The Incidental Parameters Problem in Testing for Remaining Cross-Section Correlation

Artūras Juodis\(^{a,b}\) and Simon Reese\(^{c}\)

\(^{a}\)Amsterdam School of Economics, University of Amsterdam, Amsterdam, Netherlands; \(^{b}\)Tinbergen Institute, Amsterdam, Netherlands; \(^{c}\)Department of Economics, Lund University, Lund, Sweden

**ABSTRACT**

In this article, we consider the properties of the Pesaran CD test for cross-section correlation when applied to residuals obtained from panel data models with many estimated parameters. We show that the presence of period-specific parameters leads the CD test statistic to diverge as the time dimension of the sample grows. This result holds even if cross-section dependence is correctly accounted for and hence constitutes an example of the incidental parameters problem. The relevance of this problem is investigated for both the classical two-way fixed-effects estimator and the Common Correlated Effects estimator of Pesaran. We suggest a weighted CD test statistic which re-establishes standard normal inference under the null hypothesis. Given the widespread use of the CD test statistic to test for remaining cross-section correlation, our results have far reaching implications for empirical researchers.

**ARTICLE HISTORY**

Received March 2020
Accepted March 2021

**KEYWORDS**

Cross-section dependence; Factor model; Panel data; Time fixed effects; U-statistic

1. **Introduction**

Given that economic agents rarely act entirely independently of each other, modeling cross-section dependence plays a prominent role in panel data econometrics. Time fixed effects are probably the simplest way of addressing this issue and allow controlling for a common trend whose effect is homogeneous across cross-section units. During the last decade, interactive fixed effects models have become a popular, more general alternative. They assume the presence of a factor error structure, that is, a small number of unobserved common trends interacted with entity-specific slope coefficients. Using either of the two modeling possibilities begs the question whether cross-section dependence is adequately accounted for.

In this article, we show that the application of tests for cross-section dependence to regression residuals obtained from two-way fixed-effects models or interactive effects models is problematic. We use the popular CD test statistic of Pesaran (2004, 2015a) and show that the inclusion of period-specific parameters introduces a bias term of order \(\sqrt{T}\). In order to avoid erroneous rejection of the null hypothesis of no unaccounted cross-section dependence, we suggest a modified test statistic that reweights cross-section covariances with Rademacher distributed weights. This weighted CD test statistic is asymptotically standard normal and has very good size under appropriate regularity conditions on the chosen weights.

The use of the CD test statistic as a misspecification test for regression models that already account for cross-section dependence has repeatedly been observed in empirical studies (see, e.g., Holly, Pesaran, and Yamagata 2010; Everaert and Pozzi 2014, Bailey, Holly, and Pesaran 2016; Eberhardt and Presbitero 2015 among others). In some applications, the CD test statistic is explicitly used as a model-selection tool, interpreting a reduction in the absolute value of the CD test statistic as an indication of a better model. In other cases, only specifications not rejected by the CD test statistic (given some significance level) were considered. However, to the best of our knowledge, there are currently no theoretical results in the literature that could justify this practice. Furthermore, the scenario of testing for cross-section dependence in cross-sectionally demeaned or defactored data is usually completely ignored in any of the large-scale Monte Carlo studies integral to theoretical and empirical articles in the field. Only very recently, Mao (2018) reported the size of three tests for cross-section dependence in a subset of his simulation experiments. However, despite clear evidence of excessive over-rejections, these results are neither discussed nor investigated theoretically. Therefore, this study is the first one to investigate the properties of cross-section dependence tests applied to residuals of models that control for sources of common variation across cross sections. In particular, given its popularity in the applied panel data literature, we restrict our attention to the CD test statistic of Pesaran (2004, 2015a). Our interest lies in residuals of models that characterize cross-section dependence as driven by a small number of unobservable factors. This comprises both the two-way fixed effects (2WFE) model and models with a multifactor error structure where factors are interacted with unit-specific slope coefficients. The results that we obtain are summarized as follows:

1. The application of the CD test to residuals obtained from a model where common factors enter either as time fixed effects or through a multifactor error structure renders the test statistic biased for any fixed \(T\), and divergent as \(T\to\infty\).

2. In addition to the mean of the CD test statistic, even its variance may be affected. This can result in an asymptotically degenerate distribution of the test statistic.
3. A simple way of eliminating bias is to construct the CD test statistic from specifically weighted cross-section covariances rather than correlations. This leads to a valid test statistic for remaining cross-section correlation with good small sample properties in simulations.

The degeneracy of the CD statistics can be seen as a manifestation of the incidental parameters problem (IPP) of Neyman and Scott (1948). In this respect, this article contributes to this branch of the literature. So far, the major focus in the panel data literature has been related to the IPP stemming from estimated individual specific effects. Our article is the first one to document an asymptotically nonnegligible impact of estimated time-specific common parameters which cannot be ruled out by restrictions on the relative expansion rate of time-specific common parameters which cannot be ruled out by.

Furthermore, since the CD statistic can be seen as a time-series average of second degree (degenerate) U-statistics, our results shed some light on the potential impact of the IPP beyond simple cross-section averages. Finally, while this article only considers linear models, the average correlation approach to testing for cross-section dependence was extended to nonlinear and nonparametric panel data models by Hsiao, Pesaran, and Pick (2012) and Chen, Gao, and Li (2012), respectively. Hence, problems documented in this article carry over to post-estimation properties of nonlinear models discussed in for example, Fernández-Val and Weidner (2016), Boneva and Linton (2017), or Chen, Fernández-Val, and Weidner (2021).

The remainder of this article is structured as follows: Section 2 introduces the testing problem. In Section 3, we present asymptotic results for models with two-way fixed effects and multifactor error structures. In Section 4, we discuss standard approaches for bias correction and propose a weighted CD test statistic that achieves this goal. Sections 5 and 6 illustrate the problem documented in this article by means of simulated and real data. Section 7 concludes. Additional technical discussions, Monte Carlo studies, and all proofs are relegated to the supplementary appendix.

Notation: $I_m$ denotes an $m \times m$ identity matrix and the subscript is sometimes disregarded for the sake of simplicity. $0$ denotes a vector of zeros while $\mathbf{O}$ stands for a matrix of zeros. $s_m$ denotes a selection vector all of whose elements are zero except for element $m$ which is one. $t$ is a vector entirely consisting of ones. The dimension of these latter vectors and matrices is generally suppressed for the sake of simplicity and needs to be inferred from context. For a generic $m \times n$ matrix $A$, $P_A = A(A'A)^{-1}A'$ projects onto the space spanned by the columns of $A$ and $M_A = I_m - P_A$. Furthermore, $\text{rk}(A)$ denotes the rank of $A$, $\text{tr}(A)$ its trace and $\|A\| = (\text{tr}(A'A))^{1/2}$ the Frobenius norm of $A$. For a set of $m \times n$ matrices $\{A_1, \ldots, A_N\}$, $\mathcal{A} = N^{-1} \sum_{i=1}^{N} A_i$, $\delta$ and $M$ stand for a small and large positive real number, respectively. For two real numbers $a$ and $b$, $a \vee b = \max(a, b)$. Lastly, $\mathcal{O}_p(\cdot)$ and $o_p(\cdot)$ express stochastic order relations.

2. The Testing Problem

Let $z_i$ be a $T$-dimensional data vector observed over $N$ cross-section units indexed by $i$. Combining all $z_i$ we obtain a two-dimensional data array of panel data (or longitudinal data). In empirical research, it is common to investigate whether $z_i$ can be regarded as independent over $i$ in order to select a model that can properly characterize the statistical properties of the data. In particular, researchers might be interested in the statistical hypothesis

$$H_0 : z_i \perp z_j, \quad \text{for all } i, j = 1, \ldots, N,$$

where we use the $\perp$ notation to denote independence. Most often $z_1, \ldots, z_N$ contain residuals obtained from a regression model that does not allow for cross-section dependence, for example, an entity fixed-effects model or linear regression.

By far the most widely used test for cross-section dependence (correlation, to be precise) is the CD test of Pesaran (2004, 2015a), which is based on a simple rescaled sum of all pairwise cross-section correlation coefficients, formally denoted

$$\text{CD} = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \hat{\rho}_{ij} = \sqrt{\frac{TN(N-1)}{2}} \hat{\rho}. \quad (2)$$

Here,

$$\hat{\rho}_{ij} = \frac{T^{-1} \sum_{t=1}^{T} (z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j)}{\sqrt{T^{-1} \sum_{t=1}^{T} (z_{it} - \bar{z}_i)^2} \sqrt{T^{-1} \sum_{t=1}^{T} (z_{jt} - \bar{z}_j)^2}} \quad (3)$$

is the pairwise sample correlation coefficient between units $i$ and $j$. Obviously, computing the CD test statistic involves obtaining $N(N-1)/2$ parameter estimates, each of which converges to the true parameter value at rate $\sqrt{T}$ only. These circumstances are reminiscent of the panel data setup considered in, for example, Phillips and Moon (1999) or Hahn and Kuersteiner (2002), where estimation of many (incidental) parameters in linear regression models turns out to have distributional effects on the asymptotic properties of common parameters. That is, they cause the incidental parameter problem. By contrast, the asymptotic distribution of the CD test statistic is unaffected by the estimation of all $N(N-1)/2$ cross-section correlation coefficients involved in its construction. In fact, applied to the residuals of a linear regression model with strictly exogenous regressors and individual specific means, the CD test statistic is asymptotically normal as long as $N, T \to \infty$ (see Pesaran 2015b, Theorem 2), under $H_0$ of independence (or even limited local dependence). However, as shown below, this result does not hold when the model specification includes period-specific parameters.

2.1. Heuristic Discussion of the Main Result

Consider a linear model where cross-section dependence is due to time fixed effects, so that unobserved heterogeneity over both cross-sections and time enters the model additively. That is, the relation between the $T \times 1$ vector $y_i$ and the $T \times m$ matrix $X_i$ is formally denoted

$$y_i = X_i \beta + \tau + \mu_i + \varepsilon_i, \quad i = 1, \ldots, N. \quad (4)$$

Here, $\mu_i$ and $\tau$ denote an entity-specific intercept and a $T \times 1$ vector of time-specific common parameters $\tau$, respectively. $\varepsilon_i$ is a vector of idiosyncratic error terms, independent across cross-section units. In the example of a difference-in-differences framework, $\tau$ is the common trend affecting both treated and
untreated individuals, while the treatment indicator as well as other covariates are contained in \( X_i \). For the sake of simplicity, we assume \( \beta \) and \( \mu_i \) to be known, so that \( \beta = 0 \) and \( \mu_i = 0 \) \( \forall i \) can be imposed without loss of generality. This highly restrictive assumption leaves the leading terms in the analysis below unaffected and is hence innocuous for the expository purpose of this section. Model (4) reduces to

\[
y_{i,t} = \tau_i + \varepsilon_{i,t}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{5}
\]

We additionally assume that the variance of \( \varepsilon_{i,t} \) is known and fixed to \( \sigma^2 = 1 \). In this setup, \( y_{i,t} \) are clearly cross-sectionally dependent because of \( \tau_i \); however, we are interested in testing whether \( \varepsilon_{i,t} \) are cross-sectionally uncorrelated. Given that \( \varepsilon_{i,t} \) are unobserved, the common effects \( \tau_i \) need to be estimated. The most natural approach is to estimate \( \tau_i \) by OLS so that \( \hat{\tau}_i = \bar{y}_i = \frac{1}{N} \sum_{i=1}^{N} y_{i,t} \). Using the residuals

\[
\hat{\varepsilon}_{i,t} = y_{i,t} - \hat{\tau}_i, \tag{6}
\]

the CD test statistic is given by

\[
CD = \sqrt{\frac{2}{TN(N-1)}} \sum_{i=1}^{T} \sum_{j=1}^{N} \sum_{i=1}^{N-j} \hat{\varepsilon}_{i,t} \hat{\varepsilon}_{j,t} = \sqrt{\frac{1}{2TN(N-1)}} \sum_{i=1}^{T} \left( \sum_{i=1}^{N} \hat{\varepsilon}_{i,t} \right)^2 - \sqrt{\frac{1}{2TN(N-1)}} \sum_{i=1}^{T} \sum_{i=1}^{N} \hat{\varepsilon}_{i,t}^2. \tag{7}
\]

Given this definition,

\[
\sum_{i=1}^{N} \hat{\varepsilon}_{i,t} = \sum_{i=1}^{N} y_{i,t} - \sum_{i=1}^{N} y_{i,t} = 0, \tag{8}
\]

implying that the first term in Equation (7) cancels out. The expression \( \sum_{i=1}^{T} \sum_{i=1}^{N} \hat{\varepsilon}_{i,t}^2 \) in the second term is nothing more than the sum of squared residuals (SSR) of the estimated model, a statistic that is of order \( O_p(NT) \). Consequently,

\[
CD = O_p(\sqrt{T}), \tag{9}
\]

even though the error-terms \( \varepsilon_{i,t} \) are cross-sectionally independent. Hence, the procedure commonly used by practitioners is prone to finding spurious cross-section dependence in the data.

The results shown for a model with additive time effects and homoscedastic errors are not coincidental. In fact, as we show later in this article, they carry over to a more complex model where cross-section dependence is generated by a multifactor error structure and/or where the errors are heteroscedastic.

### 3. Asymptotic Results

In this section, we provide formal asymptotic results for the CD statistic based on the residuals obtained from a 2WFE model or a model with multifactor error structure. For the sake of simplicity, we continue to assume that \( \beta \) is known throughout this section. If the vector of slope coefficients \( \beta \) were unknown, its estimator would converge (under correct model specification) to the true parameter value at the conventional rate \( \sqrt{NT} \). Hence, convergence to the true parameter values is fast enough to ensure that any terms present in the CD test statistic involving \( \beta - \beta_0 \) are asymptotically negligible. We shall not attempt to prove this claim formally but rely on Monte Carlo results of Section 5 to support this conjecture. By contrast, the documented effects on the first two moments of the CD test statistic resulted from the fact that estimates of common, period-specific parameters converge at a rate slower than \( \sqrt{NT} \).

#### 3.1. Two-Way Error Component Structure

As in Section 2.1 we assume that the true model is given by Equation (4). Additionally, we make the following assumptions on the model errors.

**Assumption 1 (Errors).** 1. Let \( \varepsilon_{i,t} = \sigma_i \eta_{i,t} \) where \( \eta_{i,t} \) is independently and identically distributed across both \( i \) and \( t \) with \( E[\eta_{i,t}] = 0 \), \( E[\eta_{i,t}^2] = 1 \) and \( E[|\eta_{i,t}|^8] < M \) for some \( M < \infty \).

2. \( \sigma_i \) is defined over an interval \( [\delta, M] \) with \( 0 < \delta < M < \infty \). It is independently and identically distributed across \( i \), and \( \sigma_i \) is independent of \( \eta_{i,t} \) for all \( i \) and \( t \).

For technical reasons, we assume that all stochastic variables in this article have finite eighth moments. This is a sufficient condition, which facilitates proving joint convergence of the test statistics considered in this article, see also Demetrescu and Homm (2016). Assumption (1) is general enough to cover several models of conditional heteroscedasticity in \( \varepsilon_{i,t} \). Alternatively, this assumption can be formulated in terms of unconditional variances, where \( \sigma_i \) are fixed numbers. However, in addition to having severe conceptual shortcomings (as discussed in Gagliardini, Ossola, and Scaillet 2016) such an approach would lead to incorrect conclusions concerning the power of a CD test statistic for remaining cross-section correlation. See Section S.3 in the supplementary appendix for a detailed discussion on power.

Natural examples for \( \sigma_i \) are either a standard exponential skedastic function

\[
\sigma_i = \exp(\alpha + \gamma \mu_i), \tag{10}
\]

or a location-scale model with

\[
\sigma_i = \alpha + \gamma \mu_i. \tag{11}
\]

Both satisfy the required restrictions, as long as \( \mu_i \) has a bounded support. Furthermore, we denote different cross-section averages of \( \sigma_i \) by \( \hat{\sigma}^T = N^{-1} \sum_{i=1}^{N} \sigma_i^k \) for \( k \in \mathbb{Z} \), and the corresponding population quantities by \( E[\sigma_i^k] \). Assumption 1 guarantees that these quantities are well defined for all finite \( k \).

Using these definitions, the CD test statistic obtained from a model with unknown period-specific effects can be characterized as follows:
Theorem 1. Suppose that $\beta$ is known. Under Assumption 1,

$$CD = \sqrt{\frac{2}{TN(N-1)}} \sum_{i=2}^{N} \sum_{j=1}^{T} \sum_{i=1}^{T} \varepsilon_{ti} \varepsilon_{tj} \left( \sigma^{-1}_{ij} - \sigma^{-1} \right),$$

$$\left( \sigma^{-1}_{ij} - \sigma^{-1} \right) + \sqrt{T} \varepsilon_i + \mathcal{O}(\sqrt{NT^{-1}}) + \mathcal{O}(T^{-1/2})$$

$$+ \mathcal{O}(P(N^{-1/2}) + \mathcal{O}(P(N^{-1/2})),$$  \hspace{1cm} (12)

where

$$\varepsilon = \sqrt{\frac{N}{2(N-1)}} \sum_{i=1}^{N} \sum_{j=1}^{T} \varepsilon_{ij} \left( \left( \sigma^{-1} \right)^2 - 2 \sigma^{-1} \sigma^{-1} \right).$$  \hspace{1cm} (13)

Furthermore, let $\Omega = \left( 1 - 2E[\sigma_i]E[\sigma^{-1}_i] + E[\sigma^2_i](E[\sigma^{-1}_i])^2 \right)^2$. Then,

$$CD - \sqrt{T} \varepsilon \overset{d}{\longrightarrow} N(0, \Omega),$$  \hspace{1cm} (14)

as $N, T \to \infty$ jointly provided that $\sqrt{T} N^{-1} \to 0$ and $\sqrt{T} N^{-1} \to 0$.

Theorem 1 shows that the inclusion of time fixed effects into the model specification has an asymptotically nonnegligible effect on the first two moments of the CD test statistic. Via expansion with appropriate functions of $\sigma_i$, it can be shown that $\varepsilon$ indicates the presence of a deterministic bias of order $\sqrt{T}$. Abstracting from this bias, Equation (14) is dominated by an expression that reflects the CD test statistic obtained from the true model errors but imposing an incorrect normalization.

Restrictions on the expansion rates of $N$ and $T$, that is, $\sqrt{NT^{-1}} \to 0$ and $\sqrt{T} N^{-1} \to 0$, are satisfied if one considers diagonal asymptotic expansion schemes as in Fernández-Val and Weidner (2016), where $NT^{-1} \to \kappa^2$ with $\kappa \in (0; \infty)$. This way $N/T + T/N$ remains a bounded constant asymptotically. The condition $\sqrt{NT^{-1}} \to 0$ exceeds requirements for distributional results of the pooled CCE estimator in linear regression models (see Pesaran, 2006, Theorem 4) and results from the fact that the feasible CD statistic is a nonlinear function in $\hat{\sigma}_i$. Hence, the rate restrictions derived for general nonlinear panel data models, as studied by Fernández-Val and Weidner (2016), are naturally applicable.

The behavior of the CD test statistic in a model with time fixed effects can be seen as an example of the incidental parameters problem (IPP) of Neyman and Scott (1948), since the bias of the CD test statistic is due to the estimation of $T$ period-specific intercepts $\tau_t$. Interestingly, in the context of estimating linear dynamic panel data models, estimation of the time effects $\tau_t$ does not introduce any asymptotic bias into the FE estimator with strictly or weakly exogenous regressors (see, e.g., Hahn and Moon 2006). In nonlinear models, estimation of the time effects $\tau_t$ affects the asymptotic mean of the estimator for slope parameters associated with explanatory variables (see, e.g., Fernández-Val and Weidner 2016) with the corresponding bias being proportional to $\sqrt{T} N^{-1}$. In this sense, our result adds new insights into the literature in that it highlights a scenario where the inclusion of time fixed effects into a linear model has nonstandard implications for the asymptotic distribution of the statistic of interest.

The results of Theorem 1 suggest that asymptotically standard normal inference can be recovered by bias-correcting and rescaling (12). Before considering this remedy in Section 4, it is important to emphasize the following special case:

**Corollary 1.** Under Assumption 1 and given $P(\sigma_i = \sigma) = 1,$

$$CD = -\sqrt{\left( \frac{T - T}{N} \right) / 2} + \mathcal{O}(R_{N,T}),$$  \hspace{1cm} (15)

where $R_{N,T} = (N^{-1/2} \vee T^{-1/2} \vee N^{-1} \sqrt{T} \vee T^{-1} \sqrt{N})$.

Corollary 1 provides more intuition about the approximate value of the bias term $\sqrt{T} \varepsilon$, suggesting that it should be reasonably close to $-\sqrt{T}/2$. More importantly, the result reveals that the leading stochastic component in the CD test statistic, the first term on the right-hand side of Equation (12), cancels out when error variances are homogeneous across $i$. Instead, random variation around the bias term $-\sqrt{T}/2$ is of order $\mathcal{O}(P(N^{-1/2} \vee T^{-1/2}))$, rendering the distribution of a modified version of the test statistic asymptotically degenerate.

The special case of homogeneous error variances hence entails consequences for the CD test statistic that are qualitatively different from those in the more general case where $\sigma_i$ may differ across $i$. Again, it would be possible to allow for asymptotically normal inference by adequately rescaling the modified test statistic. However, the resulting statistic would be of little practical use since the main source of variation is not related to error covariances across cross-sections, but is simply driven by variance of the idiosyncratic components.

### 3.2. Multifactor Error Structure

Following Pesaran (2006), we consider a model where cross-section dependence enters the model via a multifactor error structure. This model is described by the $T \times 1$ vector $y_i$ and the $T \times m$ matrix $X_i$ defined as follows:

$$y_i = X_i \beta + F \xi_i + \varepsilon_i,$$  \hspace{1cm} (16)

$$X_i = FA_i + E_i,$$  \hspace{1cm} (17)

where $F = \left[ f_1, f_2, \ldots, f_T \right]'$ denotes a $T \times r$ matrix of unobserved common factors and where $E_i = \left[ e_{i1}, e_{i2}, \ldots, e_{iT} \right]'$ is a $T \times m$ matrix of idiosyncratic variation in $X_i$. The most popular estimator designed to estimate the parameter vector $\beta$ in this specific model is the common correlated effects (CCE) estimator of Pesaran (2006) which amounts to augmenting a linear regression model with cross-section averages of potentially all variables available to the researcher in order to account for the effect of unobservable common factors. In a model with homogeneous slope coefficients, we have

$$\hat{\beta}_{CCE} = \left( \sum_{i=1}^{N} X_i' M \hat{F} X_i \right)^{-1} \left( \sum_{i=1}^{N} X_i' M \hat{F} y_i \right)$$  \hspace{1cm} (18)

where

$$\hat{F} = \left[ \hat{y}, \hat{X} \right] = \left( F \hat{\lambda}, \hat{\lambda} \right) + [\varepsilon, \quad E] \left[ \begin{array}{c} 1, \quad 0 \end{array} \right] = FC + U.$$  \hspace{1cm} (19)
Here, \(\mathcal{C}(\mathcal{U})\) are defined implicitly in terms of \(\beta\) and the corresponding average factor loadings (average errors). While the CCE estimator is agnostic about the true number of factors that affect the data, it has been shown that consistent estimation of the parameters of interest requires that the number of cross-section averages is at least as large as the true number of factors that drive the data (Westerlund and Urbain 2013).

Given that the factor estimator \(\hat{F}\), defined in Equation (19), is constructed based on observed data, restrictions on the DGP of both \(y_{it}\) and \(x_{it}\) need to be imposed.

Assumption 2. 1. Let \(\varepsilon_{it} = \sigma_i \eta_{it}\) where \(\eta_{it}\) is independently and identically distributed across both \(i\) and \(t\) with \(E[\eta_{it}] = 0\) and \(E[||\eta_{it}||^4] < M\) for some \(M < \infty\).

2. \(\sigma_i\) is defined over an interval \([\delta_i, M]\) with \(0 < \delta < M < \infty\).

It is independently and identically distributed across \(i\), and \(\sigma_i\) is independent of \(\eta_{it}\) for all \(i\) and \(t\).

3. The \(m \times 1\) random vector \(e_{it}\) is independently distributed across both \(i\) and \(t\) with \(E[e_{it}] = 0\), \(E[e_{it}e_{it}'] = \Sigma\) with the latter being a positive-definite matrix and \(E[||e_{it}||^4] < M\).

Assumption 3. \(f_i\) is a covariance stationary \(r \times 1\) random vector with positive-definite covariance matrix \(\Sigma_F\), absolutely summable autocovariances and \(E[||f_i||^4] < M\).

Assumption 4. \(\lambda_i\) is iid over \(i\) with \(E[\lambda_i] = \mu_\lambda\) and \(E[||\lambda_i||^4] < M\). Furthermore, \(A_i\) is iid over \(i\) with \(E[A_i] = \mu_A\) and \(E[||A_i||^4] < M\).

Assumption 5. \(f_{i'}, \{\lambda_i, A_i, \sigma_i\}, \eta_{i't}, e_{i't}, e'_{i't'}\) are mutually independent for all \(i, i', t, t'\) and \(t''\).

Assumption 6. \(rk((\mu_\lambda, \mu_A)) = r = m + 1\).

This set of assumptions above are a slightly more restrictive version of the framework considered in Pesaran (2006). For example, the assumption of common \(\Sigma\) can be straightforwardly relaxed. However, unlike \(\sigma_i\), \(\Sigma\) plays no major role for asymptotic results of this article. Most importantly, we rule out the presence of serial correlation as this is in line with the assumptions made for the CD test to work. Furthermore, we restrict ourselves to a classical panel data regression model instead of considering heterogeneous slope coefficients. Moreover, any dependence between \(\varepsilon_{it}\) and \(e_{it}\) is assumed away in order to allow for a tractable proof of the main theoretical result in this article. Lastly, the fact that we assume \(rk((\mu_\lambda, \mu_A)) = r = m + 1\) to hold suggests that we consider an ideal setup where none of the rank condition-related problems documented in Karabiyik, Reese, and Westerlund (2017), or Juodis, Karabiyik, and Westerlund (2021) apply.

In analogy with the previous sections, we assume that \(\beta\) is known. Thus, we disregard the estimation error \(\hat{\beta}_{CCE} - \beta\), which is generally of order \(O_p(NT^{-1/2}) + O_p(N^{-1})\), see, for example, Pesaran (2006) and Juodis, Karabiyik, and Westerlund (2021).

We begin our asymptotic analysis, by noting that in the model with assumed (known) homogeneous \(\sigma\) the result follows directly as in the model with time effects only. In particular, while it is not generally emphasized in the CCE literature, the residuals from CCE estimation which are formally given by

\[
\tilde{e}_i = M_F y_i = e_i - P_F e_i - (P_F - P_T)F \lambda_i
\]

satisfy

\[
\sum_{i=1}^{N} \tilde{e}_{i,t} = 0.
\]

In that respect, the standard 2WFE estimator is similar to the CCE estimator. More formally we formulate the following result

**Proposition 1.** Under Assumptions 2–6 and \(P(\sigma_i = \sigma) = 1\):

\[
CD = -\sqrt{\frac{T}{N}} \Omega_{R_N,T} + O_p(R_{N,T}),
\]

where \(R_{N,T} = (N^{-1/2} \times T^{-1/2} \times N^{-1/2} \times T^{-1/2} \times N^{-1})\).

Proposition 1 shows that the result we derived previously for the 2WFE estimator in Corollary 1 continues to hold for models with a factor error structure, as long as \(N \to \infty\) and \(T \to \infty\).

It is worth mentioning that the above order effect is only valid if \(\hat{F}\) contains cross-section averages of the regressand as well as all regressors. If either of those variables is omitted (without affecting the rank condition in Assumption 6), the zero mean residual condition in Equation (20) is violated, and consequently the CD test will have a \(O_p(1)\) term. However, this result is of limited empirical importance as in most cases researchers include all available cross-section averages. While this practice ensures that the estimator is invariant to \(\beta\), inclusion of too many cross-section averages can potentially have detrimental effects on the asymptotic properties of the estimator, see the corresponding discussion in Juodis, Karabiyik, and Westerlund (2021) and Juodis (2020).

In the homoscedastic case, two-way fixed effects and multifactor error models have similar asymptotic effects on CD test statistic. The conclusions of the next theorem, which is the main result of this article, indicate that this equivalence does not hold in the heteroscedastic case. In order to proceed, we introduce some useful notation. For \(t = 1, \ldots, T\), let \(u_{i,t} = [\varepsilon_{i,t} + \beta' e_{i,t}, e_{i,t}']\), and

\[
(C)^{-1} f_i - f_t = \frac{1}{N} \sum_{i=1}^{N} \psi_{i,t},
\]

where \(\psi_{i,t} = (C)^{-1} u_{i,t}\) is the influence function of the corresponding factor estimator, in this case cross-section averages of \(y_{i,t}\) and \(x_{i,t}\). Generally, the influence function depends on the joint process \([y_{i,t}, x_{i,t}]\), as long as all observed variables are used to form cross-section averages. Equipped with this notation we formulate the main result of this article.

**Theorem 2.** Suppose that \(\beta\) is known. Under Assumptions 2–6,

\[
CD = \sqrt{\frac{2}{TN(N-1)}} \sum_{i=1}^{N} \sum_{j=1}^{T} \xi_{iN,t} \xi_{jN,t} + \sqrt{T} \Phi_1
\]

\[
- 2\sqrt{T} \Phi_2 + O_p(R_{N,T}),
\]
where
\[ \hat{\xi}_{\text{N},t} = \sigma_i^{-1} \xi_{\text{N},t} \] (25)
and
\[ \Phi_1 = \sqrt{\frac{N}{2(N-1)}} \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_i^{-1} \xi_i \right) \left( \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \xi_{i,t} \psi_{i,t} \right) \] (26)
\[ \Phi_2 = \sqrt{\frac{N}{2(N-1)}} \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_i^{-1} \xi_i \right) \left( \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \psi_{i,t} \sigma_i^{-1} \xi_{i,t} \right). \] (27)

Here, as previously, \( R_{\text{N,T}} = (N^{-1/2} \sqrt{T^{-1/2}} \sqrt{N^{-1} \sqrt{T} \sqrt{T^{-1}} \sqrt{N}} \). In line with all previous results, the CD test statistic in this setup has two diverging components. However, unlike all previous results, in particular Theorem 1, these bias terms are not solely non-linear functions of \( \xi_i \). Instead, they are also influenced by the first (rescaled) moments of factor loadings in \( y_{i,t} \) and \( x_{i,t} \), as well as corresponding variances of the idiosyncratic components in \( x_{i,t} \). Thus, the influence function \( \psi_{i,t} \) directly alters distributional properties of the CD statistic. The result is qualitatively similar to any parametric two-step estimation procedure with a plug-in first-step estimator. Thus, by including cross-section averages as factor proxies, one is implicitly testing that both \( \hat{\xi}_{\text{N},t} \) and \( \psi_{i,t} \) are jointly cross-sectionally uncorrelated. Notice, that while \( \hat{\xi}_{\text{N},t} \) is usually assumed to be uncorrelated over \( t \) (e.g., for models with pre-determined regressors), the same is not true for \( u_{i,t} \), for example, if \( x_{i,t} = y_{i,t-1} \). Thus, if Assumption 2 is appropriately relaxed, then \( \hat{\xi}_{\text{N},t} \) will be serially correlated.

**Remark 1.** One can easily see that expressions for \( \hat{\xi}_{\text{N},t}, \Phi_1, \Phi_2 \) are the same if one assumes that \( \lambda_i \) is known. Thus, it is only the influence function of common factors, and not those of factor loadings, that has an impact on the asymptotic properties of the test statistic. This conclusion is the same as in the two-way error components model.

As alluded in Proposition 1, the leading \( O_p(1) \) term is degenerate if \( \sigma_i = \sigma \), as in this case:
\[ \sum_{i=1}^{N} \xi_{\text{N},t} = 0, \quad \forall t = 1, \ldots, T. \] (28)

One can easily see that the expressions for \( \Phi_1 \) and \( \Phi_2 \) derived for CCE coincide with the corresponding terms of Theorem 1 (up to a negligible remainder term), upon setting \( \lambda_i = 1 \) (thus \( \chi_i^{-1} = 1 \)), and \( \psi_{i,t} = \hat{\xi}_{i,t} \). This situation is equivalent to using \( \hat{f}_i = y_i - \bar{y}_i \beta_0 \) as factor proxies, for example, as suggested by Westerlund, Karabiyik, and Narayan (2017) in the context of predictability testing with cross-section dependence.

Recall that the result in Theorem 2 considers only the CCE setup where the number of observable factor proxies equals the number of the true factors. If Assumption 6 is relaxed and there are more observables than factors, then following Karabiyik, Reese, and Westerlund (2017) and Juodis, Karabiyik, and Westerlund (2021), one can show that the expressions for \( \Phi_1 \) and \( \Phi_2 \) will contain additional terms related to this discrepancy. In particular, \( \Phi_1, \Phi_2, \) and \( \hat{\xi}_{\text{N},t} \) are functions of an unknown rotation matrix, which cannot be consistently estimated from the data. However, as already emphasized by Juodis, Karabiyik, and Westerlund (2021), this problem can be circumvented via the use of a non-parametric bias-correction method. We will come back to this issue in Section 4.1.

### 4. Re-Establishing Standard Normal Inference

The diverging bias in the CD test statistic applied to residual from models with common, period-specific parameters is fundamentally problematic for its use in the context of testing for remaining cross-section correlation. Still, the literature review in Section 1 suggests that the underlying question of correct model specification is of high relevance for empirical researchers. For this reason, it is relevant not to discard completely the CD test statistic but instead to discuss possible modifications aimed at ensuring an asymptotically standard normal inference under the null hypothesis. Thus, methods aimed at addressing the IPP detailed in Sections 2 and 3 need to be considered.

#### 4.1. Analytic Bias Correction

Given that parametric expressions for the bias of CD have been derived in Theorems 1 and 2, analytic bias correction is feasible. As detailed in Section S.4.1 in the supplementary appendix, estimates of the unknown model components that constitute either \( \Xi \) or \( \Phi_1 \) and \( \Phi_2 \) can be obtained in order to eliminate the diverging component in CD. In addition, Propositions S.1 and S.2 in the supplementary appendix suggest that these estimated bias terms would eliminate equivalent terms \( \Xi_{H_{1,1}} \) or \( \Phi_{H_{1,1}} \) and \( \Phi_{H_{1,2}} \) under the alternative hypothesis without necessarily reducing the rate at which CD diverges.

Still, problems with its implementation in practice lead us to discard analytic bias correction and to consider it merely as a benchmark approach when evaluating our weighted CD test statistic, as introduced below, in Monte Carlo experiments. In particular, the construction of plug-in estimates of \( \Phi_1 \) and \( \Phi_2 \) proves to be tedious since both terms are constructed from estimates of \( \sigma_i^{2,1}, \ldots, \sigma_i^{2,N}, \lambda_1, \ldots, \lambda_N \) and \( \beta \). Additionally, estimation of the cross-section sum \( N^{-1} \sum_{i=1}^{N} \sigma_i^{-1} \lambda_i \) poses a tradeoff between generality and accuracy of the plug-in estimate. To be specific, independence between factor loadings \( \lambda_i \) and error variances \( \sigma_i^2 \) needs to be imposed to ensure that estimation error in \( \hat{\lambda}_i \) and \( \hat{\sigma}_i^2 \) does not dominate the small-sample distribution of the bias-corrected version of CD.

The need to impose further assumptions in order to improve the accuracy of the bias estimates is effectively a consequence of the fact that the unknown bias itself diverges at rate \( \sqrt{T} \). This means that any error in the estimation of the two aforementioned components is scaled by a factor that increases as \( N, T \to \infty \). Still, despite improving the approximation of plug-in estimates for \( \Phi_1 \) and \( \Phi_2 \) the assumption of independence between factor loadings and error variances can lead to size
4.2. Weighted Cross-Section Covariances

The method for bias correction favored in this article is to construct a CD test statistic from estimated cross-section covariances that are weighted with individual-specific, random draws $X$ from a Rademacher distribution, that is, $P(X = 1) = P(X = -1) = 0.5$. This approach is based on noticing that the cross-section correlation estimator $\hat{\rho}_{ij}$, as defined in (3) with $z_{it} = \hat{\epsilon}_{it}$, is merely a unit-specific weighting of $\text{cov} \left[ \hat{\epsilon}_{it}, \hat{\epsilon}_{jt} \right]$ with advantageous properties: Under the assumptions made in Pesaran (2004), it ensures that the CD test statistic has unit variance under the null hypothesis of no cross-section correlation, allowing for standard normal inference without the need to obtain a variance estimate. However, the case is different in models with unknown, period-specific parameters $\tau_i$ or $f_t$. As shown in Theorems 1 and 2, studentization of the model residuals fails to ensure a test statistic with a variance of one. This justifies the use of an alternative weighting scheme which we construct as a weighted average of individual-specific covariances,

$$\sqrt{\frac{2}{TN(N-1)}} \sum_{i=1}^{T} \sum_{j=2}^{N} \sum_{i=1}^{N-i+1} (w_i \hat{\epsilon}_{ij})(w_j \hat{\epsilon}_{ij}),$$

for some set of weights $w_1, \ldots, w_N$. Notice that Equation (29) would coincide with the CD test statistic of Pesaran (2004) for $w_i = \hat{\sigma}_i^{-1} \forall i$. This is not the case we consider here. Instead we make the following additional assumption:

**Assumption 7 (Weights).** $w_1, \ldots, w_N$ are identically and independently Rademacher distributed. Furthermore, $w_i$ is independent of $\lambda_i, \sigma_i, \eta_{it}$ and $e_{it}$ for all $i$ and $t$.

Rademacher distributed weights amount to random sample splitting, a method for breaking dependence that is not new to econometrics. For example, Altonji and Segal (1996, p. 358) already considered it an old concept. Its effects are most obvious in a simple modification of Theorem 2 where we replace $\sigma_i^{-1}$ with $w_i$. Under Assumption 7, the expected value of $N-1 \sum_{i=1}^{N} \lambda_i w_i$ can be conveniently split into the expectations of its two constituents. Given the zero expected value of $w_i$, an appropriate LLN applies and leads the cross-section average of $\lambda_i w_i$ to converge to zero. In an even simpler form, the same reasoning applies to the cross-section averages involving $w_i$ in an adapted version of Theorem 1. In both cases, this reduces the order of the leading bias components $\Xi$, $\Phi_1$, and $\Phi_2$ by a factor of $N$ and entails asymptotic unbiasedness of the weighted CD test statistic, subject to the restriction $\sqrt{TN}^{-1} \rightarrow 0$.

Additional rescaling with the asymptotic variance of expression (29) results in a weighted CD test statistic which, as formally stated in Theorem 3, converges to a standard normal distribution.

**Theorem 3.** Consider the weighted CD test statistic

$$CD_W = \left( \frac{1}{NT} \sum_{i=1}^{T} \sum_{j=2}^{T} \sum_{i=1}^{N} (w_i \hat{\epsilon}_{ij})(w_j \hat{\epsilon}_{ij}) \right)^{-1} \left( \sqrt{\frac{2}{TN(N-1)}} \sum_{i=1}^{T} \sum_{j=2}^{N} \sum_{i=1}^{N-i+1} (w_i \hat{\epsilon}_{ij})(w_j \hat{\epsilon}_{ij}) \right),$$

and assume that either of the following two points hold.

1. The data are generated by the time fixed-effects model (4) such that Assumption 1 holds, $\hat{\epsilon}_{it}$ is given by Equation (6).
2. The data are generated by the latent common factor model (16) such that Assumptions 2–6 hold. $\hat{\epsilon}_{it}$ is defined by Equation (20).

Under either of the two sets of assumptions above as well as Assumption 7, it holds that

$$CD_W \overset{d}{\rightarrow} N(0,1),$$

as $N, T \rightarrow \infty$ jointly subject to the restriction $\sqrt{T}N^{-1} \rightarrow 0$.

As stated by Theorem 3, the use of independent Rademacher distributed weights, analogous to weights drawn from many other distributions with zero mean, re-establishes asymptotic standard normal inference under the null hypothesis of the CD test statistic. However, asymptotic unbiasedness of $CD_W$ comes at the cost of power. More specifically, our approach to bias correction centers the leading components of $CD_W$ around zero, irrespective of whether cross-section correlation in the data is completely controlled for or not. As a consequence, only increases in the variance of $CD_W$ under its alternative hypothesis lead to power against the null hypothesis of this test. We further discuss this point in Section S.3 of the supplementary appendix.

To improve the power properties of $CD_W$, we suggest a refinement of this test which follows the power enhancement approach of Fan, Liao and Yao (2015). The authors suggest improving the power of high-dimensional cross-section tests by adding to the test statistic of interest a screening statistic. This screening statistic is equal to zero with probability approaching one under the null hypothesis of the test, but diverges at a fast rate under the alternative. In our case we choose the absolute sum of thresholded cross-section correlation coefficients. This results in a power-enhanced weighted CD test statistic which is defined as

$$CD_{W+} = CD_W + \sum_{i=1}^{N} \sum_{j=1}^{i-1} |\hat{\rho}_{ij}| 1 \left( |\hat{\rho}_{ij}| > 2\sqrt{\ln(N)/T} \right),$$

where $\hat{\rho}_{ij}$ is as in (3) with $z_{it} = \hat{\epsilon}_{it}$ and where $1(A)$ is the indicator function for event $A$. The screening statistic on the right-hand side of (32) has an asymptotically negligible effect on the size of $CD_{W+}$ because individual cross-section correlation coefficients are still consistent for their true value under $H_0$. For additional discussion on $CD_{W+}$, we refer to Section S.3 of the supplementary appendix.
Remark 2. An approach to reducing the dependence of $CD_W$ on a specific set of random weights would consist of averaging several weighted $CD$ test statistics. Denote by $CD_W^{(g)}$ the $CD$ test statistic obtained for a given draw of $N$ Rademacher distributed weights, the latter being indexed by $g$. For a total of $G$ different draws, an averaged weighted $CD$ test can be constructed as

$$\overline{CD_W} = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} CD_W^{(g)},$$

where the total number of draws $G$ should be chosen sufficiently small (e.g., $G = 30$) to avoid size distortions that may arise from scaling up lower-order terms in $CD_W^{(g)}$.

Remark 3. An initially appealing alternative to external random numbers as weights for a bias-corrected $CD$ test statistic would be a set of $N$ statistics derived from the dataset available to the researcher. When considering such internal (sample-specific) weights, it is desirable to opt for functions of the data that are optimal in the sense that they maximize the rate at which the $CD$ test statistic of a given draw of $N$ random weights is the best possible alternative.

5. Monte Carlo Study

We investigate the properties of the $CD$ test as well as its weighted alternatives in a small set of simulation experiments. We consider the common factor model

$$y_{it} = \beta x_{it} + \lambda f_t + \sqrt{\sigma_i^2} \varepsilon_{it},$$

$$x_{it} = A f_t + \varepsilon_{it}.$$  

We consider two alternative specifications for the factor loadings on the regressand $y_{it}$. We define $\lambda_i = t_r + \tilde{\lambda}_i$, where the $r$ elements of $\tilde{\lambda}_i$ are drawn (I) from $U(-0.75, 0.75)$ or (II) from a standardized $\chi^2(2)$ distribution that has zero mean and a variance of $1/6$. The latter case is designed to match the first $\chi^2$ standardization of $\lambda_i$. Since the moment of $\chi^2$ is very similar when model errors are drawn from a standardized $\chi^2$ distribution and corresponding tables can be found in Section 3.1 and 3.2.

Table 1 reports the first two moments of $CD$ when applied either to 2WFE residuals when the true model is one with two-way fixed effects or to CCE residuals in a model with multifactor error structure and two factors. The results for both cases are identical and show that $CD$ has a bias that diverges toward $-\infty$ as $T \rightarrow \infty$ and a variance that is considerably below 1. The bias term is reasonably close to the benchmark value of $-\sqrt{T}/2$ indicated by Proposition 1 but tends toward zero as heterogeneity among the individual-specific error variances is amplified.

Given the theoretical results of Sections 3.1 and 3.2 are confirmed by Table 1, we turn to the properties of $CD$ under its alternative hypothesis. For this purpose, we simulate a model with multifactor error structure and three factors, implying that either 2WFE or CCE estimation completely accounts for all sources of cross-section correlation in the simulated data. Heterogeneity among unit-specific error variances is kept constant by only considering $c_g = 1$. Instead we consider cases (I) and (II) for factor loadings as well as cases (a) and (b) for error variances. Again, results are presented only for normally distributed model errors. The moments of $CD$ are very similar when model errors are drawn from a standardized $\chi^2(2)$ distribution and corresponding tables can be found in Section S.2 of the supplementary appendix.

Tables 2 and 3 report the mean and variance of $CD$ in the presence of remaining cross-section correlation. Both are very similar to the numbers reported in Table 1 as long as factor loadings are symmetrically distributed, an observation which is in line with the theoretical results provided in Section S.3 of the supplementary appendix. Since symmetric loadings center the leading stochastic component of $CD$ around zero, the latter
across unit-specific error variances fixed at 5.2. Weighted CD Statistic dimension for this to happen is higher. We conjecture that the same sign reversal will eventually be substantial negative divergence of CD as

\[ T \text{ is dominated by a bias equivalent that diverges toward } -\infty \text{ as } T \to \infty. \] Qualitatively very different results can be observed when factor loadings are drawn from a skewed distribution, particularly when applying the CD test statistic to 2WFE residuals. As one can observe in columns two and four of Table 2, the strong negative divergence of CD as T increases is mitigated and it is reasonable to assume that it will be turned into positive divergence for \( N > 200 \). This pattern, which is amplified if error variances are a function of factor loadings, reflects the presence of a nonzero mean in the leading stochastic component of CD that diverges to \( +\infty \) at rate \( N\sqrt{T} \). In the case of 2WFE residuals, its magnitude is sufficiently large to counteract the negative bias equivalent already for \( N = 200 \). By contrast, columns two and four in Table 3 suggest that this term is considerably smaller when the CD test is applied to CCE residuals since the mean of CD continues to diverge toward \( -\infty \) for \( N = 200 \). We conjecture that the same sign reversal will eventually be obtained even in this case, but that the required cross-section dimension for this to happen is higher.

### 5.2. Weighted CD Statistic

Next, we proceed with investigating the properties of our weighted CD test statistic. As previously, we keep heterogeneity across unit-specific error variances fixed at \( c_\sigma = 1 \) and consider two different specifications each for factor loadings and error variances. We report size and power for our weighted test statistic CD\( W \) as well as the power-enhanced refinement CD\( W^{*} \). As a benchmark test, we include a CD test statistic with analytic bias correction which corrects CD with sample equivalents of the asymptotic bias terms in Theorems 1 and 2. Details on its implementation can be found in Section S.4 of the supplementary appendix. The original CD test is left out for the sake of saving space and in particular since rejection rates are 100% for most cases.

Table 2 reports the size and power of all three bias corrected test statistic when applied to 2WFE residuals. The size of CD\( W \) and CD\( W^{*} \) is very close to the nominal level of 5% as long as \( N \geq T \). Size distortions are given in cases where \( T \) is considerably larger than \( N \) and the effect of power enhancement on size is generally negligible. The analytically bias-corrected
Table 4. Rejection rates for weighted CD test statistic when applied to 2WFE residuals.

| Part A: Size | \( \lambda_i: \) Symmetric | \( \sigma_i^2: \) \( \perp \lambda_i \) | \( f(\lambda_i) \) | Part B: Power | \( \lambda_i: \) Skewed | \( \sigma_i^2: \) \( \perp \lambda_i \) | \( f(\lambda_i) \) |
|---|---|---|---|---|---|---|---|
| \( N \) | \( T \) | \( CD_W \) | \( CD_W^+ \) | \( CD_BC \) | \( CD_W \) | \( CD_W^+ \) | \( CD_BC \) | \( CD_W \) | \( CD_W^+ \) | \( CD_BC \) | \( CD_W \) | \( CD_W^+ \) | \( CD_BC \) |
| 25 | 25 | 5.3 | 5.5 | 2.7 | 6.2 | 6.4 | 2.8 | 11.5 | 16.3 | 17.8 | 12.5 | 14.9 | 27.4 |
| 50 | 50 | 6.6 | 6.9 | 4.3 | 5.8 | 6.1 | 4.1 | 18.9 | 52.9 | 35.1 | 24.2 | 56.9 | 50.7 |
| 25 | 50 | 7.8 | 8.5 | 3.7 | 7.0 | 9.1 | 4.6 | 31.0 | 90.6 | 50.7 | 38.5 | 95.8 | 71.1 |
| 200 | 25 | 9.8 | 10.3 | 5.6 | 9.0 | 10.0 | 6.1 | 44.3 | 99.4 | 33.5 | 52.8 | 99.9 | 83.1 |
| 50 | 25 | 4.6 | 4.7 | 2.8 | 4.7 | 4.7 | 3.7 | 12.3 | 77.2 | 42.3 | 14.2 | 145 | 65.1 |
| 50 | 50 | 5.5 | 5.9 | 4.7 | 5.4 | 5.8 | 4.8 | 19.8 | 79.1 | 65.6 | 22.8 | 666 | 89.1 |
| 50 | 25 | 6.1 | 6.2 | 3.4 | 5.2 | 5.4 | 3.6 | 32.2 | 99.8 | 85.0 | 37.1 | 99.9 | 97.2 |
| 200 | 25 | 4.5 | 4.8 | 4.2 | 4.6 | 4.6 | 4.4 | 44.0 | 4.6 | 4.6 | 4.9 | 4.9 | 4.9 |
| 200 | 50 | 4.1 | 5.3 | 4.4 | 5.1 | 5.3 | 4.6 |

NOTES: In Part A the model has two factors which are restricted as noted in Table 1. Part B corresponds to a model with factor error structure and 3 factors. For details on all other model parameters, see Table 2. \( CD_W \) is the weighted CD test statistic introduced in Theorem 3. \( CD_W^+ \) is its power-enhanced refinement. \( CD_BC \) is a CD test statistic with analytic bias correction.

CD test statistic \( CD_BC \) exhibits hardly any tendency to over-reject but is rather conservative, in particular when \( T \) is small.

Panel B in Table 4 report results on power. Here, we see that the power of \( CD_W \) is in general low and increases only in \( T \). This is improved upon considerably by power enhancement, leading the refined test statistic \( CD_W^+ \) to reliably reject when it should as long as the number of time periods is large enough. The performance of \( CD_W^+ \) is considerably above that of the benchmark statistic \( CD_BC \) when factor loadings are drawn from a symmetric distribution and is on par in most other cases. An exception is the case of skewed loadings with small \( T \) and large \( N \) where \( CD_BC \) has markedly higher rejection rates. It can also be noted that the performance of \( CD_BC \) largely depends on whether factor loadings are drawn from a symmetric distribution or not. If this is the case, analytic bias correction leads to rejection rates under the alternative hypothesis that are worst among all three tests considered.

When testing for remaining cross-section correlation in CCE residuals we observe similar results as in Table 4. Interestingly, it can be observed that the benchmark statistic \( CD_BC \) exhibits size distortions in samples with large \( T \) when loadings are drawn from a skewed distribution and if factor loadings and error variances are dependent. This results from the nature of our bias correction which assumes independence between \( \sigma_i^2 \) and \( \lambda_i \) to considerably improve the accuracy of the bias estimate.

The power properties of \( CD_W \) and \( CD_W^+ \) when either of these tests is applied to CCE residuals mirror those seen in the 2WFE case, even though rejection rates are generally somewhat lower. However, the power-enhanced statistic \( CD_W^+ \) now performs best in all cases without ever being inferior to \( CD_BC \).

### 6. Empirical Illustration: R&D Investments

In this section we illustrate the applicability of the standard and the weighted CD statistics using the R&D investments data of Eberhardt, Helmers, and Strauss (2013). Information on up to twelve manufacturing industries in 10 countries (Denmark, Fin-
Table 5. Rejection rates for weighted CD test statistic when applied to CCE residuals.

| \( \lambda_i \) | \( \varphi_i^2 \) | \( \perp \lambda_i \) | \( f(\lambda_i) \) | \( \perp \lambda_i \) | \( f(\lambda_i) \) |
|----------------|----------------|-----------------|-----------------|-----------------|-----------------|
|                |                | Symmetric       | CD\(_W\) CD\(_W^+\) CD\(_BC\) | CD\(_W\) CD\(_W^+\) CD\(_BC\) | CD\(_W\) CD\(_W^+\) CD\(_BC\) | CD\(_W\) CD\(_W^+\) CD\(_BC\) |
| N T            |               |                 |                 |                 |                 |                 |
| 25 25          | 5.9           | 5.9             | 4.2             | 5.3            | 5.6            | 4.6            |
| 25 50          | 6.4           | 6.9             | 4.5             | 5.7            | 6.2            | 5.3            |
| 25 100         | 8.3           | 9.0             | 6.1             | 7.9            | 8.4            | 6.6            |
| 25 200         | 11.2          | 12.2            | 6.8             | 10.3           | 11.3           | 8.2            |
| 50 25          | 5.9           | 6.1             | 4.1             | 6.2            | 6.4            | 5.0            |
| 50 50          | 6.4           | 6.7             | 4.4             | 5.8            | 5.9            | 4.0            |
| 50 100         | 6.1           | 6.3             | 5.4             | 5.9            | 6.0            | 5.9            |
| 50 200         | 7.3           | 7.2             | 5.0             | 6.4            | 6.3            | 6.5            |
| 100 25         | 5.2           | 5.2             | 4.5             | 5.5            | 5.5            | 4.2            |
| 100 50         | 5.0           | 5.1             | 3.8             | 4.8            | 4.9            | 4.9            |
| 100 100        | 5.6           | 5.8             | 4.9             | 6.0            | 5.9            | 5.7            |
| 100 200        | 5.4           | 5.5             | 4.9             | 5.9            | 5.9            | 6.2            |
| 200 25         | 5.3           | 5.3             | 3.8             | 5.1            | 5.1            | 3.9            |
| 200 50         | 4.3           | 4.4             | 4.4             | 4.9            | 4.9            | 4.9            |
| 200 100        | 4.5           | 4.6             | 4.5             | 5.6            | 5.6            | 4.4            |
| 200 200        | 4.4           | 4.7             | 5.1             | 5.8            | 5.8            | 5.1            |

Part A: Size

| \( \lambda_i \) | \( \varphi_i^2 \) | \( \perp \lambda_i \) | \( f(\lambda_i) \) | \( \perp \lambda_i \) | \( f(\lambda_i) \) |
|----------------|----------------|-----------------|-----------------|-----------------|-----------------|
|                |                | Skewed           | CD\(_W\) CD\(_W^+\) CD\(_BC\) | CD\(_W\) CD\(_W^+\) CD\(_BC\) | CD\(_W\) CD\(_W^+\) CD\(_BC\) |
| N T            |               |                 |                 |                 |                 |                 |
| 25 25          | 10.0          | 11.8            | 5.4             | 8.9            | 9.5            | 6.2            |
| 25 50          | 13.8          | 13.6            | 10.4            | 12.5           | 18.9           | 9.4            |
| 25 100         | 21.3          | 21.4            | 15.4            | 17.1           | 44.4           | 14.6           |
| 25 200         | 31.8          | 22.9            | 22.9            | 26.1           | 85.1           | 20.8           |
| 50 25          | 9.6           | 11.9            | 5.9             | 8.0            | 8.3            | 6.9            |
| 50 50          | 14.3          | 13.7            | 8.6             | 10.3           | 16.0           | 8.4            |
| 50 100         | 19.6          | 13.7            | 13.7            | 16.5           | 59.5           | 15.3           |
| 50 200         | 29.4          | 23.0            | 23.0            | 20.8           | 98.3           | 24.5           |
| 100 25         | 8.2           | 8.1             | 5.1             | 8.4            | 8.4            | 5.3            |
| 100 50         | 12.3          | 9.4             | 10.9            | 15.5           | 15.0           | 10.0           |
| 100 100        | 20.2          | 15.7            | 15.7            | 15.5           | 76.4           | 15.3           |
| 100 200        | 29.3          | 23.8            | 23.8            | 22.8           | 100            | 24.8           |
| 200 25         | 8.9           | 8.9             | 7.0             | 8.2            | 8.2            | 5.9            |
| 200 50         | 13.7          | 8.3             | 10.8            | 15.0           | 10.5           | 10.5           |
| 200 100        | 17.5          | 14.6            | 13.4            | 8.67           | 14.5           | 14.5           |
| 200 200        | 29.7          | 22.8            | 22.8            | 22.5           | 100            | 25.3           |

Part B: Power

NOTES: The model has as factor error structure with two (Part A) or three (Part B) factors. For details on all other model parameters, see Table 2. For an explanation of the tests, see Table 4.

land, Germany, Italy, Japan, The Netherlands, Portugal, Sweden, UK, and the United States) over a time period from 1980 to 2005 was used to construct the dataset. After some minor modification to the original dataset (see Section S.1 of the supplementary appendix), we are left with a panel dataset of \( N = 82 \) and \( T = 25 \).

In this application, serial correlation is important from an economic point of view. In Section S.1 of the supplementary appendix, we outline how the CD statistic can be modified to account for serial correlation, given a set of known weights \( w_i \).

Eberhardt, Helmers, and Strauss (2013) question whether R&D can be estimated in a standard Griliches-type “knowledge production function” framework ignoring the potential presence of knowledge spillovers between cross-section units as well as other cross-section dependencies. Among other things they document a

\[ \vdots \]

strong evidence for cross-section dependence and the presence of a common factor structure in the data, which we interpret as indicative for the presence of knowledge spillovers and additional unobserved cross-section dependencies. (Eberhardt, Helmers, and Strauss 2013, p.437).

Cross-section dependence was measured by means of a CD test. In this section, we will primarily revisit some of the results in Table 5 of Eberhardt, Helmers, and Strauss (2013) for pooled (static) production function estimates. Our goal is to investigate how the divergent properties of the CD test statistic might have influenced the choice between the First Difference (FD) estimator with yearly dummies and Pooled CCE (CCEP) estimators, as presented in columns 3 and 4, respectively, of the table mentioned above. Which of the two models is considered to be correctly specified has important consequences for the conclusions that can be drawn from the entire table. With regard to the coefficient of private R&D investments, Eberhardt, Helmers, and Strauss (2013) report that a significance test for the corresponding slope coefficient cannot reject the null hypothesis in the FD model while it can in the CCE model.

As the results for CD\(_W\) statistic allowing for serial-correlation correction, are similar to those without serial correlation, we
only focus on the latter option. First of all, as we can see from Table 6 the conclusions that we can draw from the original CD test are almost identical to those of Eberhardt, Helmers, and Strauss (2013), despite adjustments made in terms of the sample size. In particular, while the value of CD statistic based on CCE residuals imply rejection of the null hypothesis, no such conclusion is implied by FD (at least at a 5% significance level). However, as given that the original test statistic suffers from serial-correlation adjustments, thus serial-correlation cannot be seen from Table 6, the conclusion is unaltered after appropriate adjustments, as the conclusions are irrelevant for the FD estimator with time-effects only.

### 7. Conclusion

This article documents how the estimation of common time-specific parameters using panel data causes the CD test of Pesaran (2004, 2015a) to break down. Using popular additively and multiplicatively interacted specifications for individual- and time-specific components in the model errors, we show that the CD test statistic applied to residuals of correctly specified regression models is divergent under null hypothesis of cross-section independence. We can find an equivalent term under the alternative hypothesis which may balance out the leading diverging component of CD and can lead to low power in small samples. The results documented in this article are interpreted as a manifestation of the incidental parameter problem (IPP) since they ultimately follow from the estimation of $O(T)$ period-specific parameters. Our main theorems illustrate the pervasive nature of the IPP in this setup, given that the consequences of estimating time-specific parameters do not disappear as the sample size increases.

Our proposed weighted CD test statistic achieves our primary goal of re-establishing asymptotic standard normal inference and hence constitutes an alternative to popular bias correction methods which circumvent the problems these approaches have in the present context. Our results have far reaching implications for empirical panel data analysis, where CD test has been widely used as a model selection/diagnostic tool. An illustration of how our theoretical results translate into applications is given via simulations and real datasets.

Finally, in this article, we assumed that the parameters in the linear model are estimated using a least-square objective function. If one deviates from this setup, and instead uses an (over-identified) GMM criterion function to estimate parameters, the usual GMM $J$-statistic is readily available for the purpose of testing residual cross-section correlation. Examples in a fixed-$T$ framework are given by Sarafidis, Yamagata, and Robertson (2009), Ahn, Lee, and Schmidt (2013), and Juodis (2018) among others. In the large $N$, $T$ setup average $J$-statistic as a model specification tool was explicitly used, for example, by Everaert and Pozzi (2014) for the CCE-GMM estimator. Accordingly, these alternative procedures (if applicable) may be a suitable complement to the weighted CD test statistic proposed in this article.

### Acknowledgments

We would like to thank the Co-Editor (Jianqing Fan), an anonymous Associate Editor and two anonymous referees for numerous suggestions that greatly improved this article. We would also like to thank participants at AMES (Seoul, 2018 and Xiamen, 2019), IPDC (Seoul, 2018 and Vlijmius, 2019), IAAE (Montreal, 2018), Nordic Econometric Meeting (Stockholm, 2019), and seminar participants at Copenhagen Business School, Copenhagen University, Erasmus University Rotterdam, KU Leuven, Monash University, New Economic School, Queen Mary University London, University of Amsterdam, University of Bern, University of Birmingham, University of Groningen, University of Southern California, and Uppsala University. We express special thanks to Geert Dhaene, Jiti Gao, Cheng Hsiao, Frank Kleibergen, Braise Melly, Roger Moon, Andreas Pick, Geert Ridder, Vasilis Sarafidis, Peter Schmidt, Yongcheol Shin, Vanessa Smith, Bas Werker, and Joakim Westerlund for helpful comments. Simon Reese was a Postdoctoral Scholar Research Associate at USC Dornsife INET while working on this project.

### Funding

Financial support from the Netherlands Organization for Scientific Research (NWO) under research grant number 451–17–002 is gratefully acknowledged.

### Table 6. Cross-section dependence testing with R&D data.

| Serial Estimator | Correlation | $q_{0.1}(CD_{Wy})$ | $q_{0.5}(CD_{Wy})$ | $q_{0.9}(CD_{Wy})$ | $CD_{Wy}$ | $CD_{Wy}+$ |
|------------------|-------------|--------------------|--------------------|--------------------|----------|----------|
| FD               | No          | −1.90              | −1.41              | 0.04               | 0.78     | −0.23    | 1.49     |
|                  | Yes         | NA                 | −1.24              | 0.04               | 0.69     | −0.20    | 1.52     |
| CCEP             | No          | −2.95              | −1.72              | −0.54              | 1.72     | −1.30    | 7.38     |
|                  | Yes         | NA                 | −0.91              | −0.29              | 0.91     | −0.69    | 7.98     |

NOTES: All results are obtained for Weighted Covariance Bias-corrected CD statistic based on $G = 30$ Rademacher draws of $w_t$, $CD_{WC}$ denotes the analytical bias-corrected statistic. Serial correlation with “Yes” option stands for the adjusted test statistic robust to the serial correlation, as described in Section S.4.2 in the supplementary appendix.
