Estimating Heterogeneous Coefficients in Panel Data Models with Endogenous Regressors and Common Factors

By Timothy Neal *

This article extends the Common Correlated Effects (CCE) approach of Pesaran (2006) and Chudika and Pesaran (2015) by replacing OLS with 2SLS/GMM, and using lags of the variables to form the instrument set. By so doing the estimator is now robust to the presence of endogenous regressors, and Monte Carlo simulations show that in the case of dynamic panel data models it possesses better small sample properties regardless of whether the regressors are exogenous or endogenous. The estimator is then applied to the topic of a long run relationship between inequality and crime using U.S. state data, and finds evidence for a positive relationship between inequality and violent crime.

JEL: C23
Keywords: Panel Data, Common Factors, Endogeneity, Dynamic, GMM

The use of panel time series data (large N and T) has become increasingly popular over the last two decades, partly due to the availability of large international macroeconomic databases such as the Penn World Tables, the World Bank’s World Development Indicators, the IMF’s International Financial Statistics, and the World Top Incomes Database. Another contributing factor has been the development of the first generation of panel time series estimators which allowed for heterogeneity in the slope coefficients between panel units (a common feature of panels of this type). These include Mean Group OLS (Pesaran and Smith (1995)), Pooled Mean Group (Pesaran, Shin and Smith (1999)), Panel Fully Modified OLS (Pedroni (2000)), and Panel Dynamic OLS (Pedroni (2001)).

The literature has since shown that these estimators are inconsistent in the presence of cross-sectional dependence in the data, an issue also known as common factors or common shocks. Unobserved common factors in the panel can lead to correlation in the residuals across panel units, as well as correlation between the residuals and the regressors themselves. Left uncorrected it can cause severe bias in the coefficients, and accordingly it has garnered a great deal of attention in the literature. Recent work that has focused on correcting for these unobservable common factors in the estimation procedure can be categorised into those that assume slope homogeneity between panel units and those that do not. Techniques that rely on the assumption of slope homogeneity, including Bai (2009) and Moon

* University of New South Wales, timothy.neal@unsw.edu.au. Thank you to Luke Hartigan for constructive comments that improved the article.
and Widner (2015), can potentially lead to inconsistent estimation when the panel contains heterogeneous slopes.¹

A number of panel time series estimators have been proposed that are robust to both cross-sectional dependence as well as slope heterogeneity. Pesaran (2006) introduced Common Correlated Effects (‘CCE’) estimation, which approximates the projection space of the unobserved factors with cross section averages of the variables, and accounts for slope heterogeneity by using mean group instead of pooled regression. Other notable estimators in this class include the ‘CoVariance’ estimator (Li and Lina (2014)) and Augmented Mean Group (Bond and Eberhardt (2013)). Like CCE they are applied to a static panel model with strictly exogenous regressors. Chudika and Pesaran (2015) and Song (2013) extended the work of Pesaran (2006) and Bai (2009) (respectively) to allow for dynamic heterogeneous panel models that feature lags of the dependent variable as well as weakly exogenous regressors. Chudika and Pesaran (2015) noted that improving the small sample properties of the estimator is a challenge to be tackled in the future, as they found bias in the slope coefficient of the lagged dependent variable in small samples.

This article extends the CCE estimation approach of Pesaran (2006) and Chudika and Pesaran (2015) in order to overcome some of the few remaining issues in large panel data estimation. It replaces the use of Least Squares (‘OLS’) in the individual-specific regressions to Generalised Method of Moments (‘GMM’), and uses lagged observations of the variables to form the instrument set. A Monte Carlo simulation study demonstrates that exchanging OLS for GMM allows CCE to be robust to endogenous regressors in both static and dynamic panel data models, and it also significantly improves the small sample properties of the estimator in dynamic panel data models regardless of whether the regressors are strictly exogenous, weakly exogenous, or endogenous. These improvements are measured in terms of mean bias and root mean squared error (‘RMSE’). As in the original papers this instrumental variable approach is also robust to cross-sectional dependence and slope heterogeneity.

This estimator, along with the others discussed in this paper, is then applied to the topic of a long run relationship between inequality and crime using a panel time series dataset of U.S. states. The exercise finds evidence of a positive relationship between inequality and violent crime, but no evidence of a positive relationship with property crime. It illustrates how sensitive mean coefficient estimates can be between panel time series estimators in real datasets, and accordingly the importance of carefully considering the appropriate estimator for any panel data being analysed.

¹See Pesaran and Smith (1995) for a discussion on pooled panel estimation, as well as the simulation results in Kapetanios, Pesaran and Yamagata (2011).
The rest of the article is organised in the following way. Section 1 outlines the instrumental variable extension to CCE introduced in this article through a multifactor panel structure. Section 2 presents a Monte Carlo simulation study that seeks to thoroughly test this extension against a range of existent panel estimators. Section 3 applies these estimators to the topic of a potential long run relationship between inequality and crime in U.S. states. Concluding remarks are in Section 4, while the Appendix provides mathematical proofs and additional notation.

1. Panel Time Series Estimation with Common Factors

1.1. Multifactor Error Structure

Consider the following heterogeneous panel data model:

\[ y_{it} = \phi_i y_{it-1} + \beta'_i x_{it} + u_{it} \]

\[ u_{it} = c_{yi} + \gamma'_i f_t + \epsilon_{it} \]

\[ x_{it} = c_{xi} + \Gamma'_i f_t + v_{it} \]

where \( y_{it} \) is the observation associated with the \( i \)th panel unit at time \( t \) for \( i = 1, 2, ..., N \) and \( t = 1, 2, ..., T \). \( N \) and \( T \) correspond to a panel time series structure where both are medium to large in size. The error term \( u_{it} \) and the vector of regressors \( x_{it} \) (\( k \times 1 \) in dimension) are both determined by the individual specific fixed effects \( c_i = (c_{yi}, c_{xi}) \) and an \( m \times 1 \) vector of unobserved common factors \( f_t = (f_{1t}, f_{2t}, ..., f_{mt}) \). The factor loadings are given by \( C_i = (\gamma_i, \Gamma_i) \), where \( \gamma_i \) is a \( m \times 1 \) vector of factor loadings for the dependent variable and \( \Gamma_i \) is a \( m \times k \) matrix of factor loadings for the regressors. \( \epsilon_{it} \) and \( v_{it} \) are the idiosyncratic error terms.

Let \( \tau_{it} = (y_{it}, x_{it}) \) and write the above model compactly as:

\[ A_{0i} \tau_{it} = c_i + A_{1i} \tau_{it-1} + C_i f_t + e_{it} \]

where \( e_{it} = (\epsilon_{it}, v_{it}) \) is the error process, \( A_{0i} = \begin{pmatrix} 1 & -\beta_i \\ 0 & I_k \end{pmatrix} \) and \( A_{1i} = \begin{pmatrix} \phi_i & 0 \\ 0 & 0 \end{pmatrix} \). Assuming that \( A_{0i} \) is invertible this reduces to:
\begin{equation}
\tau_{it} = A_{0i}^{-1} c_i + A_{0i}^{-1} A_{1i} \tau_{i,t-1} + A_{0i}^{-1} C_i f_t + A_{0i}^{-1} e_{it}
\end{equation}

Note that if $A_{1i} = 0$ (i.e. $\phi_i = 0 \ \forall \ i$), then the model (1) - (3) reduces to a static heterogeneous multifactor panel data model akin to that considered in Pesaran (2006). Otherwise, it is a dynamic heterogeneous panel data model similar to that in Chudik and Pesaran (2015). The model abstracts away from several features including observed common effects (such as a trend term), lags of the regressors, and further lags of the dependent variable. These could be added at the cost of notational complexity, but will not meaningfully affect the results. Following Chudik and Pesaran (2015) we limit the behaviour of the above model with the following assumptions.

**Assumption 1: Common Effects** The $m \times 1$ vector of unobserved common factors $f_t = (f_{1t}, f_{2t}, ..., f_{mt})$ is assumed to be covariance stationary with absolute summable autocovariances and bounded fourth order moments. It is distributed independently of the elements in $e_{it} = (\epsilon_{it}, \nu_{it})$.

**Assumption 2: Factor Loadings and Rank** Factor loadings $\gamma_i$ and $\Gamma_i = (\Gamma_{1i}, \Gamma_{2i}, ..., \Gamma_{ki})$ are independent of the factor series $f_t$ and across $i$. They are generated from the following random loadings model:

\begin{equation}
\gamma_i = \gamma + \eta_{\gamma}, \ \eta_{\gamma} \overset{iid}{\sim} N(0, \Omega_{\gamma})
\end{equation}

and

\begin{equation}
\Gamma_{ji} = \Gamma + \eta_{\Gamma}, \ \eta_{\Gamma} \overset{iid}{\sim} N(0, \Omega_{\Gamma})
\end{equation}

for $i = 1, 2, ..., N$ and $j = 1, 2, ..., k$, where $\Omega_{\gamma}$ and $\Omega_{\Gamma}$ are symmetric non-negative definite matrices. Further assume that the $(k + 1) \times m$ factor loading matrix $C = (\gamma, \Gamma)$ has full column rank. In this context, full column rank indicates that the number of observed variables is at least as large as the number of unobserved common factors (i.e. $k + 1 \geq m$). This assumption is required for later results.\(^2\)

**Assumption 3: Coefficients** Define a coefficient vector $\pi_i = (\phi_i, \beta_i)$ which follows a random coefficient model:

\(^2\)If it is suspected that $k + 1 < m$, Pesaran, Smith and Yamagata (2013) suggest adding to the regression cross section averages of covariates that are correlated with the unobserved common factors, but not correlated with the dependent variable.
\( \pi_i = \pi + \eta_i, \eta_i \sim IID(0, \Omega_\pi) \)

where \( \Omega_\pi \) is a symmetric non-negative definite matrix, and the random error \( \eta_i \) is distributed independently of the error terms of the dependent variable \( \epsilon_{it} \) and the regressor \( v_{it} \), the factor loadings \( \gamma_i \) and \( \Gamma_i \), and finally the factor series \( f_t \). Furthermore, all elements of \( \phi_i \) are assumed to lie strictly inside the unit circle.

**Assumption 4: Error term** Let the idiosyncratic error terms \( \epsilon_{it} \) and \( v_{it} \) be independent from one another at all \( i \) and \( t \). Furthermore, let the vector of errors across \( i \), \( \epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{Nt}) \), possess weak cross-sectional dependence with:

\[
\epsilon_t = R\varsigma_t
\]

and also:

\[
v_{it} = \sum_{l=0}^{\infty} S_{il}\varsigma_{v_{i-t-l}}
\]

where \( \varsigma_t = (\varsigma_{\epsilon_{1t}}, \varsigma_{\epsilon_{2t}}, ..., \varsigma_{\epsilon_{Nt}}) \) and the \( K \times 1 \) vector of variables \( \varsigma_{v_{it}} \) are IID with mean 0, unit variances, and finite fourth-order moments. The \( N \times N \) matrix \( R \) has bounded row and column norms and the diagonal elements of \( RR' \) are bounded away from zero. Furthermore:

\[
\|Var(v_{it})\| = \left\| \sum_{l=0}^{\infty} S_{il}S_{il}' \right\| \leq K < \infty,
\]

for all \( i = 1, 2, ..., N \).

**Assumption 5: Exogeneity of the Regressors** Assume that the regressors are strictly exogenous, with \( E(x_{is}\epsilon_{it}) = 0 \) for all \( s \) and \( t \).

### 1.2. Common Factor Representation

The chief contribution of Pesaran (2006) and Chudika and Pesaran (2015) is to demonstrate that cross section averages of observed variables and their lags are able to adequately approximate the projection space of the unobserved common factors \( f_t \) (under several conditions) for static (i.e. \( \phi_i = 0 \ \forall \ i \)) and dynamic panel
Following Chudika and Pesaran (2015), first define a series of weights for the cross section averages $w = (w_1, w_2, ..., w_N)$ of $N \times 1$ dimension that satisfy the conditions that the spectral norm is $\|w\| = O(N^{-\frac{1}{2}})$, $\frac{w_i}{\|w\|} = O(N^{-\frac{1}{2}})$, and finally $\sum_{i=1}^N w_i = 1$. Next, assuming that the support of the eigenvalues of $A_{0i}^{-1}A_{1i}$ lies strictly within the unit circle, express (5) as the following covariance stationary process:

$$\tau_{it} = \sum_{l=0}^{\infty} (A_{0i}^{-1}A_{1i})^l (A_{0i}^{-1}c_i + A_{0i}^{-1}C_i f_{t-l} + A_{0i}^{-1}e_{it-l})$$

for each $i$. Define $\tilde{\tau}_{wt} = \tau_{wt} - A_{0i}^{-1}c_w$ as the de-trended cross section averages, where $\tilde{\tau}_{wt} = (\tilde{y}_{wt}, \tilde{x}_{wt}) = \sum_{i=1}^N w_i \tau_{wt}$. Take the weighted cross-sectional average of (12) and assuming Assumptions 1-3 hold yields:

$$\tilde{\tau}_{wt} = \Lambda(L)Cf_t + O_p(N^{-\frac{1}{2}})$$

where $C = E(C_i) = (\gamma, \Gamma)$, $\Lambda(L) = \sum_{l=0}^{\infty} E((A_{0i}^{-1}A_{1i})^lA_{0i}^{-1})L^l$, and $L$ is the lag operator. The approximation error $O_p(N^{-\frac{1}{2}})$ arises from the idiosyncratic error term $e_{it}$, relying on Assumptions 2-3 and the fact that it possesses weak cross-sectional dependence, such that:

$$\sum_{i=1}^N \sum_{l=0}^{\infty} w_i (A_{0i}^{-1}A_{1i})^lA_{0i}^{-1}e_{it-l} = O_p(N^{-\frac{1}{2}})$$

Rearranging (13), we obtain an expression of the unobserved common factors with the factor loading matrix:

$$Cf_t = \Lambda^{-1}(L)\tilde{\tau}_{wt} + O_p(N^{-\frac{1}{2}})$$

Since, according to Assumption 5, the regressor $x_{it}$ is strictly exogenous, the elements of $\Lambda^{-1}(L)$ can be expressed directly. Start by multiplying (1) by $(1 - \phi_tL)^{-1}$ which gives:

$$y_{it} = \sum_{l=0}^{\infty} \phi_t^l c_{yi} + \sum_{l=0}^{\infty} \phi_t^l \beta_t^l x_{it} + \sum_{l=0}^{\infty} \phi_t^l \gamma_t^l f_t + \sum_{l=0}^{\infty} \phi_t^l e_{it}$$
Taking the cross section average of $y_{it}$ and the vector of regressors $x_{it}$ gives:

\[(17) \quad \bar{y}_{wt} = \bar{c}_{yw} + a(L)\beta'\bar{x}_{wt} + a(L)\gamma'f_t + O_p(N^{-\frac{1}{2}})\]

and

\[(18) \quad \bar{x}_{wt} = \bar{c}_{xw} + \Gamma'f_t + O_p(N^{-\frac{1}{2}})\]

where $a(L) = \sum_{l=0}^{\infty} E(\phi_l)L^l$, $\bar{c}_{yw} = \sum_{i=1}^{N} w_i c_{yi} (1 - \phi_i)^{-1}$, $\bar{c}_{xw} = \sum_{i=1}^{N} w_i c_{xi}$, and let $b(L) = a^{-1}(L)$. After rearranging, this becomes:

\[(19) \quad \gamma'f_t = b(L)\bar{y}_{wt} - b(1)\bar{c}_{yw} - \beta'\bar{x}_{wt} + O_p(N^{-\frac{1}{2}})\]

and

\[(20) \quad \Gamma'f_t = \bar{x}_{wt} - \bar{c}_{xw} - O_p(N^{-\frac{1}{2}})\]

Stacking (19) and (20) together yields (15), and finally:

\[(21) \quad f_t = (C'C)^{-1}C'\Lambda^{-1}(L)\bar{\tau}_{wt} + O_p(N^{-\frac{1}{2}})\]

where $\Lambda^{-1}(L) = \begin{pmatrix} b(L) & -\beta \\ 0 & I_k \end{pmatrix}$. This is the expression of the unobserved common factors that will be used to obtain a robust estimator in subsection 1.3. Note that in a static model, where $\phi_i = 0 \forall i$, (16) becomes:

\[(22) \quad y_{it} = c_{yi} + \beta_i x_{it} + \gamma_i f_t + \epsilon_{it}\]

and it can be seen that the process described in (17) - (20) produces (21) where $\Lambda^{-1}(L) = \begin{pmatrix} 1 & -\beta \\ 0 & I_k \end{pmatrix}$.

1.3. Estimation with Exogenous Regressors

Substituting the expression for the unobserved common factors in (21) into (1) yields:
\( y_{it} = c^*_{y_i} + \phi_i y_{it-1} + \beta'_i x_{it} + \sum_{l=0}^{\infty} \delta_{il} \bar{\tau}_{wt} + O_p(N^{-\frac{1}{2}}) + \epsilon_{it} \) \hspace{1cm} (23)

where \( \sum_{l=0}^{\infty} \delta_{il} = (C'C)^{-1}C'\Lambda^{-1}(L) \) and \( c^*_{y_i} = c_{y_i} - \delta'_i(1)\bar{\epsilon}_{\tau w} \). Since the support of \( \phi_i \) lies strictly within the unit circle, the coefficients of \( \sum_{l=0}^{\infty} \delta_{il} \) will decay at an exponential rate. Accordingly, the unobserved common factors are able to be approximated using a cross section average of \( \tau_{it} \) and its lags. Let \( p_T \) be the number of lags of the cross section averages to be used, and the regression equation becomes:

\( y_{it} = c^*_{y_i} + \phi y_{it-1} + \beta'_i x_{it} + \sum_{l=0}^{p_T} \delta_{il} \bar{\tau}_{wt} + \sum_{l=p_T+1}^{\infty} \delta_{il} \bar{\tau}_{wt} + O_p(N^{-\frac{1}{2}}) + \epsilon_{it} \) \hspace{1cm} (24)

The error is comprised of three components. \( O_p(N^{-\frac{1}{2}}) \) reflects the approximate nature of using cross section averages to proxy unobserved common factors. \( \sum_{l=p_T+1}^{\infty} \delta_{il} \bar{\tau}_{wt} \) comes from the truncation of the infinite polynomial distributed lag function \( \delta_i(L) \). The larger the lag length selected for the cross section averages, defined as \( p_T \), the smaller this component of the error term will become (with the trade-off that adding additional lags removes observations from the sample). Lastly, \( \epsilon_{it} \) is the idiosyncratic error term described above. Should the model be static (i.e. \( \phi_i = 0 \ \forall \ i \)) (24) reduces to:

\( y_{it} = c^*_{y_i} + \beta'_i x_{it} + \delta_i \bar{\tau}_{wt} + O_p(N^{-\frac{1}{2}}) + \epsilon_{it} \) \hspace{1cm} (25)

since the coefficients for any lag length is equal to zero. Therefore, the projection space of the unobserved common factors can be approximated through a cross section average of \( \tau_{it} \) without any lags. To express the estimator of \( \beta_i \) first define the following matrices:

\[ \Xi_i = \begin{pmatrix} y_{i,p_T} & x_{i,p_T+1} \\ y_{i,p_T+1} & x_{i,p_T+2} \\ \vdots & \vdots \\ y_{i,T-1} & x_{i,T} \end{pmatrix}, \quad \bar{Q}_w = \begin{pmatrix} 1 & \bar{\tau}_{w,p_T+1} & \bar{\tau}_{w,p_T} & \cdots & \bar{\tau}_{w,1} \\ 1 & \bar{\tau}_{w,p_T+2} & \bar{\tau}_{w,p_T+1} & \cdots & \bar{\tau}_{w,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \bar{\tau}_{w,T} & \bar{\tau}_{w,T+1} & \cdots & \bar{\tau}_{w,T-p_T} \end{pmatrix}, \]

and the projection matrix \( \bar{M}_q = I_{T-p_T} - \bar{Q}_w'\bar{Q}_w^{-1}\bar{Q}_w \). Now Assumption 6 is defined to ensure that these matrices are well defined.\(^3\)

\(^3\)Analogous to Assumption 7 in Chudika and Pesaran (2015).
Assumption 6: Asymptotic Behaviour of Matrices There exists a $N_0$ and $T_0$ such that $\forall N \geq N_0$ and $T \geq T_0$, the matrix $(\Xi_i' \bar{M}_q \Xi_i / T)^{-1}$ exists for all $i$. Second, the matrix $\Sigma_{i\xi}$ defined in (A2) is invertible and $\|\Sigma_{i\xi}^{-1}\| < K < \infty$ for all $i$. Lastly, the elements of the matrix $\Xi_i M_h$, where $M_h$ is defined in (A1), all contain uniformly bounded fourth moments (for all $t$, $i$, and for each regressor).

The coefficients contained in $\hat{\pi}_i$ can be expressed as:

\[(27)\]
\[
\hat{\pi}_i = (\Xi_i' \bar{M}_q \Xi_i)^{-1} \Xi_i' \bar{M}_q y_i
\]

The mean group estimate of (27) can be found by $\hat{\pi} = \frac{1}{N} \sum_{i=1}^{N} \hat{\pi}_i$. This is the Dynamic Common Correlated Effects (‘DCCE’) estimator, introduced in Chudika and Pesaran (2015), for dynamic panel models. The Common Correlated Effects (‘CCE’) estimator, introduced in Pesaran (2006) for static panel models, is equivalent to (27) except that:

\[(28)\]
\[
\bar{Q}_w = \begin{pmatrix}
1 & \bar{\tau}_{w,1} \\
1 & \bar{\tau}_{w,2} \\
\vdots & \vdots \\
1 & \bar{\tau}_{w,T}
\end{pmatrix}
\]

Theorems that detail the conditions required for the asymptotic performance of the unit-specific and mean group coefficients will now be outlined.

Theorem 1: Consistency of $\hat{\pi}_i$ and $\hat{\pi}$

For the panel model outlined in (1) - (3) in cases where Assumptions 1-6 hold, as $(N, T, pT) \to \infty$ such that $\frac{pT}{T} \to \chi$, $0 < \chi < \infty$ it is true that:

\[(29)\]
\[
\hat{\pi}_i - \pi_i \overset{p}{\to} 0
\]

and also

\[(30)\]
\[
\hat{\pi} - \pi \overset{p}{\to} 0
\]

The proofs of both equations can be found in Chudika and Pesaran (2015).4 Briefly, they showed that:

\[4\text{See Proof of Theorem 1 and 2 in the Appendix of the paper.}\]
\[
\hat{\pi}_i - \pi_i = \left( \frac{\Xi'_i \bar{M}_q \Xi_i}{T} \right)^{-1} \frac{\Xi'_i \bar{M}_q}{T} (\epsilon_i + \eta_i + \Upsilon_i)
\]

where \( \left( \frac{\Xi'_i \bar{M}_q \Xi_i}{T} \right)^{-1} = O_p(1) \), \( \epsilon_i = (\epsilon_{i,pT+1}, \epsilon_{i,pT+2}, ..., \epsilon_{iT}) \), \( \eta_i = \sum_{l=pT+1}^{\infty} \delta_{il} \bar{\tau}_{wl} \), and \( \Upsilon_i \) is \( O_p(N^{-\frac{3}{2}}) \). They further show that this reduces to

\[
\hat{\pi}_i - \pi_i = \left( \frac{\Xi'_i \bar{M}_q \Xi_i}{T} \right)^{-1} \frac{\Xi'_i \bar{M}_q \epsilon_i}{T}
\]

With the assumption of exogeneity of the regressors (Assumption 5), it can be seen that:

\[
\frac{\Xi'_i \bar{M}_q \epsilon_i}{T} \xrightarrow{p} 0
\]

and therefore there is consistency in the case of exogenous regressors with the CCE/DCCE estimators. The asymptotic variance of \( \hat{\pi} \) will now be outlined in a second theorem.

**Theorem 2: Asymptotic Variance of \( \hat{\pi} \)**

For the panel model outlined in (1) - (3) in cases where Assumptions 1-6 hold, as \((N, T, pT) \xrightarrow{d} \infty\) such that \( \frac{N}{T} \to \chi_1 \) and \( \frac{pT}{T} \to \chi_2 \), where \( \chi_1 > 0 \) and \( 0 < \chi_2 < \infty \) it is true that:

\[
\sqrt{N} (\hat{\pi} - \pi) \xrightarrow{d} N(0, \Omega_{\pi})
\]

where \( \Omega_{\pi} = Var(\pi_i) = Var(\eta_\pi) \) as defined in (8).\(^5\) Note that unlike Theorem 1, the asymptotic variance of DCCE requires that \( N/T \) converge to a fixed number, and is therefore not appropriate for panels where \( T \) is small relative to \( N \). This is true for the case of DCCE, but CCE does not require this condition (see Pesaran (2006)).

1.4. **Estimation with Endogenous Regressors**

Now suppose that Assumption 5 does not hold, and \( E(x_{it} \epsilon_{it}) \neq 0 \). This is the case of endogenous regressors, where the regressor is correlated with the error

\(^5\)For a formal proof please see the Proof of Theorem 3 in Chudika and Pesaran (2015).
term beyond the presence of unobserved common factors (which are controlled by the CCE/DCCE augmentations). In this case, (33) no longer holds and the CCE/DCCE estimators will be inconsistent in static/dynamic panel models.

Define a set of $J$ instruments:

$$
Z_{iw} = \begin{pmatrix}
    z_{1i,pT+1} & \cdots & z_{Ji,pT+1} \\
    z_{1i,pT+2} & \cdots & z_{Ji,pT+2} \\
    \vdots & \ddots & \vdots \\
    z_{1i,T} & \cdots & z_{Ji,T}
\end{pmatrix},
$$

and make the following assumption regarding their characteristics.

**Assumption 7: Instrument Set** The set of instruments $Z_{iw}$ are exogenous ($E(Z_{it}\epsilon_{it}) = 0$), linearly independent ($\text{rank}(Z_{iw}Z_{iw}) = J$), are sufficiently correlated with the regressors to contain full rank ($\text{rank}(Z_{iw}'\Xi_i) = K + 1$), and finally satisfy the order condition $J \geq K + 1$ for the complete identification of coefficients.

The instrument set will be populated with any exogenous regressors, the cross section averages, and lags of the endogenous regressors and/or dependent variable (given that the panel is long possessing a sufficient number of exogenous lags will be achievable). Define the exogenous regressor set $\hat{\Xi}_i$ as:

$$
\hat{\Xi}_i = Z_i\zeta^{-1}Z_i\Xi_i
$$

where $\zeta$ is a positive semi-definite weight matrix. Next, estimate CCE/DCCE with a GMM estimator by substituting (36) into (27) to obtain:

$$
\hat{\pi}_{GMM}^i = (\Xi_i'Z_i\zeta^{-1}Z_i'M_qZ_i(\zeta^{-1}Z_i'\Xi_i)\Xi_i'Z_i\zeta^{-1}Z_i'y_i
$$

or more compactly as:

$$
\hat{\pi}_{GMM}^i = (\hat{\Xi}_i'M_q\hat{\Xi}_i)^{-1}\hat{\Xi}_i'M_qy_i
$$

One choice of weight matrix is $\zeta^{-1} = (Z_i'Z_i)^{-1}$, which is equivalent to estimating the regression equations through 2SLS:

$$
\hat{\pi}_{2SLS}^i = (\Xi_i'Z_i(Z_i'Z_i)^{-1}Z_i'M_qZ_i(Z_i'Z_i)^{-1}Z_i'\Xi_i)\Xi_i'Z_i(Z_i'Z_i)^{-1}Z_i'y_i
$$
Alternatively, in order to obtain a more efficient weight matrix in cases of heteroskedasticity and/or autocorrelation, first estimate through 2SLS in order to obtain the residuals $\tilde{u}_{it}$. Then, estimate the covariance of the second moments $\text{Var}(Z_i'\tilde{u}_{it})$. Use the inverse of this estimate as the weight matrix to obtain a consistent GMM estimator with an efficient HAC weight matrix.

The theorems that establish the asymptotic characteristics of estimating CCE/DCCE with 2SLS will now be outlined.

**Theorem 3: Consistency of $\hat{\pi}^{2SLS}_i$ and $\hat{\pi}^{2SLS}$**

For the panel model outlined in (1) - (3) in cases where Assumptions 1-4 and 6-7 hold, as $(N,T,p_T) \xrightarrow{j} \infty$ such that $\frac{p_T}{T} \rightarrow \chi$, $0 < \chi < \infty$ it is true that:

(40) \[ \hat{\pi}^{2SLS}_i - \pi \xrightarrow{p} 0 \]

and also

(41) \[ \hat{\pi}^{2SLS} - \pi \xrightarrow{p} 0 \]

**Theorem 4: Asymptotic Variance of $\hat{\pi}^{2SLS}$**

For the panel model outlined in (1) - (3) in cases where Assumptions 1-4 and 6-7 hold, as $(N,T,p_T) \xrightarrow{j} \infty$ such that $N/T \rightarrow \chi_1$ and $\frac{p_T}{T} \rightarrow \chi_2$, where $\chi_1 > 0$ and $0 < \chi_2 < \infty$ it is true that:

(42) \[ \sqrt{N}(\hat{\pi}^{2SLS} - \pi) \xrightarrow{d} N(0,\Omega_\pi) \]

where $\Omega_\pi = \text{Var}(\pi_i) = \text{Var}(\eta_\pi)$ as defined in (8). The proof of both Theorem 3 and 4 can be found in the Appendix of this article.

2. Monte Carlo Simulation Results

This section tests the finite sample performance of the instrumental variable extension to the CCE/DCCE estimator introduced in section 1 against a range of other panel time series estimators using Monte Carlo simulation analysis. A number of scenarios are constructed from a general data generating process that features a panel time series structure (with $N$ and $T$ between 30 and 100 to represent the typical size of panel time series datasets) and unobserved common factors. The chief differentiation between scenarios is the inclusion of a lagged dependent variable and the degree of endogeneity in the single regressor $x_{it}$. 
The performance of each estimator under consideration will be measured using the mean bias and the root mean squared error, which is defined as:

\[
RMSE = \sqrt{\frac{1}{S} \sum_{s=1}^{S} (\hat{\beta}_s - \beta)^2}
\]

where \( \beta \) is the true mean group coefficient, \( \hat{\beta}_s \) is the estimated mean group coefficient for simulation \( s \), and \( S \) is the total number of simulation repetitions which was set to 2,000 for this study.

2.1. Data Generating Process

Analogous to (1) the dependent variable is defined by:

\[
y_{it} = \phi_i y_{i,t-1} + \beta_i x_{it} + u_{it}
\]

where \( i = 1, 2, ..., N \) and \( t = -99, ..., 0, 1, ..., T \). In all scenarios, we generate heterogeneous coefficients \( \beta_i = 1 + \eta \beta \) with \( \eta \beta \sim U[-0.25, 0.25] \). Accordingly, an unbiased estimate of \( \beta \), being the group mean of \( \beta_i \), will be equal to 1. \( \phi_i \) will either be set to zero in the static models, or \( \phi_i = 0.5 + \eta \phi \) with \( \eta \phi \sim U[-0.25, 0.25] \) in the dynamic models. The error term takes on a multifactor form with fixed effects, analogous to (3):

\[
u_{it} = c_{yi} + \gamma'f_t + \epsilon_{it}
\]

where \( \epsilon_{it} \sim N(0, \sigma_{\epsilon}) \), \( \sigma_{\epsilon} = 0.0025 \), and \( c_{yi} \sim U[0, 1] \). The regressor takes the following form:

\[
x_{it} = c_{xi} + \Gamma'f_t + \delta \epsilon_{i,t-1} + \kappa \epsilon_{it} + v_{it}
\]

\[
v_{it} = \rho_x v_{i,t-1} + \epsilon_{it}
\]

Once the serially correlated error term \( v_{it} \) is brought into the main equation, the regressor becomes:

\[
x_{it} = (1 - \rho_x) c_{xi} + \rho_x x_{i,t-1} + \Gamma'f_t - \rho_x \Gamma'f_{t-1} + \\
\delta (\epsilon_{i,t-1} - \rho_x \epsilon_{i,t-2}) + \kappa (\epsilon_{it} - \rho_x \epsilon_{i,t-1}) + \epsilon_{it}
\]
where $e_{it} \sim N(0, \sigma_e)$ and $\sigma_e \sim U[0.001, 0.003]$. The series (48) begins with $x_{i,-99} = c_{xi}$ where $c_{xi} \sim N(0.5, 0.5)$ and continues along the relevant process with $t = -99, ..., 0, 1, ..., T$. The first 100 observations are discarded prior to estimation to ensure that the results are not sensitive to the selection of the initial value.

The vector of $M$ common factors $f_t = (f_{1t}, ..., f_{Mt})$ is generated under the following process:

\[(49) \quad f_{mt} = \mu_m + \rho_f f_{m, t-1} + v_{mt}, v_{mt} \sim N(0, \sigma_f)\]

where $\sigma_f = 0.0025$, and $\mu_m = (0.015, 0.012)$. As before, the process is applied to $t = -99, ..., 0, 1, ..., T$, with $f_{m,-99} = 0$ and the first 100 observations are discarded prior to estimation. The number of unobserved common factors, $M$, has been set to two for this study.\(^6\) The factor loading vectors in the error and regressor terms are generated independently with $\gamma_{mi} = \gamma + \eta_\gamma$ and $\Gamma_{mi} = \Gamma + \eta_\Gamma$ for $M = 2$ where $\gamma = \Gamma = 0.5$, $\eta_\Gamma \sim U[-0.25, 0.25]$, and $\eta_\gamma \sim U[-0.25, 0.25]$.

This is a flexible DGP that closely mirrors the original estimation problem in section 1. A number of parameters have yet to be defined that will determine the order of integration of the variables ($\rho_x$, and $\rho_f$), and also the severity of endogeneity in the regressor ($\delta$ and $\kappa$). For the set of results in this section, the regressor is nonstationary with $\rho_x = 1$ and the factors are stationary with $\rho_f = 0.5$. The remaining parameters will vary depending on the specific scenario.

Six scenarios are considered, that principally vary the degree of endogeneity of the single regressor and whether the dependent variable is autoregressive. The scenario design is presented in Table 1. The first three scenarios are static models that do not feature a lagged dependent variable, with $\phi_i = 0$ for all $i$. Scenarios 4 to 6 are dynamic panel models, where $\phi_i = 0.5 + \eta_\phi$. Scenarios 1 and 4 contain strictly exogenous regressors with $\delta = \kappa = 0$. Scenarios 2 and 5 feature weakly exogenous regressors with $\delta = 0.5$ while $\kappa = 0$. Scenarios 3 and 6 contain endogenous regressors where $\delta = \kappa = 0.5$.

\[2.2. \text{Summary of Estimators}\]

All estimators considered in the Monte Carlo study are mean group estimators, meaning a regression is run for each panel unit and the individual coefficients are averaged to obtain a mean group estimate i.e. $\beta = N^{-1} \sum_{i=1}^N \beta_i$. The mean group estimators tested in this simulation study are briefly outlined below.

\(^6\)Adding additional unobserved common factors will break the assumption of full column rank in the factor loading matrix, and necessitate additional regressors or the use of cross section averages of covariates (that are correlated with the common factors but not $y_{it}$ into the DGP as in Pesaran, Smith and Yamagata (2013)). It would not meaningfully affect the relative results between estimators.
### Table 1—Scenario Design

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\phi_i$</th>
<th>$\delta$</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>0.5 + $\eta_\phi$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0.5 + $\eta_\phi$</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0.5 + $\eta_\phi$</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Note: In all scenarios the number of replications is set to 2,000.

## Panel Dynamic OLS

Panel Dynamic OLS (‘PDOLS’), introduced in Pedroni (2001) and based on the time series estimator for cointegrated systems seen in Stock and Watson (1993), runs the following regression for each panel unit:

\[
y_{it} = \beta_i x_{it} + \sum_{s=-l}^{l} \delta_{is} \Delta x_{is} + e_{it}
\]

where the regressor is augmented with $l$ lags and leads of its differences. The data can be time demeaned with $\bar{x}_{it} = x_{it} - \sum_{i=1}^{N-1} x_{it}$ which will replace $x_{it}$ in the regression and is equivalent to adding time dummies to the regression (in order to account for unobserved common factors with homogeneous factor loadings). Both the standard (‘PDOLS’) and time-demeaned version (‘PDOLS-DUM’) of this estimator are considered in the static models, with an a priori rule of $l = 2$ used in all results.

## Common Correlated Effects

The CCE estimation method (‘CCE’), first introduced in Pesaran (2006), seeks to approximate the projection space of the unobserved factors through the cross-sectional averages (i.e. the average across panel units over one period of time) of the dependent and explanatory variables. As such, the individual regressions are:

\[
y_{it} = \beta_i \bar{x}_{it} + \delta_{xi} \bar{x}_t + \delta_{yi} \bar{y}_t + e_{it}
\]

where $\bar{x}_t = N^{-1} \sum_{i=1}^{N} x_t$ is the cross-sectional average of the regressor(s), and $\bar{y}_t = N^{-1} \sum_{i=1}^{N} y_t$ is the cross-sectional average of the dependent variable. Further details can be seen in section 1.3. CCE will be tested in both the static and dynamic scenarios.
Dynamic Common Correlated Effects

Everaert and Groote (2013) showed that the standard CCE estimation method is unsuitable in models with a lagged dependent variable, due to a number of biases. In response, Chudika and Pesaran (2015) extend the CCE approach to allow for models with lagged dependent variables and weakly exogenous regressors. Called Dynamic Common Correlated Effects (‘DCCE’), it uses lags of the cross section averages in the regression, as follows:

\[ y_{it} = \phi_i y_{i,t-1} + \beta_i x_{it} + \sum_{p=0}^{p_T} \delta_{xip} \bar{x}_{t-p} + \sum_{p=0}^{p_T} \delta_{yip} \bar{y}_{t-p} + e_{it} \]

where the number of lags included in the cross section averages, \( p_T \), is variable. For further details see section 1.3. DCCE will be tested in the dynamic scenarios of this study. The \textit{a priori} rule \( p_T = T^{1/3} \) will be used for all results.

Common Correlated Effects with 2SLS

The estimator introduced in section 1.4 expands on CCE and DCCE by using an instrument set to further account for endogenous regressors in static and dynamic models. In static models:

\[ y_{it} = \beta_i \hat{x}_{it} + \delta_{xi} \bar{x}_t + \delta_{yi} \bar{y}_t + e_{it} \]

and dynamic models:

\[ y_{it} = \phi_i \hat{y}_{i,t-1} + \beta_i \hat{x}_{it} + \sum_{p=0}^{p_T} \delta_{xip} \bar{x}_{t-p} + \sum_{p=0}^{p_T} \delta_{yip} \bar{y}_{t-p} + e_{it} \]

where \( \hat{x}_{it} \) and \( \hat{y}_{it-1} \) is defined in (36). Considering the structure of the data generating process, the study will only present results for CCE estimated with 2SLS (‘CCE-2SLS’) in the static scenarios, and DCCE estimated with 2SLS (‘DCCE-2SLS’) in the dynamic scenarios. The instrument set used for these estimators are listed below each results table.

Augmented Mean Group

Augmented Mean Group, introduced in Bond and Eberhardt (2013), uses a two-step regression that includes a common dynamic effect to the individual panel unit regressions in the second stage. The dynamic effect is estimated through time dummies included in the first stage first-difference pooled regression. The set up is as follows:
Stage 1:

\[(55) \Delta y_{it} = \beta \Delta x_{it} + \sum_{t=2}^{T} c_t \Delta D_t + e_{it}\]

In this pooled first-difference regression, \(D_t\) represents time dummies (starting from the second period as they are differenced). The coefficients to the time dummies, \(c_t\), are turned into a variable shared across panel units \(\hat{\mu}_t\), as a coefficient estimate will exist for each time period in the panel.

Stage 2:

\[(56) y_{it} = a_i + \beta_i x_{it} + d_i \hat{\mu}_t + e_{it}\]

The time dummy coefficient variable included in Stage 2 approximates the unobserved common factors that are potentially driving the variables in each panel unit.\(^7\)

2.3. Results

The results for scenario 1, the baseline DGP, are presented in Table 2. It is a static model with unobserved common factors and strictly exogenous regressors. Of all the estimators considered here PDOLS performs the worst, due to the bias exhibited from failing to account for cross-sectional dependence. Time-demeaning the data prior to estimation successfully removes the bias from the mean group parameter in this scenario. However, that is unlikely to be true if the distribution of factor loadings were more complex. AMG is the most efficient estimator in this scenario, very closely followed by CCE. Predictably, with strictly exogenous regressors and a static specification estimating CCE through 2SLS is less efficient.

The results for scenario 2 are presented in Table 3, which is identical to Scenario 1 except it features a weakly exogenous regressor. Bias is introduced in PDOLS-DUM, indicating that is not capable of unbiasedly estimating the coefficients with unobserved common factors and regressors that are not strictly exogenous. Moderate bias is also present in CCE and AMG, but it declines rapidly as \(T\) expands. CCE-2SLS successfully removes the bias at any value of \(T\) considered here. In terms of RMSE, however, CCE-2SLS perform roughly on par with CCE. The reduction in bias in the former is offset by being less efficient than the latter. Accordingly, in a static estimation problem with weakly exogenous regressors there appears to be a bias-efficiency trade-off between estimating CCE with OLS or with 2SLS.

\(^7\)Note that it only accounts for the unobserved heterogeneity from the common factors as the regressors were also included in the first stage regression.
Table 2—Simulation Results - Scenario 1

<table>
<thead>
<tr>
<th>(N = 50, T)</th>
<th>RMSE (x100)</th>
<th>Bias (x100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>PDOLS</td>
<td>16.12</td>
<td>12.91</td>
</tr>
<tr>
<td>PDOLS-DUM</td>
<td>4.83</td>
<td>3.53</td>
</tr>
<tr>
<td>CCE</td>
<td>3.32</td>
<td>2.82</td>
</tr>
<tr>
<td>AMG</td>
<td>3.14</td>
<td>2.70</td>
</tr>
<tr>
<td>CCE-2SLS</td>
<td>5.19</td>
<td>3.80</td>
</tr>
</tbody>
</table>

Note: 2,000 Monte Carlo Simulations with N = 50 and variable T. The instruments \(\Delta x_t, x_{t-2},\) and \(x_{t-3}\) were used in CCE-2SLS.

Table 3—Simulation Results - Scenario 2

<table>
<thead>
<tr>
<th>(N = 50, T)</th>
<th>RMSE (x100)</th>
<th>Bias (x100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>PDOLS</td>
<td>19.21</td>
<td>15.67</td>
</tr>
<tr>
<td>PDOLS-DUM</td>
<td>7.11</td>
<td>5.43</td>
</tr>
<tr>
<td>CCE</td>
<td>4.36</td>
<td>3.33</td>
</tr>
<tr>
<td>AMG</td>
<td>4.38</td>
<td>3.42</td>
</tr>
<tr>
<td>CCE-2SLS</td>
<td>4.35</td>
<td>3.32</td>
</tr>
</tbody>
</table>

Note: 2,000 Monte Carlo Simulations with N = 50 and variable T. The instruments \(\Delta x_t\) and \(x_{t-2}\) were used in CCE-2SLS.

The results for scenario 3 are presented in Table 4, which features an endogenous regressor. Strong bias is now present in CCE and AMG at all lengths of T considered here. CCE-2SLS successfully removes the bias from the endogenous regressor and provides a very substantial improvement in RMSE over all other estimators considered in this study. PDOLS-DUM follows behind in performance terms, performing similarly here than in Scenario 2. Figure 1 illustrates the relative performance in terms of RMSE between CCE, PDOLS-DUM, and CCE-2SLS as \(\sigma_e\) increases. One of the effects of a rising \(\sigma_e\) is an increase in the severity of the endogeneity between \(x_{it}\) and \(\varepsilon_{it}\), and therefore the graph determines how sensitive the relative performance is to more severe endogenous effects. The degree of bias found in CCE-2SLS remains very small as \(\sigma_e\) increases, while the RMSE increases very slowly. In contrast, the amount of bias in CCE and PDOLS-DUM increases very fast at first and then begins to temper off. The difference between the three estimators diverge rapidly until around \(\sigma_e = 0.3\) when they begin to separate more slowly. Therefore, it is clear that CCE-2SLS provides a very significant improvement over CCE and PDOLS-DUM in static panel models with cross-sectional dependence and endogenous regressors, and this improvement
only becomes more significant as the degree of endogeneity increases.

Table 4—Simulation Results - Scenario 3

<table>
<thead>
<tr>
<th>(N = 50, T)</th>
<th>RMSE (x100)</th>
<th>Bias (x100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>PDOLS-DUM</td>
<td>6.59</td>
<td>5.36</td>
</tr>
<tr>
<td>CCE</td>
<td>25.29</td>
<td>22.20</td>
</tr>
<tr>
<td>CCE-2SLS</td>
<td>4.34</td>
<td>3.31</td>
</tr>
</tbody>
</table>

Note: 2,000 Monte Carlo Simulations with N = 50 and variable T. The instruments \(x_{t-1}, x_{t-2},\) and \(y_{t-2}\) were used for CCE-2SLS.

Figure 1. Root Mean Squared Error by the Variance of the Error Term

Notes: The figure shows the Root Mean Squared Error of CCE, PDOLS-DUM, and CCE-2SLS as \(\sigma_\varepsilon\) varies in Scenario 3.

The results for scenario 4, which is a dynamic panel model with strictly exogenous regressors, are presented in Table 5. The table separates results by the
AR-coefficient $\phi$ and the short run effect of the regressor $\beta$. DCCE possesses bias for both $\phi$ and $\beta$ at all sample lengths considered in the study. It does, however, offer an improvement over standard CCE when $T > 30$ (it appears that the advantage of using cross-sectional lags in DCCE relative to CCE is overwhelmed from the removal of observations when the sample size is very small). In contrast, DCCE-2SLS possesses only minor bias for $\phi$ until $T \leq 50$ (when it disappears entirely), while bias for $\beta$ is insignificant at any sample length. It also features a significantly lower RMSE compared to CCE or DCCE. Accordingly, even with strictly exogenous regressors, the results suggest that using instruments in the DCCE regression improves the performance of the estimator significantly.

### Table 5—Simulation Results for $\phi$ and $\beta$ - Scenario 4

<table>
<thead>
<tr>
<th>$(N = 50, T)$</th>
<th>RMSE (x100)</th>
<th>Bias (x100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Results for $\phi$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCE</td>
<td>8.77</td>
<td>7.07</td>
</tr>
<tr>
<td>DCCE</td>
<td>10.62</td>
<td>6.99</td>
</tr>
<tr>
<td>DCCE-2SLS</td>
<td>5.47</td>
<td>3.73</td>
</tr>
<tr>
<td>Results for $\beta$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCCE</td>
<td>12.14</td>
<td>9.19</td>
</tr>
<tr>
<td>DCCE-2SLS</td>
<td>7.83</td>
<td>5.43</td>
</tr>
</tbody>
</table>

*Note:* 2,000 Monte Carlo Simulations with $N = 50$ and variable $T$. The instruments $\Delta x_t$, $x_{t-2}$, $x_{t-3}$, and $x_{t-4}$ were used in DCCE-2SLS.

The results for scenario 5 are presented in Table 6. Bias for $\phi$, which is present for all estimators considered here, decreases significantly as $T$ increases for all estimators save CCE. Evidently, the error introduced from the truncation of the infinite polynomial distributed lag function, as seen in (24), will not diminish as time increases unless lags of the cross section averages are used. For both $\phi$ and $\beta$, DCCE-2SLS shows an improvement over DCCE in terms of mean bias and RMSE.

The results for scenario 6 are presented in Table 7, a dynamic model featuring endogenous regressors. The endogeneity of the regressors has noticeably increased the level of bias found in all estimators for both $\phi$ and $\beta$ relative to scenario 4 or 5. Nevertheless, DCCE-2SLS displays a significant improvement in terms of bias and RMSE relative to CCE or DCCE. For $T \leq 50$ DCCE performs worse than CCE. Furthermore, while the degree of bias only lessens slightly as $T$ increases in both CCE and DCCE, it declines rapidly for DCCE-2SLS. This demonstrates the
Table 6—Simulation Results for $\phi$ and $\beta$ - Scenario 5

<table>
<thead>
<tr>
<th>$(N = 50, T)$</th>
<th>RMSE (x100)</th>
<th>Bias (x100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Results for $\phi$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCE</td>
<td>8.57</td>
<td>6.96</td>
</tr>
<tr>
<td>DCCE</td>
<td>9.92</td>
<td>6.53</td>
</tr>
<tr>
<td>DCCE-2SLS</td>
<td>7.00</td>
<td>4.67</td>
</tr>
</tbody>
</table>

Results for $\beta$

| CCE          | 7.68        | 7.63        | 7.78        | 8.19        | 8.50        | 6.47        | 6.66        | 6.89        | 7.32        | 7.65        |
| DCCE         | 6.31        | 5.32        | 5.03        | 3.83        | 3.12        | 3.89        | 3.69        | 3.52        | 2.46        | 1.56        |
| DCCE-2SLS    | 7.08        | 5.07        | 4.31        | 3.26        | 2.94        | 2.57        | 1.44        | 0.78        | 0.05        | 0.07        |

Note: 2,000 Monte Carlo Simulations with $N = 50$ and variable $T$. The instruments $\Delta x_t$, $x_{t-2}$, $x_{t-3}$, and $x_{t-4}$ were used in DCCE-2SLS.

The importance of a large sample size and adopting an instrumental variable approach to estimating the DCCE equation in any study using a dynamic panel time series model with endogenous regressors.

Table 7—Simulation Results for $\phi$ and $\beta$ - Scenario 6

<table>
<thead>
<tr>
<th>$(N = 50, T)$</th>
<th>RMSE (x100)</th>
<th>Bias (x100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Results for $\phi$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCE</td>
<td>18.49</td>
<td>17.45</td>
</tr>
<tr>
<td>DCCE</td>
<td>20.64</td>
<td>18.41</td>
</tr>
<tr>
<td>DCCE-2SLS</td>
<td>16.47</td>
<td>10.85</td>
</tr>
</tbody>
</table>

Results for $\beta$

| CCE          | 47.78       | 44.87       | 43.11       | 39.86       | 38.30       | 45.58       | 42.80       | 41.20       | 38.18       | 36.75       |
| DCCE         | 52.07       | 48.30       | 46.47       | 41.32       | 38.84       | 49.67       | 45.90       | 44.15       | 39.15       | 36.76       |
| DCCE-2SLS    | 32.49       | 19.27       | 15.09       | 8.61        | 6.20        | 24.23       | 15.61       | 12.66       | 7.17        | 4.65        |

Note: 2,000 Monte Carlo Simulations with $N = 50$ and variable $T$. The instruments $x_{t-1}$, $x_{t-2}$, and $y_{t-2}$ were used for DCCE-2SLS.

In conclusion, the results indicate that estimating CCE or DCCE with 2SLS (or GMM), instead of OLS, succeeds in removing bias and substantially improving RMSE in static estimation models with endogenous regressors, while standard CCE performs adequately (albeit with slightly more bias) when the regressors are merely weakly exogenous. For dynamic estimation problems, however, the
results show an unqualified improvement for estimating with 2SLS/GMM, often quite substantially, regardless of whether the regressors are strictly exogenous, weakly exogenous or endogenous.

3. Empirical Application: Inequality and Crime

The topic of inequality and crime is a fitting empirical application for panel time series econometrics, as it demonstrates the degree to which robust estimation can significantly impact empirical results. There is a long history of studies that investigate the impact that inequality has on crime (e.g. Ehrlich (1973)), Kelly (2000), and Fajnzylber, Lederman and Loayza (2002)). Studies finding a positive relation, usually using cross-country statistics, agree with the theoretical literature (such as Becker (1968)) that high inequality increases the economic incentives for criminal activity.

Chintrakarn and Herzer (2012) is potentially the first article in the literature to estimate the relationship using a panel time series approach. Their conclusion, quite surprisingly, is that the top 10% income share and the Gini coefficient has a negative relationship with the violent crime rate (an elasticity of -0.9 to -1.0). The estimator they used was Panel Dynamic OLS (both a pooled and mean group version), which they argue is robust to serial correlation and endogeneity. However, as shown above the unbiasness of this estimator relies on cross-sectional independence. The assumption that inequality and crime do not share any common factors is problematic, considering the number of factors that could easily be related across both variables and across panel units (e.g. education, unemployment, government policy, and other national trends). It is possible that time-demeaning the data will remove the bias, but relies on a number of assumptions on the structure of the factor loadings.

In order to determine the existence and direction of a long run relationship between inequality and crime, this application will test a variety of estimators on the following reduced form equations:

\[
\begin{align*}
\text{log}(\text{ViolentCrimeRate}_{it}) &= \beta_0 + \beta_1 \text{log}(\text{Top10\%incomeshare}_{it}) + e_{it} \\
\text{log}(\text{PropertyCrimeRate}_{it}) &= \beta_0 + \beta_1 \text{log}(\text{Top10\%incomeshare}_{it}) + e_{it}
\end{align*}
\]

The same dataset used in Chintrakarn and Herzer (2012) will be utilised here. According to the results in their paper the variables are nonstationary and cointe-

---

8For instance, if the common factors are stationary and follow a random factor loadings model such as the first scenario in the simulation results above.
Crime data is taken from the FBI Uniform Crime Reports, and measures for income inequality on a US state level is taken from Frank (2009). The data ranges from 1960-2012 across all U.S. States (including the District of Columbia), and accordingly \( T = 52 \) and \( N = 51 \). This is a reasonable size and very similar to what was seen in the Monte Carlo simulations above. All variables will be transformed into logs before estimation in order to compute the long run elasticities.

To determine if common factors are present in the data, a formal test for cross section dependence is undertaken. The specific test was formulated by Pesaran (2004), under a null hypothesis of no cross-sectional dependence. The violent crime rate and property crime rate will be tested separately, as well as the top 10% income share and the residuals from pooled OLS regressions of (57) and (58). All variables (save the residuals) receive a log transformation before testing. Table 3 presents the results. Given the value of the test statistics, it is clear that cross section dependence pervades both the variables and the residuals of this dataset, and therefore using a panel time series estimator that is not robust to the existence of common factors will likely yield incorrect inference.

Table 8—Pesaran (2004) Test for Cross Section Dependence

<table>
<thead>
<tr>
<th>Variable</th>
<th>Test Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Top 10% Income Share)</td>
<td>225.32</td>
<td>0.00</td>
</tr>
<tr>
<td>log(Violent Crime Rate)</td>
<td>226.84</td>
<td>0.00</td>
</tr>
<tr>
<td>log(Property Crime Rate)</td>
<td>222.79</td>
<td>0.00</td>
</tr>
<tr>
<td>Residuals from Violent Crime Regression</td>
<td>199.65</td>
<td>0.00</td>
</tr>
<tr>
<td>Residuals from Property Crime Regression</td>
<td>217.96</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 9 presents the estimation results from a variety of estimators on the two basic equations outlined above. The results are highly inconsistent across estimators. Panel Dynamic OLS offers a different prediction, depending on whether time dummies are used or not used. Without time dummies, PDOLS reports a positive average elasticity with the violent crime rate of over 1, but an insignificant average elasticity for the property crime rate. With time dummies, the elasticity is negative, statistically significant, and larger than 1 for both violent crime and property crime. CCE reports a more mild positive elasticity with the violent crime rate, relative to PDOLS, and also reports an insignificant coefficient for the property crime rate. CCE-2SLS and CCE-GMM report an average elasticity of around 1 with the violent crime rate, and an insignificant coefficient for the property crime rate.

What information can be gained from these results? First of all, it demon-

---

9Available at: http://www.disastercenter.com/crime/
strates that with real world data the results can be highly sensitive to the type of panel time series estimator used, and accordingly it is important to ensure the most appropriate estimator is utilised. It is also clear that the negative elasticity reported in Chintrakarn and Herzer (2012) relies on the use of an inappropriate estimator. In these results there is evidence for a positive relationship between violent crime and inequality in US states, but no evidence for a positive relationship with property crime. Further analysis is required in order for the relationship between inequality and crime to be better understood.

**Table 9—Regression Table - Inequality and Crime**

<table>
<thead>
<tr>
<th>Dep. Var.: log(Violent Crime Rate)</th>
<th>Estimator</th>
<th>$\beta$</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDOLS</td>
<td>1.172</td>
<td>17.6***</td>
<td></td>
</tr>
<tr>
<td>PDOLS-DUM</td>
<td>-1.252</td>
<td>-4.9***</td>
<td></td>
</tr>
<tr>
<td>CCE</td>
<td>0.450</td>
<td>4.68***</td>
<td></td>
</tr>
<tr>
<td>CCE-2SLS</td>
<td>0.987</td>
<td>11.37***</td>
<td></td>
</tr>
<tr>
<td>CCE-GMM</td>
<td>1.059</td>
<td>11.86***</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dep. Var.: log(Property Crime Rate)</th>
<th>Estimator</th>
<th>$\beta$</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDOLS</td>
<td>-0.19</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>PDOLS-DUM</td>
<td>-1.47</td>
<td>-17.3***</td>
<td></td>
</tr>
<tr>
<td>CCE</td>
<td>-0.056</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td>CCE-2SLS</td>
<td>-0.124</td>
<td>-0.69</td>
<td></td>
</tr>
<tr>
<td>CCE-GMM</td>
<td>-0.145</td>
<td>-0.68</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Inequality is measured by the top 10% income share. Data provided by Frank (2009) and the FBI’s Uniform Crime Reports. All mean group point estimates are weighted by the standard error of the individual coefficient estimates. The instruments $x_{t-1}$, $x_{t-2}$, $x_{t-3}$, $y_{t-1}$, $y_{t-2}$, and $y_{t-3}$ were used in CCE-2SLS and CCE-GMM. Cross section averages of $\log(\text{Population})$ and $\log(\text{DisposableIncomePerCapita})$ were also used as covariates in CCE, CCE-2SLS, and CCE-GMM.

4. Conclusion

This article proposed an extension to the Common Correlated Effects approach to estimating static and dynamic panel data models suffering from cross-sectional dependence. It replaces OLS with an instrumental variable approach to estimation through 2SLS or GMM, using lags of the variables as instruments for any weakly exogenous or endogenous regressors. Monte Carlo simulation results show a significant improvement over existent estimators in a number of scenarios with unobserved common factors and heterogeneous slope coefficients. In static panel data models it is able to remove all of the bias found in CCE with weakly exogenous or endogenous regressors, and significantly reduces the size of the RMSE with endogenous regressors. Furthermore, in dynamic panel data models it shows
a significant improvement over the original DCCE estimator in small samples regardless of whether the regressors were strictly exogenous, weakly exogenous, or endogenous.

This estimator, along with others, was then applied to the empirical issue of the existence of a long run relationship between inequality and crime. Using a panel time series dataset of U.S. states over the last fifty years, the results found evidence for a positive relationship between inequality and violent crime, but no evidence for a positive relationship between inequality and property crime. This result is contrary to previous research that concluded (inappropriately) from the same dataset that there is evidence for a negative relationship, illustrating how sensitive inference can be across estimators in large panel datasets. Accordingly, it is important to carefully consider the most appropriate estimator for the data when conducting an applied panel time series study.

REFERENCES


Further Notation

This subsection of the mathematical appendix adds extra notation for Assumption 6 following Chudika and Pesaran (2015). First, define the projection matrix:

\[(A1) \quad M_h = I_{T-P_T} - H_w(H_w' H_w)^+ H_w'\]

where
\[
H_w = \begin{pmatrix}
1 & h_{w,pt+1} & h_{w,pt} & \cdots & h_{w,1} \\
1 & h_{w,pt+2} & h_{w,pt+1} & \cdots & h_{w,2} \\
\vdots & \vdots & \vdots & & \vdots \\
1 & h_{w,T} & h_{w,T-1} & \cdots & h_{w,T-pt}
\end{pmatrix},
\]

and \(h_{wt} = \sum_{i=1}^{N} w_i(I_{k+1} - A_{0i}^{-1} A_{1i} L)^{-1} A_{0i}^{-1} c_i + \sum_{i=1}^{N} w_i(I_{k+1} - A_{0i}^{-1} A_{1i} L)^{-1} c_i\).

Further define the positive definite matrix \(\Sigma_{i\xi}\) as:

(A2) \[
\Sigma_{i\xi} = \text{Var}[S'\Psi_{i\xi}(L)e_{it}] + \text{Var}[S'\Psi_{i\xi}(L)C^*_i f_i]
\]

where \(C^*_i = I_{k+1} - CC^+\) is the orthogonal projector onto the orthogonal complement of \(\text{Col}(C)\), the selection matrix \(S\) is:

\[
S = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & I_k \\
0 & I_k & 0
\end{pmatrix},
\]

and

(A3) \[
\Psi_{i\xi}(L) = \begin{pmatrix}
0 \\
S'_{x}
\end{pmatrix} A_{0i}^{-1} + \begin{pmatrix}
S'_{yx}(I_{k+1} - A_{0i}^{-1} A_{1i} L)^{-1} L \\
S'_{x}[(I_{k+1} - A_{0i}^{-1} A_{1i} L)^{-1} - I_{k+1}] A_{0i}^{-1}
\end{pmatrix}
\]

where \(S'_{x} = (0 \quad I_k \quad 0)\) and \(S'_{yx} = \begin{pmatrix}
1 & 0 & 0 \\
0 & I_k & 0
\end{pmatrix}\).

**Proof of Theorem 3**

The proof will lean heavily on the proof of Theorem 1 and Theorem 2 in Chudika and Pesaran (2015), as it is an identical problem save for the instrument set \(Z_{iw}\) being used to instrument the regressors \(\Xi_i\). From Assumption 7 it is true that the matrix \((Z'_{iw} \Xi_i)\) has full rank, and accordingly the projection of \(\Xi_i\) onto \(Z_{iw}\) yields a nonsingular matrix \(\hat{\Xi}_i = Z_i(Z'_{iw} Z_i)^{-1} Z'_{iw} \Xi_i\). Assuming the number of instruments exceeds the number of regressors plus the lag(s) of the dependent variable (here \(L \geq K + 1\)), the 2SLS coefficient vector is completely identified:

(A4) \[
\hat{\pi}^{2SLS}_i = (\Xi'_i Z_i(Z'_{iw} Z_i)^{-1} Z'_{iw} \tilde{M}_q Z_i(Z'_{iw} Z_i)^{-1} Z'_{iw} \Xi_i) \Xi'_i Z_i(Z'_{iw} Z_i)^{-1} Z'_{iw} y_i
\]

To see (40) follow the proof of Theorem 1 in Chudika and Pesaran (2015) with \(\hat{\Xi}_i\) in place of \(\Xi_i\). The statement of consistency reduces to:
\[ \hat{\pi}_i^{2SLS} - \pi_i = \left( \frac{\hat{\Xi}_i' \bar{M}_q \hat{\Xi}_i}{T} \right)^{-1} \frac{\hat{\Xi}_i' \bar{M}_q \epsilon_i}{T} \]

and from Assumption 7 it is true that the instruments are exogenous (with \( E(Z_i \epsilon_{it}) = 0 \)) which leads, as it does in Lemma A.4 of Chudika and Pesaran (2015), to:

\[ \hat{\Xi}_i' \bar{M}_q \epsilon_i \xrightarrow{p} 0 \]

and therefore:

\[ \hat{\pi}_i^{2SLS} - \pi_i \xrightarrow{p} 0 \]

To see (42) first note that \( \hat{\pi}_i^{2SLS} = N^{-1} \sum_{i=1}^{N} \hat{\pi}_i^{2SLS} \) and from (8) it is true that \( \pi_i = \pi + \eta_\pi \) where \( \eta_\pi \sim IID(0, \Omega_\pi) \). Accordingly, from Theorem 1 and the above:

\[ \hat{\pi}_i^{2SLS} - N^{-1} \sum_{i=1}^{N} \pi_i \xrightarrow{p} 0 \]

and

\[ N^{-1} \sum_{i=1}^{N} \pi_i - \pi = N^{-1} \sum_{i=1}^{N} \eta_\pi \xrightarrow{p} 0 \]

(42) is established from (A8) and (A9).

**Proof of Theorem 4**

From (8), (A5), and \( \hat{\pi}_i^{2SLS} = N^{-1} \sum_{i=1}^{N} \hat{\pi}_i^{2SLS} \) there is the following relation:

\[ \hat{\pi}_i^{2SLS} - \pi = N^{-1} \sum_{i=1}^{N} \pi + N^{-1} \sum_{i=1}^{N} \eta_\pi + N^{-1} \sum_{i=1}^{N} \left( \frac{\hat{\Xi}_i' \bar{M}_q \hat{\Xi}_i}{T} \right)^{-1} \frac{\hat{\Xi}_i' \bar{M}_q \epsilon_i}{T} - \pi \]

which reduces to:
\begin{equation}
\hat{\pi}^{2SLS} - \pi = N^{-1} \sum_{i=1}^{N} \eta_\pi + N^{-1} \sum_{i=1}^{N} \left( \frac{(\hat{\Xi}_i' \bar{M}_q \hat{\Xi}_i)}{T} \right)^{-1} \frac{\hat{\Xi}_i' \bar{M}_q \epsilon_i}{T}
\end{equation}

Premultiplying this equation with $\sqrt{N}$ produces the following:

\begin{equation}
\sqrt{N}(\hat{\pi}^{2SLS} - \pi) = \sqrt{N}^{-1} \sum_{i=1}^{N} \eta_\pi + \sqrt{N}^{-1} \sum_{i=1}^{N} \left( \frac{(\hat{\Xi}_i' \bar{M}_q \hat{\Xi}_i)}{T} \right)^{-1} \frac{\hat{\Xi}_i' \bar{M}_q \epsilon_i}{T}
\end{equation}

yet due to (32) it converges in distribution to:

\begin{equation}
\hat{\pi}^{2SLS} - \pi \xrightarrow{d} \sqrt{N}^{-1} \sum_{i=1}^{N} \eta_\pi
\end{equation}

As defined in Assumption 3, the variance of $\eta_\pi$ is $\Omega_\pi$, establishing the chief result of Theorem 4.