition in (A.16) is hold, the asymptotic risk for the CCE Stein-like combined estimator, for any user specific positive definite choice matrix $\mathbb{W}$, and for each $i = \{1, \ldots, N\}$, is

$$\begin{aligned} \rho\big(\widetilde{\beta}_i, \mathbb{W}\big) = \rho\big(\widetilde{\beta}_{i(2)}, \mathbb{W}\big) &- \frac{1}{k-2}\left[\Big(\frac{\mathrm{tr}(A_i)\ \eta_i' M_i \eta_i}{\eta_i' A_i \eta_i} - 2\Big)^2 \ \Big(\frac{\eta_i' A_i \eta_i}{\eta_i' M_i \eta_i}\Big)\right]\left[e^{-\mu_i}{}_1F_1\Big(\frac{k}{2} - 1; \frac{k}{2}; \mu_i\Big)\right] \\ &- \frac{1}{k-2}\left[\Big(\frac{\mathrm{tr}(A_i)\ \eta_i' M_i \eta_i}{\eta_i' A_i \eta_i}\Big)^2 - 4\right]\left[\frac{\eta_i' A_i \eta_i}{\eta_i' M_i \eta_i} - \frac{\mathrm{tr}(A_i)}{k}\right]\left[e^{-\mu_i}\ {}_1F_1\Big(\frac{k}{2} - 1; \frac{k}{2} + 1; \mu_i\Big)\right]. \end{aligned} \tag{A.17}$$

This shows that the asymptotic risk of the CCE Stein-like combined estimator is less than the asymptotic risk of the CCE post-break estimator for all values of localizing parameter δ_i, even very large values of break sizes. Thus, the CCE Stein-like combined estimator dominates the CCE post-break estimator.

We note that for when $\mathbb{W} = (V_{i(2)} - V_{i,Full})^{-1}$, the asymptotic risk of the CCE Stein-like combined estimator presented in (A.13) simplifies to the results of Theorem 3.2. Also, in this case, the optimal value of the shrinkage parameter becomes $\tau^* = k - 2$, and the upper bound in (A.14) becomes $k > 2$. Therefore, the CCE Stein-like combined estimator dominates the CCE post-break estimator so long as $k > 2$. This completes the proof of Theorem 3.2. ■

A.3 Proof of Theorem 4.1:

Under Assumption 8 of a random coefficient model, and the local alternative assumption, $\beta_{(1)} - \beta_{(2)} = \delta_1/\sqrt{T}$, the asymptotic distribution of the CCEMG full-sample estimator is

$$\begin{aligned}
\widetilde{\beta}_{MG,Full} &= \frac{1}{N}\sum_{i=1}^{N}\widetilde{\beta}_{i,Full} \\
&= \frac{1}{N}\sum_{i=1}^{N}\beta_{i(2)} + \frac{1}{N}\sum_{i=1}^{N}\Big(\widetilde{X}_i'\widetilde{X}_i\Big)^{-1}\Big(\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}(\beta_{i(1)} - \beta_{i(2)}) + \widetilde{X}_i'\widetilde{\epsilon}_i + \widetilde{X}_{i(1)}'M_{w_{(1)}}F_{(1)}\gamma_i \\
&\quad + \widetilde{X}_{i(2)}'M_{w_{(2)}}F_{(2)}\gamma_i\Big).
\end{aligned} \tag{A.18}$$

Therefore,

$$\begin{aligned}
\sqrt{N}\big(\widetilde{\beta}_{MG,Full} - \beta_{(2)}\big) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nu_{i,\beta_{(2)}} + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\Big)^{-1}\Big(\frac{\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}\ b_1}{T_1}\sqrt{T}(\beta_{i(1)} - \beta_{i(2)}) \\
&\quad + \frac{\widetilde{X}_i'\widetilde{\epsilon}_i}{\sqrt{T}} + \sqrt{\frac{b_1}{T_1}}\widetilde{X}_{i(1)}'M_{w_{(1)}}F_{(1)}\gamma_i + \sqrt{\frac{1-b_1}{T-T_1}}\widetilde{X}_{i(2)}'M_{w_{(2)}}F_{(2)}\gamma_i\Big) \\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nu_{i,\beta_{(2)}} + \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\Big)^{-1}\Big(\frac{\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}}{T_1}b_1\delta_1\Big) \\
&\quad - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\Big)^{-1}\Big(\frac{\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}}{T_1}b_1(\nu_{i,\beta_{(2)}} - \nu_{i,\beta_{(1)}})\Big) + O_p\big(\frac{1}{\sqrt{N}}\big) + O_p\big(\frac{1}{\sqrt{T}}\big) \\
&\xrightarrow{d} N\big(b_1 c\mathcal{Q}\delta_1, V_{MG,Full}\big),
\end{aligned} \tag{A.19}$$

where $\sqrt{\frac{N}{T}} \to c$ as $(N,T) \to \infty$ in which c is a fixed constant, $\frac{1}{N}\sum_{i=1}^{N}\big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\big)^{-1}\big(\frac{\widetilde{X}_{i(1)}'\widetilde{X}_{i(1)}}{T_1}\big) \xrightarrow{p} \mathcal{Q}$, and $V_{MG,Full}$ is the asymptotic variance of the CCEMG full-sample estimator. The variance estimator for $V_{MG,Full}$ suggested by Pesaran (2006) is given by $\frac{1}{N-1}\sum_{i=1}^{N}(\widetilde{\beta}_{i,Full} - \widetilde{\beta}_{MG,Full})(\widetilde{\beta}_{i,Full} - \widetilde{\beta}_{MG,Full})'$.

Besides, the order of the third term is derived as

$$Var\Big(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\Big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\Big)^{-1}\frac{\widetilde{X}_i'\widetilde{\epsilon}_i}{\sqrt{T}}\Big) = \frac{1}{NT}\sum_{i=1}^{N}\Big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\Big)^{-1}\ \frac{\widetilde{X}_i'Var(\widetilde{\epsilon}_i)\widetilde{X}_i}{T}\Big(\frac{\widetilde{X}_i'\widetilde{X}_i}{T}\Big)^{-1} = O_p\big(\frac{1}{T}\big), \tag{A.20}$$

implying that $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left(\frac{\widetilde{X}_i' \widetilde{X}_i}{T}\right)^{-1} \frac{\widetilde{X}_i' \widetilde{\epsilon}_i}{\sqrt{T}} = O_p(\frac{1}{\sqrt{T}})$. Besides, using the results of (A.10), the order of the forth term can be derived as

$$\begin{aligned}\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Big(\frac{\widetilde{X}_i' \widetilde{X}_i}{T}\Big)^{-1} \sqrt{\frac{b_1}{T_1}} \widetilde{X}_{i(1)}' M_{w_{(1)}} F_{(1)} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} O_p(1) \Big[O_p(\frac{\sqrt{T}}{N}) + O_p(\frac{1}{\sqrt{N}})\Big] \\ &= O_p(\frac{1}{\sqrt{N}}) + O_p(\frac{1}{\sqrt{T}}).\end{aligned} \tag{A.21}$$

The order of the fifth term can be derived similarly.

Furthermore, the asymptotic distribution of the CCEMG post-break estimator is

$$\begin{aligned}\widetilde{\beta}_{MG(2)} &= \frac{1}{N} \sum_{i=1}^{N} \widetilde{\beta}_{i(2)} \\ &= \beta_{(2)} + \frac{1}{N} \sum_{i=1}^{N} \nu_{i,\beta_{(2)}} + \frac{1}{N} \sum_{i=1}^{N} \Big(\widetilde{X}_{i(2)}' \widetilde{X}_{i(2)}\Big)^{-1} \Big(\widetilde{X}_{i(2)}' \widetilde{\epsilon}_{i(2)} + \widetilde{X}_{i(2)}' M_{w_{(2)}} F_{(2)} \gamma_i\Big).\end{aligned} \tag{A.22}$$

Therefore,

$$\begin{aligned}\sqrt{N}\big(\widetilde{\beta}_{MG(2)} - \beta_{(2)}\big) &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \nu_{i,\beta_{(2)}} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \Big(\frac{\widetilde{X}_{i(2)}' \widetilde{X}_{i(2)}}{T - T_1}\Big)^{-1} \Big(\frac{\widetilde{X}_{i(2)}' \widetilde{\epsilon}_{i(2)}}{\sqrt{1-b_1}\sqrt{T-T_1}} + \frac{\widetilde{X}_{i(2)}' M_{w_{(2)}} F_{(2)} \gamma_i}{\sqrt{1-b_1}\sqrt{T-T_1}}\Big) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \nu_{i,\beta_{(2)}} + O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{\sqrt{N}}) \\ &\xrightarrow{d} N\big(0, \Sigma_{\beta_{(2)}}\big),\end{aligned} \tag{A.23}$$

where the orders of the second and third terms can be derived similar to the above. Using (A.19) and (A.23), the joint asymptotic distribution of the CCEMG full-sample estimator and the CCEMG post-break estimator is derived. This completes the proof of Theorem 4.1. ∎

A.4 Proof of Theorem 4.2:

$$
\begin{aligned}
\rho(\widetilde{\beta}_{MG}, \mathbb{W}) &= \mathbb{E}\left[N\big(\widetilde{\beta}_{MG} - \beta_{(2)}\big)' \,\mathbb{W}\, \big(\widetilde{\beta}_{MG} - \beta_{i(2)}\big)\right] \\
&= N\,\mathbb{E}\left[\big(\widetilde{\beta}_{MG(2)} - \beta_{(2)}\big) - \alpha\big(\widetilde{\beta}_{MG(2)} - \widetilde{\beta}_{MG,Full}\big)\right]' \mathbb{W} \left[\big(\widetilde{\beta}_{MG(2)} - \beta_{(2)}\big) - \alpha\big(\widetilde{\beta}_{MG(2)} - \widetilde{\beta}_{MG,Full}\big)\right] \\
&= \rho(\widetilde{\beta}_{MG(2)}, \mathbb{W}) + \tau^2\,\mathbb{E}\left[\frac{1}{(\dot{Z}+\dot{\eta})'P(\dot{Z}+\dot{\eta})}\right] - 2\tau\,\mathbb{E}\left[\psi(\dot{Z}+\dot{\eta})'G\,\mathbb{W}\,G_2'\dot{Z}\right] \\
&= \rho(\widetilde{\beta}_{MG(2)}, \mathbb{W}) + \tau^2\,\mathbb{E}\left[\frac{1}{(\dot{Z}+\dot{\eta})'P(\dot{Z}+\dot{\eta})}\right] - 2\tau\,\mathbb{E}\,tr\left[\frac{\partial}{\partial(\dot{Z}+\dot{\eta})}\psi(\dot{Z}+\dot{\eta})'G\,\mathbb{W}\,G_2'\dot{V}\right] \\
&\leq \rho(\widetilde{\beta}_{MG(2)}, \mathbb{W}) - \tau\left[\frac{2\big(tr(\mathcal{A}) - 2\lambda_{max}(\mathcal{A})\big) - \tau}{\mathbb{E}(\dot{Z}+\dot{\eta})'P(\dot{Z}+\dot{\eta})}\right] \\
&\leq \rho(\widetilde{\beta}_{MG(2)}, \mathbb{W}) - \tau\left[\frac{2\big(tr(\mathcal{A}) - 2\lambda_{max}(\mathcal{A})\big) - \tau}{\dot{\eta}'P\dot{\eta} + tr(P\dot{V})}\right],
\end{aligned}
\tag{A.24}
$$

where the last inequality is Jensen's, $\psi(x) \equiv \frac{x}{x'Px}$, and thus $\frac{\partial}{\partial(x)}\psi(x)' = \big(\frac{1}{x'Px}\big)I - \frac{2Pxx'}{(x'Px)^2}$. Besides, using Stein's Lemma, Lemma 2 in Hansen (2016), we obtain

$$
\begin{aligned}
\mathbb{E}\left[\psi(\dot{Z}+\dot{\eta})'G\,\mathbb{W}\,G_2'\dot{Z}\right] &= \mathbb{E}\,tr\left[\frac{\partial}{\partial(\dot{Z}+\dot{\eta})}\psi(\dot{Z}+\dot{\eta})'G\,\mathbb{W}\,G_2'\dot{V}\right] \\
&= \mathbb{E}\left[\frac{tr\big(G\,\mathbb{W}\,G_2'\dot{V}\big)}{(\dot{Z}+\dot{\eta})'P(\dot{Z}+\dot{\eta})}\right] - 2\,\mathbb{E}\left[\frac{tr\big((\dot{Z}+\dot{\eta})'G\,\mathbb{W}\,G_2'\dot{V}P(\dot{Z}+\dot{\eta})\big)}{\big((\dot{Z}+\dot{\eta})'P(\dot{Z}+\dot{\eta})\big)^2}\right] \\
&= \mathbb{E}\left[\frac{tr\big(\mathcal{A}\big)}{(\dot{Z}+\dot{\eta})'P(\dot{Z}+\dot{\eta})}\right] - 2\,\mathbb{E}\left[\frac{tr\big((\dot{Z}+\dot{\eta})'\mathcal{B}_1'\mathcal{A}\mathcal{B}_1(\dot{Z}+\dot{\eta})\big)}{\big((\dot{Z}+\dot{\eta})'P(\dot{Z}+\dot{\eta})\big)^2}\right] \\
&\geq \mathbb{E}\left[\frac{tr\big(\mathcal{A} - 2\lambda_{max}(\mathcal{A})\big)}{(\dot{Z}+\dot{\eta})'P(\dot{Z}+\dot{\eta})}\right],
\end{aligned}
\tag{A.25}
$$

where $\mathcal{A} \equiv \mathbb{W}\,G_2'\dot{V}G$, $\mathcal{B}_1 = \mathbb{W}^{1/2}\,G'$, $\mathcal{B}_1'\mathcal{B}_1 = P$, and $\mathcal{B}_1'\mathcal{A}\mathcal{B}_1 \leq \mathcal{B}_1'\mathcal{B}_1\lambda_{max}(\mathcal{A})$ where $\lambda_{max}(\mathcal{A})$ denotes the maximum eigenvalues of $\mathcal{A}$. This completes the proof of Theorem 4.2. ■