STEIN-LIKE SHRINKAGE ESTIMATION OF PANEL DATA MODELS WITH COMMON CORRELATED EFFECTS

Bai Huang[a], Tae-Hwy Lee[b] and Aman Ullah[b]

[a]*Central University of Finance & Economics, China*
[b]*University of California Riverside, United States*

ABSTRACT

This chapter examines the asymptotic properties of the Stein-type shrinkage combined (averaging) estimation of panel data models. We introduce a combined estimation when the fixed effects (FE) estimator is inconsistent due to endogeneity arising from the correlated common effects in the regression error and regressors. In this case, the FE estimator and the CCEP estimator of Pesaran (2006) are combined. This can be viewed as the panel data model version of the shrinkage to combine the OLS and 2SLS estimators as the CCEP estimator is a 2SLS or control function estimator that controls for the endogeneity arising from the correlated common effects. The asymptotic theory, Monte Carlo simulation, and empirical applications are presented. According to our calculation of the asymptotic risk, the Stein-like shrinkage estimator is more efficient estimation than the CCEP estimator.

Keywords: Endogeneity; panel data; fixed effect; common correlated effects; shrinkage; model averaging; local asymptotics; Hausman test

JEL classification: C13; C33; C52

Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling
Advances in Econometrics, Volume 40A, 249–274

ISSN: 0731-9053/doi:10.1108/S0731-90532019000040A011

1. INTRODUCTION

This chapter considers a Stein-like (1956) shrinkage estimation of panel data models. The Stein shrinkage estimators have Bayesian interpretations which leads to a model combination framework parallel to Bayesian model averaging (BMA) of Jeffreys (1961), which is one of the most commonly used model combination methods in statistical learning (see Hastie, Tibshirani, and Friedman (2009), Section 8.8). It can be shown that BMA produces a Stein-type shrinkage estimator (see Judge and Bock (1978), pp. 173−176).

In this important development of shrinkage estimation, Dale Poirier has been one of the leading researchers for more than four decades. Among his numerous contributions in Bayesian analysis in econometrics and statistics, he already developed in early 1980s an idea similar to the concept of model averaging (Hansen, 2007). The BMA assigns prior probabilities of the candidate models being the true model, while there is little guidance in the literature on elicitation of prior probabilities. An exception is the *model occurrence* framework developed by Poirier and Klepper (1981) and Poirier (1988), and applied in the study by Koop and Poirier (1995). The model occurrence is related to model averaging (e.g., Hansen, 2007) as its focus is on how the underlying model occurrence probabilities of the competing models depend on characteristics of the environments in which the data subsets are generated. In the study by Poirier and Klepper (1981), classical, Bayesian, and mixed estimation approaches are developed while the Bayesian approaches are shown to be especially attractive whenever the models are nested. Recent literature on BMA includes Draper (1995) and Raftery, Madigan, and Hoeting (1997), and a survey by Hoeting, Madigan, Raftery, and Volinsky (1999).

Although the model averaging has been one of the most active topics in the recent literature, we have not seen much for panel data models. Recently, there has been increased interest in the estimation of models with error cross-sectional dependence in panel data models. A particular form that has become popular is a common factor error structure with a fixed number of unobserved common factors and individual-specific factor loadings. The most obvious implication of error cross-sectional dependence is that the standard panel data estimators are inefficient and estimated standard errors are biased and inconsistent. One popular approach to this problem is the common correlated effects (CCEs) method proposed by Pesaran (2006). The virtue of the CCE estimation is that it can be easily computed by the least squares regression augmented using the cross-sectional averages of the dependent and explanatory variables as proxies for the factors.

In this paper we consider the common correlated effects pooled (CCEP) estimator of Pesaran (2006) in comparison with the FE estimator. If there exists CCE, the CCEP estimator is consistent and the FE model is inconsistent. On the other hand, if there is no error cross-sectional dependence, both the CCEP and FE estimators are consistent while the FE estimator is efficient. We consider the combined estimator which is a weighted combination of the FE estimator and the CCEP estimator. We study the asymptotic distribution of the combined estimator in a local asymptotic framework where some factor loadings in the error term are in a local neighborhood of zero. We show that under certain conditions, the combined (shrinkage) estimator has strictly smaller risk than the CCEP estimator. The combined estimator also has smaller asymptotic risk compared to the FE estimator unless the endogeneity is very weak. Our simulation result shows that the combined estimator can reduce finite sample MSE relative to the CCEP estimator for all degrees of endogeneity, as well as relative to FE estimator for moderate to large degrees of endogeneity.

Specifically, we consider a panel data regression model of the general form

$$y_{it} = x_{it}'\beta + \alpha_i + u_{it}, \tag{1}$$

where $i = 1,\dots,n$, $t = 1,\dots,T$, x_{it} is $q \times 1$, z_{it} is $1 \times p$, β is a $q \times 1$ parameter of our interest, $m(\cdot)$ is an unknown smooth function, α_i is the individual effect, and u_{it} is the random error. We are interested in estimation of β for the case when the error term u_{it} is correlated with the regressors x_{it}. The above model suffers from endogeneity which results in the inconsistent estimation of β.

The solution to the inconsistent estimation depends on what causes the endogeneity: (i) when the individual effect α_i is correlated with the regressors x_{it} or (ii) when the error term u_{it} is correlated with the regressors x_{it}. In both cases the above model suffers from endogeneity which results in the inconsistent estimation of β. The first case endogeneity arises when the individual heterogenous effect α_i is treated as the random error component and it is correlated with the regressors. We are interested in estimation of β for the second case, where the endogeneity arises when regressors that are omitted in the model are correlated with the included regressors x_{it} such that the regression error u_{it} and the regressors x_{it} share common factors (common effects) and thus are correlated. In the first case, the solution is to treat the individual heterogenous effect α_i as fixed effects (FE) rather than random effects. In the second case, it is to remove the effect of the misspecification, i.e., either by adding them to the model or by controlling for the omitted variables. As the omitted variables may be unobservable or their data may not be available in hand, we consider the 'control function approach' suggested by Pesaran (2006), which is essentially the 2SLS estimator for the panel data models.

In the mean time however, it is important to note that the solutions to restore the consistency will be based on less efficient estimation. Therefore, there is a trade-off between consistency (bias) and efficiency (variance). In the first case of the endogeneity when the random component α_i is correlated with x_{it}, the random effects (RE) estimator is inconsistent but more efficient and the FE estimator is consistent but less efficient. In the second case when the error term u_{it} is correlated with x_{it} due to omitted variables or unobservable variables, the FE estimator (which is an OLS estimator) is inconsistent but relatively more efficient and the CCE estimator of Pesaran (2006) (which is a 2SLS estimator) is consistent but less efficient. The CCE estimator can be thought of as a control function estimator as the correlation between u_{it} and x_{it} are modeled by a common factor specification in one step and then removed in another step, and therefore, it is a 2SLS estimator for the panel data model with the common effects in u_{it} and x_{it}.

When u_{it} and x_{it} are correlated (strong endogeneity), the CCE estimator is preferred to the FE estimator. When u_{it} and x_{it} are not correlated (no endogeneity), the FE estimator is preferred to the CCE estimator. Hence, a natural question is which one to choose when the endogenous is weak. The answer is that we combine the FE and CCE estimators when u_{it} and x_{it} are weakly correlated. Hence, this chapter extends the study by Hansen (2017) for panel data models with common correlated effects.

The rest of this chapter is organized as follows. To examine and compare the alternative estimators of β and their combined estimator, Section 2 gives the models and these estimators. Section 3 presents the asymptotic theory, with all the proofs collected in Appendix (Section 7). Section 4 provides some Monte Carlo simulation results to demonstrate the asymptotic results in finite sample. An empirical application is given in Section 5. Finally, Section 6 concludes.

2. ESTIMATING PANEL DATA MODELS WITH COMMON CORRELATED EFFECTS

Consider a panel data regression model

$$y_{it} = x_{it}'\beta + \alpha_i + u_{it}, \tag{2}$$

where $i = 1,\ldots,n, t = 1,\ldots,T, x_{it}$ is a vector of q explanatory variables, β is a $q\times 1$ unknown coefficients, α_i denotes the individual-specific effects and is assumed to be fixed. The disturbance term u_{it} has a multifactor structure

$$u_{it} = \gamma_i' f_t + \varepsilon_{it}, \tag{3}$$

in which f_t is an $r\times 1$ vector of individual-invariant time-specific unobserved common effects, γ_i is an $r\times 1$ stochastic individual-specific factor loading vector, and ε_{it} are the idiosyncratic errors assumed to be independent of x_{it}. To model the correlation between the regressors x_{it} and the errors u_{it}, the regressors may contain the unobserved common factor

$$x_{it} = \Gamma_i' f_t + v_{it}, \tag{4}$$

where Γ_i is an $r\times q$ stochastic factor loading matrix, and v_{it} is the $q\times 1$ vector of idiosyncratic components of x_{it} and is independent of the common effects f_t.

In vector notation,

$$\begin{aligned} y_i &= X_i\beta + \alpha_i \iota_T + u_i, \\ u_i &= F\gamma_i + \varepsilon_i, \\ X_i &= F\Gamma_i + v_i, \end{aligned} \tag{5}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$ is $T \times 1$, $X_i = (x'_{i1}, \ldots, x'_{iT})'$ is $T \times q$, $u_i = (u_{i1}, \ldots, u_{iT})'$ is $T \times 1$,ι_T is the $T \times 1$ vector of ones, $F = (f'_1, \ldots, f'_T)'$ is $T \times r$, $v_i = (v_{i1}, \ldots, v_{iT})'$ is $T \times q$.

We make the following assumptions on the common factors, their loadings and the individual or unit specific errors.

Assumption 1. ε_{it} is independently and identically distributed (iid) across both i and t with $E(\varepsilon_{it}) = 0$, $Var(\varepsilon_{it}) = \sigma^2 > 0$ and $E\|\varepsilon_{it}\|^4 < \infty$.

Assumption 2. v_{it} is distributed independently across both i and t with $E(v_{it}) = 0$, $Var(v_{it}) = \Sigma_i$ positive definite and $E\|v_{it}\|^4 < \infty$.

Assumption 3. f_t is covariance stationary with absolute summable autocovariances, such that $E\|f_t\|^4 < \infty$.

Assumption 4. γ_i and Γ_i are iid across i and independent of ε_{it} and v_{it}, f_t for all i and t with fixed means and finite variances. In particular,

$$\gamma_i = \gamma + \eta_i, \eta_i \sim \text{iid}\left(0,\Omega_\eta\right), \tag{6}$$

where Ω_η is an $r \times r$ symmetric nonnegative definite matrix, and $\|\gamma\| < K$, $\|\Gamma\| < K$ and $\|\Omega_\eta\| < K$ for some positive constant $K < \infty$.

Assumption 5. ε_{it}, v_{it} and f_t are mutually independent.

First, let $Q_T \equiv I_T - \iota_T(\iota'_T\iota_T)^{-1}\iota'_T$, which is a $T \times T$ symmetric idempotent matrix. Further, $Q_T\iota_T = 0$, and so for ith unit, pre-multiplying Eq. (5) by Q_T gives

$$Q_T y_i = Q_T X_i \beta + Q_T u_i. \tag{7}$$

The $\hat{\beta}_{FE}$ can be expressed as

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{n} X'_i Q_T X_i\right)^{-1} \left(\sum_{i=1}^{n} X'_i Q_T y_i\right), \tag{8}$$

and

$$\text{avar}\left(\hat{\beta}_{FE}\right) = \Psi^{*-1} R^* \Psi^{*-1}, \tag{9}$$

where $R^* = \text{plim}\left(\sigma^2 \frac{1}{n}\sum_{i=1}^{n} X'_i Q_T X_i + \frac{1}{n}\sum_{i=1}^{n} X'_i Q_T F \Omega_\eta F' Q_T X_i\right)$, and $\Psi^{*-1} = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} X'_i Q_T X_i\right)$. If $\gamma_i \neq 0$ so that the error term u_{it} and x_{it} have the common correlated effects, the FE estimator is inconsistent. If $\gamma_i = 0$, $\hat{\beta}_{FE}$ is consistent and has the following asymptotic distribution

$$\sqrt{n}\left(\hat{\beta}_{FE}-\beta\right)\xrightarrow{d}N(0,\Sigma_{FE}),\tag{10}$$

where $\Sigma_{FE}=\sigma^2\left(\text{plim}\frac{1}{n}\sum_{i=1}^{n}X_i'Q_TX_i\right)^{-1}$, under the following additional assumption:

Assumption 6. $\frac{1}{n}\sum_{i=1}^{n}X_i'Q_TX_i$ is bounded and nonsingular.

Next, let us consider the common correlated effects pooled (CCEP) estimator of Pesaran (2006). The idea underlying the common correlated effects approach is that the unobservable common factors f_t can be well approximated by a linear combination of the cross-section averages of the dependent variable and those of the regressors. To illustrate this result, we write Eqs. (2) and (4) as

$$z_{it}=\begin{pmatrix}y_{it}\\x_{it}\end{pmatrix}=B_i+C_i'f_t+e_{it},\tag{11}$$

where

$$e_{it}=\begin{pmatrix}\beta'v_{it}+\varepsilon_{it}\\v_{it}\end{pmatrix},\tag{12}$$

$$B_i=\begin{pmatrix}\alpha_i\\\mathbf{0}\end{pmatrix},C_i=\begin{pmatrix}\gamma_i & \Gamma_i\end{pmatrix}\begin{pmatrix}1 & 0\\\beta & I_q\end{pmatrix},\tag{13}$$

z_{it} is $(q+1)\times 1$, B_i is $1\times(q+1)$, $\mathbf{0}$ is a $q\times 1$ vector of zeros, C_i is $r\times(q+1)$, and I_q is an identity matrix of order q. The covariance matrix of e_{it} is given by

$$E\left(e_{it}e_{it}'\right)=\Sigma_{e,i}=\begin{bmatrix}\beta'\Sigma_i\beta+\sigma_i^2 & \beta'\Sigma_i\\\Sigma_i\beta & \Sigma_i\end{bmatrix}.\tag{14}$$

Then, the cross-sectional average is

$$\bar{z}_t=\overline{B}+\overline{C}'f_t+\bar{e}_t,\tag{15}$$

where

$$\bar{z}_t=\frac{1}{n}\sum_{i=1}^{n}z_{it},\ \overline{B}=\frac{1}{n}\sum_{i=1}^{n}B_i,\ \overline{C}=\frac{1}{n}\sum_{i=1}^{n}C_i,\ \bar{e}_t=\frac{1}{n}\sum_{i=1}^{n}e_{it}.\tag{16}$$

Although not considered here, generally one can consider $\bar{z}_t=\bar{z}_{wt}=\sum_{i=1}^{n}w_iz_{it}$, where $w_i=\sigma_i^{-2}/\sum_{j=1}^{n}\sigma_j^{-2}$. If we assume

$$\text{Rank}(\overline{C}) = r \leq q + 1, \quad \text{for all } n, \tag{17}$$

it follows that

$$f_t = (\overline{C}\overline{C}')^{-1}\overline{C}(\overline{z}_t - \overline{B} - \overline{e}_t). \tag{18}$$

Therefore, f_t can be approximated by a linear combination of $\{\overline{z}_t 1\}$, if $\overline{e}_t \rightarrow^{q.m.} 0$ as $n \rightarrow \infty$ (cf. Lemma 1 in Pesaran, 2006). In such a case, we obtain

$$f_t - (CC')^{-1}C(\overline{z}_t - \overline{B}) \xrightarrow{p} 0 \text{ as } \quad n \rightarrow \infty, \tag{19}$$

where

$$\overline{C} \xrightarrow{p} C = \tilde{\Gamma}\begin{pmatrix} 1 & 0 \\ \beta & I_k \end{pmatrix} \text{as} \quad n \rightarrow \infty \tag{20}$$

and $\tilde{\Gamma} = \left(E(\gamma_i)\ E(\Gamma_i)\right) = (\gamma\Gamma)$.

From Eqs. (2) and (3), y_{it} is generated as

$$y_{it} = x_{it}'\beta + \alpha_i + \gamma_i' f_t + \varepsilon_{it}. \tag{21}$$

Next, substitute $f_t = (CC')^{-1}C(\overline{z}_t - \overline{B})$ into (21),

$$\begin{aligned} y_{it} &= x_{it}'\beta + \alpha_i + \gamma_i'(CC')^{-1}C(\overline{z}_t - \overline{B}) + \varepsilon_{it} \\ &= x_{it}'\beta + \left(\alpha_i - \gamma_i'(CC')^{-1}C\overline{B}\right) + \gamma_i'(CC')^{-1}C\overline{z}_t + \varepsilon_{it} \\ &= x_{it}'\beta + \overline{h}_t' c_i + \varepsilon_{it}, \end{aligned} \tag{22}$$

where $c_i = \left[\left(\alpha_i - \gamma_i'(CC')^{-1}C\overline{B}\right)\gamma_i'(CC')^{-1}C\right]'$ is $(q+2)\times 1$ and $\overline{h}_t = \left(1\overline{z}_t'\right)'$ is $(q+2)\times 1$. This suggests using $\overline{h}_t = \left(1\overline{z}_t'\right)'$ as an observable proxy for f_t. In vector notation,

$$y_i = X_i\beta + \overline{H}c_i + \epsilon_i, \tag{23}$$

where $\overline{H} = \left(\iota_T \overline{Z}\right)$ is $T\times(q+2)$, $\overline{Z} = (\overline{z}_1, \ldots, \overline{z}_T)$ is $T\times(q+1)$. Let $\overline{M} = I_T - \overline{H}\left(\overline{H}'\overline{H}\right)^{-1}\overline{H}'$. Since $\overline{M}\overline{H} = 0$,

$$\overline{M}y_i = \overline{M}X_i\beta + \overline{M}\epsilon_i. \tag{24}$$

Now, we state the following assumptions:

Assumption 7. $\frac{1}{n}\sum_{i=1}^{n} X_i'\overline{M}X_i$ is bounded and nonsingular.

The CCE estimator can be obtained by performing OLS on the resulting transformed model

$$\hat{\beta}_{CCEP} = \left(\frac{1}{n}\sum_{i=1}^{n} X_i'\overline{M}X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'\overline{M}y_i\right). \tag{25}$$

Following Pesaran (2006), for fixed T and $n \to \infty$, the asymptotic for CCEP estimator still holds. Under Assumptions 1–5, 7 and with the rank condition (17) satisfied,

$$\sqrt{n}\left(\hat{\beta}_{CCEP} - \beta\right) \xrightarrow{d} N(0, \Sigma_{CCEP}), \tag{26}$$

where $\Sigma_{CCEP} = \Psi^{-1} R \Psi^{-1}$, $R = \text{plim}\left(\sigma^2 \frac{1}{n}\sum_{i=1}^{n} X_i' M_f X_i + \frac{1}{n}\sum_{i=1}^{n} X_i' M_f F \Omega_\eta F' M_f X_i\right)$, $\Psi^{-1} = \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} X_i' M_f X_i\right)$, $M_f = I - F_1(F_1'F_1)^{-1}F_1'$, and $F_1 = (\iota_T F)$.

With the common correlated effects, the FE estimator (which is an OLS) does not produce consistent estimates of the coefficients β of interest. Pesaran (2006) suggests the CCEP approach that yields consistent estimation in the presence of the correlated unobserved common effects. We now extend Hansen (2017) to propose the following combined estimator of β, which is a weighted combined FE and CCEP estimator with weights depending on Hausman (1978) statistic:

$$\hat{\beta}_c = w\hat{\beta}_{FE} + (1 - w)\hat{\beta}_{CCEP}, \tag{27}$$

where

$$w = \begin{cases} \dfrac{\tau}{H_n} & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases}, \tag{28}$$

$$H_n = n(\hat{\beta}_{CCEP} - \hat{\beta}_{FE})'\left[Var\left(\sqrt{n}\left(\hat{\beta}_{CCEP} - \hat{\beta}_{FE}\right)\right)\right]^{-1}\left(\hat{\beta}_{CCEP} - \hat{\beta}_{FE}\right), \tag{29}$$

where τ is a shrinkage parameter. The degree of shrinkage depends on τ/H_n. When $H_n < \tau$ then $\hat{\beta}_c = \hat{\beta}_{FE}$, When $H_n \geq \tau$ then $\hat{\beta}_c$ is a weighted average of $\hat{\beta}_{FE}$ and $\hat{\beta}_{CCEP}$, with more weight on $\hat{\beta}_{CCEP}$ when H_n is large.

3. ASYMPTOTICS

The variable x_{it} is exogenous if $\gamma_i = 0$. We use the local asymptotic approach. For fixed T, γ_i is local to zero by setting

$$\gamma_i = \tilde{\gamma}_i \rho, \tag{30}$$

$$\rho = \frac{1}{\sqrt{n}}\delta, \tag{31}$$

where $\tilde{\gamma}_i$ is an $r \times 1$ constant, and ρ, δ are scalars. Thus the correlation between x_i and u_i is local to zero. When $\delta = 0$, x_{it} are exogenous. When $\delta \neq 0$, x_{it} are endogenous. δ controls the degree of endogeneity. We make a further assumption:

Assumption 8. X_i, $i = 1,\ldots,n$, are iid over i. $E\|x_{it}\|^{2+K} < \infty$ for some $K > 0$. $E\|x_{it}\|^4 < \infty$. $V_1 = \sigma_\varepsilon^2\left(\text{plim}\frac{1}{n}\sum_{i=1}^n X_i' Q_T X_i\right)^{-1}$ and $V_2 = \left(\text{plim}\frac{\Psi^{-1}R\Psi^{-1}}{n}\right)^{-1}$, where $\sigma_\varepsilon^2 = E\left(\varepsilon_{it}^2\right)$ is the variance of the idiosyncratic error in Eq. (3).

Theorem 1. Under Assumptions 1–8,

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{FE} - \beta \\ \hat{\beta}_{CCEP} - \beta \end{pmatrix} \xrightarrow{d} h + \xi, \tag{32}$$

where

$$h = \begin{pmatrix} \sigma_\varepsilon^{-2} V_1 \Sigma \delta \\ 0 \end{pmatrix}, \qquad \text{with} \quad \Sigma \equiv \text{plim}\frac{1}{n}\sum_{i=1}^{n} X_i' Q_T F \tilde{\gamma}_i, \tag{33}$$

and

$$\xi \sim N(0, V), \qquad \text{with} \quad V = \begin{pmatrix} V_1 & V_{21}' \\ V_{21} & V_2 \end{pmatrix}. \tag{34}$$

Furthermore,

$$H_n \xrightarrow{d} (h + \xi)' B (h + \xi), \tag{35}$$

and

$$\sqrt{n}\left(\hat{\beta}_c - \beta\right) \xrightarrow{d} \tilde{\Psi} = G_2' \xi - \left(\frac{\tau}{(h + \xi)' B (h + \xi)}\right)_1 G'(h + \xi), \tag{36}$$

where $B = G\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-1} G'$, $G_1 = (I0)'$, $G_2 = (0I)'$, $G = G_2 - G_1 = (-II)'$ *and* $(a)_1 = \min[1,a]$.

Theorem 1 gives expressions for the joint asymptotic distribution of the FE and CCEP estimator, the Hausman statistic, and the combined estimator as a transformation of the normal random vector ξ and the noncentrality parameter h under the local exogeneity assumption. As noted in Poirier (1995, p. 284) the mean and covariance matrix of the Stein-like combined estimator are complicated. See Judge and Bock (1978, pp. 172–173) for details. In our case, the asymptotic distribution of $\hat{\beta}_c$ is written as a random weighted average of the asymptotic distribution of $\hat{\beta}_{FE}$ and $\hat{\beta}_{CCEP}$, as shown in Eq. (36).

Remark 1. Theorem 1 extends Hansen (2017) for the panel data models and generalizes his results by allowing $V_1 \neq V_{12}$ and $B = G(V_1 + V_2 - V_{12} - V_{21})^{-1}G'$ to be asymmetric. If $\hat{\beta}_{FE}$ is fully efficient, then $V_1 = V_{12} = V_{21}$ and $B = G(V_2 - V_1)^{-1}G'$ as in the case of Hansen (2017). In general $\hat{\beta}_{FE}$ may not be fully efficient and so $V_1 \neq V_{12}$ and $B = G(V_1 + V_2 - V_{12} - V_{21})^{-1}G'$. In that case the derivation of V_{12}, V_{21} is required or they need to be estimated by the use of bootstrap as we do in this paper in Sections 4 and 5.

Next, we compare $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_c$ in the asymptotic risk. The asymptotic risk of any sequence of estimators β_n of β can be defined as

$$R(\beta_n, \beta, W) = \lim_{n \to \infty} E\left[n(\beta_n - \beta)' \, W(\beta_n - \beta)\right] = R(\beta_n). \tag{37}$$

Define the largest eigenvalue of the matrix $\frac{A+A'}{2}$ and $\frac{A^*+A^{*'}}{2}$

$$\lambda_1 = \lambda_{\max}\left(\frac{A + A'}{2}\right), \tag{38}$$

$$\lambda_1^* = \lambda_{\max}\left(\frac{A^* + A^{*'}}{2}\right), \tag{39}$$

where

$$A = \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}} W(V_2 - V_{21})\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-\frac{1}{2}}, \tag{40}$$

$$A^* = \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}} W\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}}. \tag{41}$$

Let

$$d = \frac{\mathrm{tr}(W(V_2 - V_{21}))}{\lambda_1}. \tag{42}$$

Theorem 2. Under Assumptions 1–8, if

$$d > 2, \tag{43}$$

and

$$0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}, \tag{44}$$

then $R\left(\hat{\beta}_c\right) = \text{tr}\left[WE\left(\tilde{\Psi}\tilde{\Psi}'\right)\right]$,

$$R\left(\hat{\beta}_{CCEP}\right) = \text{tr}(WV_2),$$

and

$$R\left(\hat{\beta}_c\right) < R\left(\hat{\beta}_{CCEP}\right) - \frac{\tau\left(2\lambda_1(d-2) - \lambda_1^*\tau\right)}{\sigma^{-4}\delta\Sigma' V_1\left[V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right]^{-1} V_1\Sigma\delta + q}. \tag{45}$$

□

Equation (45) shows that the asymptotic risk of the combined estimator is strictly less than that of the CCEP estimator, so long as τ satisfies the condition (44). τ appears in the risk bound (45) as a quadratic expression, so there is an optimal choice $\tau^* = \frac{\lambda_1(d-2)}{\lambda_1^*}$ which minimizes this bound. The assumption $d > 2$ is the critical condition needed in order for the right-hand side of Eq. (44) to be positive, which is necessary for the existence of τ satisfying Eq. (44).

Poirier (1995, p. 284) noted the fact that the asymptotic risk of the combined estimator is strictly less than that of the CCEP estimator, $R\left(\hat{\beta}_c\right) < R\left(\hat{\beta}_{CCEP}\right)$, does not imply that $MSE\left(\hat{\beta}_c\right) < MSE\left(\hat{\beta}_{CCEP}\right)$ for each element of the estimator of the $q \times 1$ coefficients of β. Poirier (1995) also wrote:

> While it may appear that [the James-Stein estimator] is some sort of mathematical trick pulled out of the air, this is not the case. [It] can in fact be given a Bayesian interpretation.

Corollary 1. $R\left(\hat{\beta}_c\right) - R\left(\hat{\beta}_{CCEP}\right) < 0$, for $d > 2$ and $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$. In the case $W = (V_2 - V_{21})^{-1}$, $0 < \tau \leq 2\left(\frac{q-2}{\lambda_1^*}\right)$ and $q > 2$ which is Stein's (1956) classic condition for shrinkage. □

See Poirier (1995, p. 283) for more discussion on Stein's (1956) classic condition for shrinkage. The following two corollaries are obtained with $W = (V_2 - V_1)^{-1}$.

Corollary 2. $R\left(\hat{\beta}_{FE}\right) = \text{tr}(WV_1) + \sigma^{-4}\delta'\Sigma' V_1 WV_1\Sigma\delta$; $R\left(\hat{\beta}_{FE}\right) \leq R\left(\hat{\beta}_{CCEP}\right)$ when $\sigma^{-4}\delta'\Sigma' V_1 WV_1\Sigma\delta \leq q$, and $R\left(\hat{\beta}_{FE}\right) > R\left(\hat{\beta}_{CCEP}\right)$ otherwise. □

Corollary 3. $R\left(\hat{\beta}_c\right) - R\left(\hat{\beta}_{FE}\right) < 0$, for $q < \sigma^{-4}\delta'\Sigma' V_1 WV_1\Sigma\delta$, $d > 2$, and $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$. □

Corollary 2 indicates that when endogeneity is weak (γ_i and hence δ_i is close to zero) the FE estimator may perform better than the CCEP estimator. Corollary 3 indicates that when endogeneity is strong, $d > 2$, $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$, the combined estimator performs best among these three estimators.

4. MONTE CARLO

We now investigate the finite sample MSE of our combined estimator in the following simulation design,

$$y_{it} = \alpha_i + \beta' x_{it} + \gamma_i' f_t + \varepsilon_{it}, \tag{46}$$

$$x_{it} = \Gamma_i' f_t + v_{it}, \tag{47}$$

where α_i is drawn from $N(0,1)$, $\varepsilon_{it} \sim$ iid $N(0,1)$, $v_{it} \sim$ iid $N(0,1)$. The factor is drawn from $N(0, I_r)$. We vary $n \in \{50, 100\}$, $T \in \{8, 16\}$, $q \in \{1, 2, 3\}$, $r \in \{1, 3\}$ and $\beta \in \{0, 1\}$. The parameters of the unobserved common effects in the x_{it} equation are generated independently across replications as $\Gamma_i = (\Gamma_{i1}\ \Gamma_{i2}\ \Gamma_{i3})$ with $\Gamma_{i1} \sim$ iid $N(0.5, 0.5)$, $\Gamma_{i2} \sim$ iid $N(0, 0.5)$, and $\Gamma_{i3} \sim$ iid $N(0, 0.5)$. Let $\gamma_i = \tilde{\gamma}_i \rho$ where $\tilde{\gamma}_{i1} \sim$ iid $N(1, 0.1)$, $\tilde{\gamma}_{i2} \sim$ iid $N(1, 0.1)$, $\tilde{\gamma}_{i3} \sim$ iid $N(1, 0.1)$. We consider ρ on a 40-point grid on $[0, 0.975]$. ρ controls the degree of endogeneity.

We generated 5,000 samples on each calculated $\hat{\beta}_{CCEP}$, $\hat{\beta}_{FE}$, $\hat{\beta}_c$, for the latter we set $\tau = 1/4$ for $q = 1$, $\tau = 1$ for $q = 2$, and $\tau = \tau^*$ otherwise. We also calculated the Hausman pre-test estimator:

$$\hat{\beta}_{PT} = \hat{\beta}_{FE} 1(H_n < c) + \hat{\beta}_{CCEP} 1(H_n \geq c),$$

where c is the 5% critical value from the χ_q^2 distribution. We compare the estimator by relative MSE

$$MSE\left(\hat{\beta}\right) = E(\hat{\beta} - \beta)'(\hat{\beta} - \beta), \tag{48}$$

which we normalize by MSE of the CCEP estimator. Thus, the values less than one indicate improved precision relative to CCEP estimator, and values greater than one indicate worse performance than the CCEP estimator.

We do a bootstrap pairs procedure that resample with replacement over i and uses all observed time periods for a given individual. For data $\{(y_i, X_i), i = 1, \ldots, n\}$, this yields B pseduo-samples, and for each pseudo-sample, we perform regression, yielding B estimates, $b = 1, \ldots, B$. The panel bootstrap estimate of the variance matrix is then given by

$$\hat{V}_{Boot}\left(\hat{\beta}_{CCEP} - \hat{\beta}_{FE}\right) = \frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}_b - \bar{\hat{\theta}}\right)\left(\hat{\theta}_b - \bar{\hat{\theta}}\right)'. \tag{49}$$

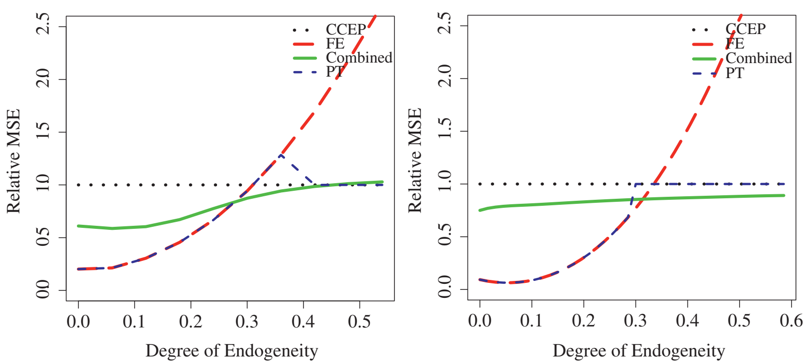
$\underline{b}$ denotes the bth of B bootstrap replications, and $\hat{\theta} = \hat{\beta}_{CCEP} - \hat{\beta}_{FE}$, $\bar{\hat{\theta}} = B^{-1}\sum_b \hat{\theta}_b$.

The dotted line (black) is the normalized MSE of the CCEP estimator, the solid line (green) is the normalized MSE of the combined estimator, the longer dashed line (red) is the normalized MSE of the FE estimator, and the shorter dashed line (blue) is the normalized MSE of the pre-test estimator.

Figs. 3, 4, 7, and 8 plots the relative MSE of $\hat{\beta}_{CCEP}$, $\hat{\beta}_{FE}$, $\hat{\beta}_c$, and $\hat{\beta}_{PT}$, respectively, with $q = \{1, 2\}$. These are the cases where the Eqs. (43) and (44) are not satisfied, which are sufficient conditions for Theorem 2 to hold, $R\left(\hat{\beta}_c\right) < R\left(\hat{\beta}_{CCEP}\right)$. Because these conditions are just sufficient but not necessary, the theorem may or may not hold for these smaller values of $q = 1, 2$. Indeed, $q = 1$ makes the results quite erratic, showing that Theorem 2 does not hold for $q = 1$, and somewhat less degree for $q = 2$.

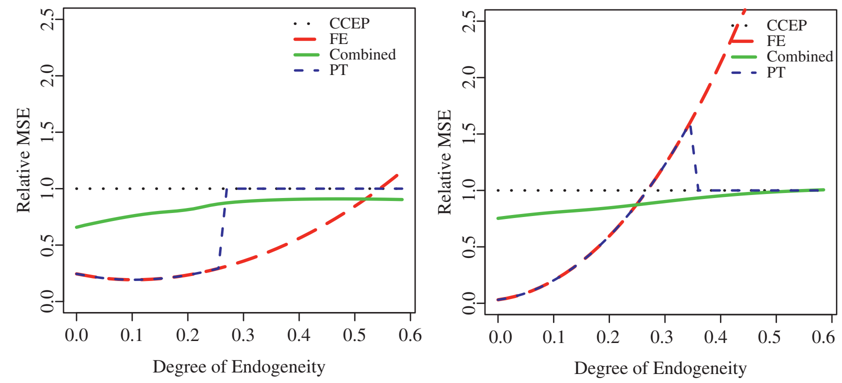
In each Figure, subfigure (a) plots the relative MSE for $n = 50$; subfigure (b) plot the relative MSE for $n = 100$. Figs. 1, 2, 3, and 4 plot the relative MSE for $t = 8$; Figs. 5, 6, 7, and 8 plot the relative MSE for $t = 16$. We see that the region of dominance for the combined estimator over the FE and CCEP estimators is greater for smaller n and smaller t.

Next consider the case of three endogenous regressors, $q = 3$. This is the case where Theorem 2 shows that the weighted asymptotic MSE of the combined estimator is uniformly smaller than that of the CCEP estimator. From Figs. 1, 2, 5, and 6, we see that the MSE of the combined estimator is uniformly smaller than that of the CCEP estimator for all factor loading values. For small ρ, the FE estimator has lower MSE than the combined estimator, but the ranking is reversed for moderate values of ρ. The FE estimator is very sensitive, which has quite low MSE for very small ρ, but very large MSE for large ρ. The combined and the pre-test estimators have much smaller MSE than CCEP for small values of ρ, but the ranking is reversed for large values of ρ. The MSE of the pre-test



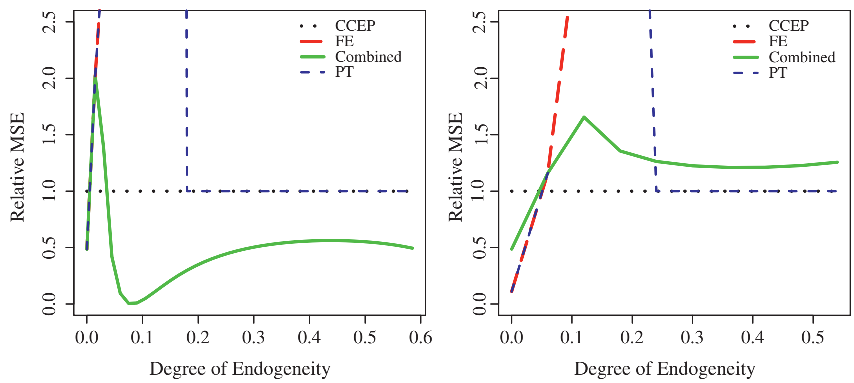
(a) $n = 50$, $T = 8$, $q = 3$, $r = 1$ (b) $n = 100$, $T = 8$, $q = 3$, $r = 1$

Fig. 1. Relative MSE of CCEP, FE, Pre-test, and Combined Estimators, $n = \{50, 100\}$, $T = 8$, $q = 3$, $r = 1$.



(a) $n = 50, T = 8, q = 3, r = 3$ (b) $n = 100, T = 8, q = 3, r = 3$

Fig. 2. Relative MSE of CCEP, FE, Pre-test, and Combined Estimators, $n = \{50, 100\}$, $T = 8$, $q = 3$, $r = 3$.

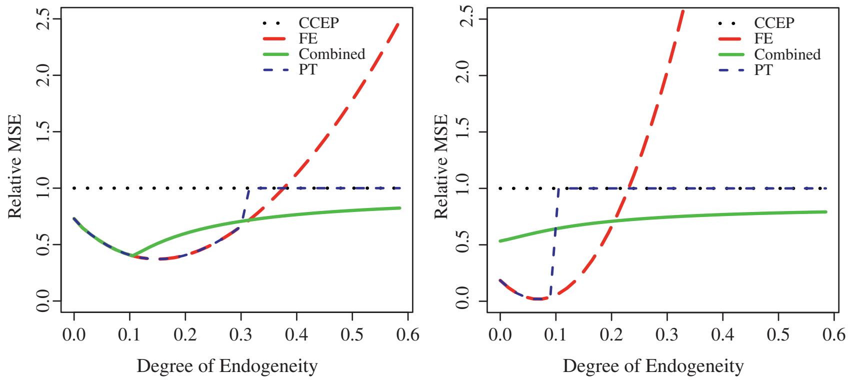


(a) $n = 50, T = 8, q = 1, r = 1$ (b) $n = 100, T = 8, q = 1, r = 1$

Fig. 3. Relative MSE of CCEP, FE, Pre-test, and Combined Estimators, $n = \{50, 100\}$, $T = 8$, $q = 1$, $r = 1$.

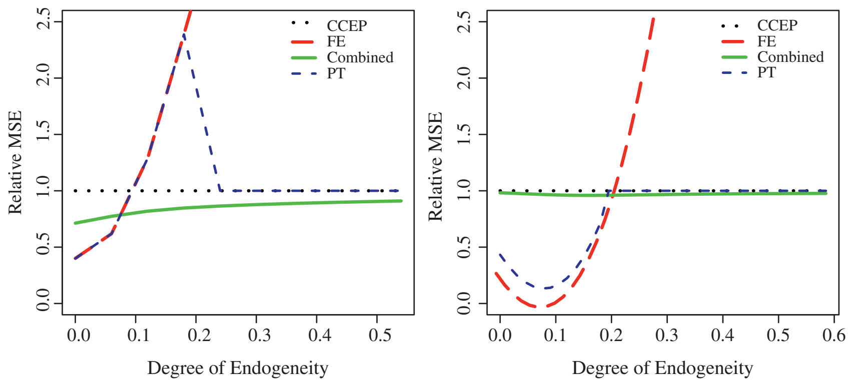
estimator is generally similar to the FE estimator for small ρ. For intermediate values of ρ, the MSE of the pre-test estimator is typically larger than the combined estimator. Following Pesaran (2006), γ_i and Γ_i are randomly generated as described earlier, while we have also tried with constant γ_i and Γ_i which give slightly better but essentially the same results (not reported).

Figs. 2 and 6 are the cases where $r = 3$. The general nature of the plot is the same, except that the gains is not as strong as in the case of $r = 1$. We see that the gains from the combined estimator are strong for small ρ, with the MSE converging to that of CCEP as ρ increases toward one. This is consistent with Theorem 2, which shows that the improvement are asymptotically local to



(a) $n = 50$, $T = 8$, $q = 2$, $r = 1$ (b) $n = 100$, $T = 8$, $q = 2$, $r = 1$

Fig. 4. Relative MSE of CCEP, FE, Pre-test, and Combined Estimators, $n = \{50, 100\}$, $T = 8$, $q = 2$, $r = 1$.

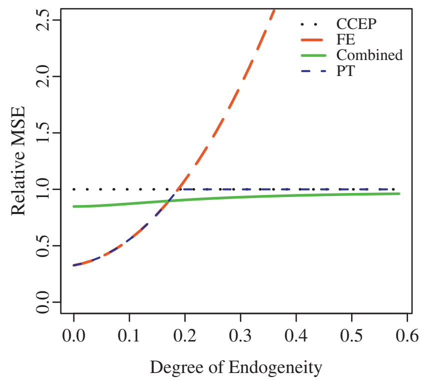


(a) $n = 50$, $T = 16$, $q = 3$, $r = 1$ (b) $n = 100$, $T = 16$, $q = 3$, $r = 1$

Fig. 5. Relative MSE of CCEP, FE, Pre-test and Combined Estimators, $n = \{50, 100\}$, $T = 16$, $q = 3$, $r = 1$.

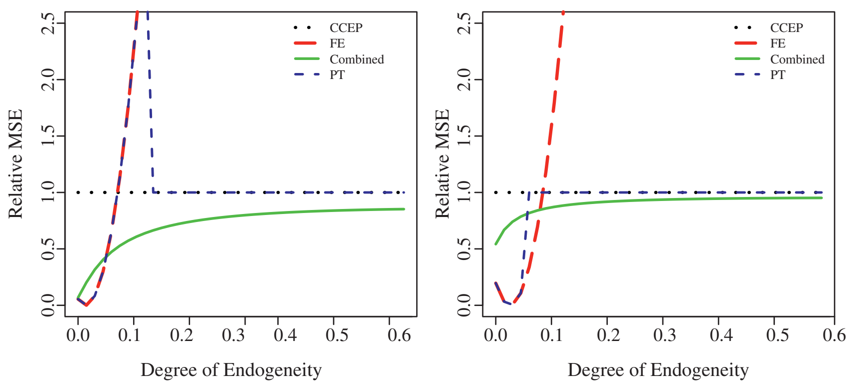
$\rho = 0$. The FE estimator has lower MSE than the combined estimator, but the MSE of the FE estimator increases dramatically after intermediate values of ρ. Still the combined estimator has uniformly smaller MSE than CCEP. In summary, the simulation results provide strong finite sample confirmation of Theorem 2 and its corollaries 1, 2, and 3.

Remark 1. It is interesting that the relative performance of using CCEP versus IFE depends on the true value of β is zero or not. We have experimented in simulation with $\beta \in \{0, 1\}$. A few figures to compare the results for $\beta \in \{0, 1\}$ are reported in Figs. 9 and 10. The figures with $\beta = 1$ were almost exactly the same as those with $\beta = 0$.



(a) $n = 50$, $T = 16$, $q = 3$, $r = 3$ (b) $n = 100$, $T = 16$, $q = 3$, $r = 3$

Fig. 6. Relative MSE of CCEP, FE, Pre-test, and Combined Estimators, $n = \{50, 100\}$, $T = 16$, $q = 3$, $r = 3$.



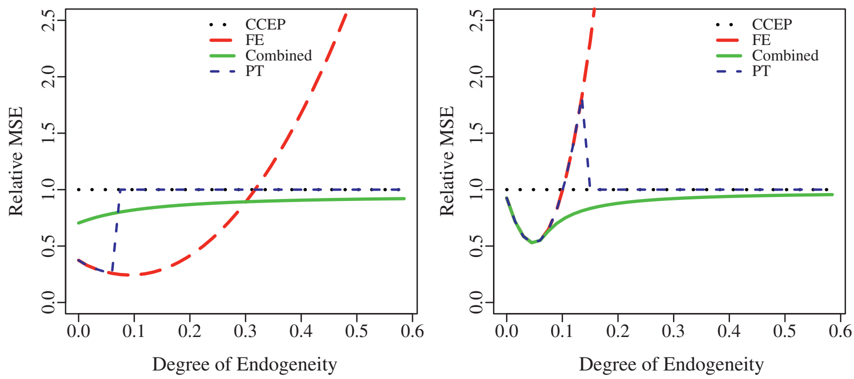
(a) $n = 50$, $T = 16$, $q = 1$, $r = 1$ (b) $n = 100$, $T = 16$, $q = 1$, $r = 1$

Fig. 7. Relative MSE of CCEP, FE, Pre-test and Combined Estimators, $n = \{50, 100\}$, $T = 16$, $q = 1$, $r = 1$.

Remark 2. We have also tried $2\tau^*$ as this choice still satisfies the classic James-Stein condition in Eq. (44). could make the MSE of the combined estimator somewhat closer to the MSE of FE when the degree of endogeneity is small. The results are reported in Figs. 11 and 12.

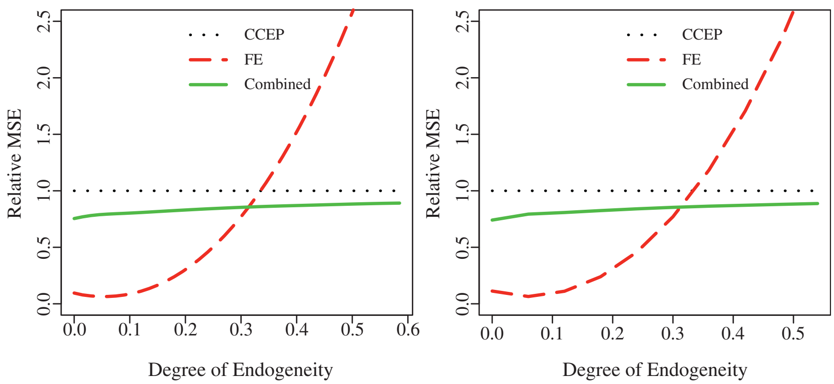
5. APPLICATION

Holly, Pesaran, and Yamagata (2010), hereafter HPY, provide an empirical analysis of changes in real house prices in United States. using state-level data. They use a panel of 49 states over the period of 1975–2003 to show that



(a) $n = 50, T = 16, q = 2, r = 1$ (b) $n = 100, T = 16, q = 2, r = 1$

Fig. 8. Relative MSE of CCEP, FE, Pre-test, and Combined Estimators, $n = \{50, 100\}$, $T = 16$, $q = 2$, $r = 1$.

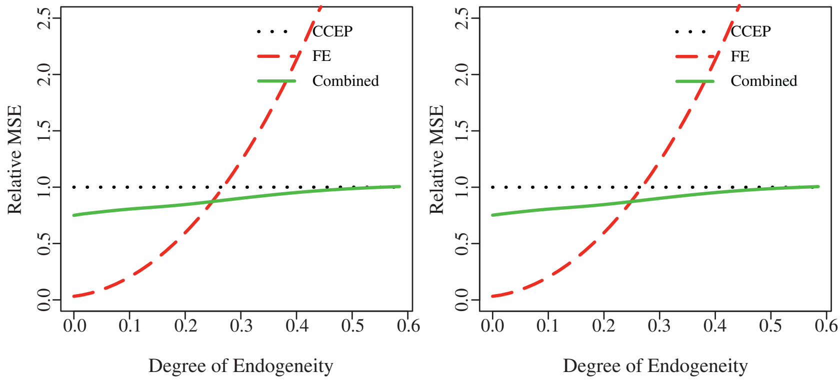


(a) $n = 100, T = 8, q = 3, r = 1, \beta = 0$ (b) $n = 100, T = 8, q = 3, r = 1, \beta = 1$

Fig. 9. Relative MSE of CCEP, FE and Combined Estimators, $n = 100$, $T = 8$, $q = 3$, $r = 1$, $\beta = \{0, 1\}$.

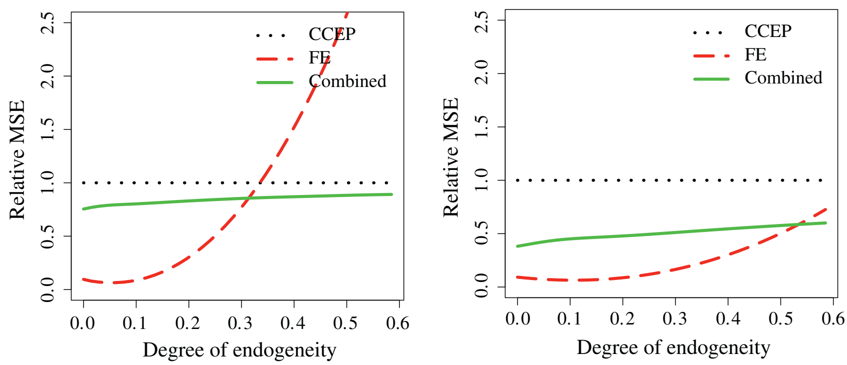
state-level real housing prices are driven by economic fundamentals, such as real per capita disposable income, as well as by common shocks, such as changes in interest rates, oil prices, and technological change. Baltagi and Li (2014) replicate their results using a panel of 381 metropolitan statistical areas observed over the period 1975–2011. Their replication shows that HPY results are fairly robust. Our empirical analysis relies upon a panel of 49 states over the period 1975–2011 to examine the performance of the combined estimator. Consider the following panel data model for US states

$$p_{i,t} = \beta_0 + \beta_1 y_{i,t} + \beta_2 g_{i,t-1} + \beta_3 c_{i,t-1} + \alpha_i + u_{it}, \tag{50}$$



(a) $n = 100, T = 8, q = 3, r = 1, \beta = 0$ (b) $n = 100, T = 8, q = 3, r = 1, \beta = 1$

Fig. 10. Relative MSE of CCEP, FE, and Combined Estimators, $n = 100$, $T = 8$, $q = 3$, $r = 3$, $\beta = \{0, 1\}$.



(a) $n = 100, T = 8, q = 3, r = 1, \tau$ (b) $n = 100, T = 8, q = 3, r = 1, 2\tau$

Fig. 11. Relative MSE of CCEP, FE, and Combined Estimators, $n = 100$, $T = 8$, $q = 3$, $r = 1$, $\{\tau, 2\tau\}$.

where $i = 1,\ldots, 49$, $t = 1,\ldots, 17$, $p_{i,t}$ is the logarithm of the real price of housing in the ith State during year t, and $y_{i,t}$ is the logarithm of the real per capita personal disposable income. The net cost of borrowing defined by $c_{i,t-1} = r_{it} - \Delta p_{it}$, where r_{it} represents the long-term real interest rate and $g_{i,t}$ represents the population growth rate. The state-specific effects can be treated as the endowment of climate, location, and culture. A more detailed description can be found in HPY. We would expect a rise in $c_{i,t}$ to be associated with a fall in the price income ratio, and hence a negative coefficient for $c_{i,t-1}$. The effect of population growth on real house prices is expected to be positive.

Table 1 suggests that the income elasticity of real house prices for the combined estimator is 1.2151, and the estimate of the coefficients on the rate of

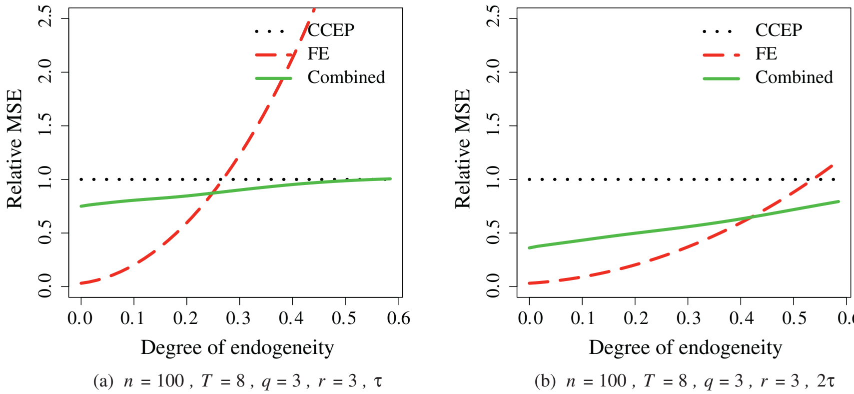


Fig. 12. Relative MSE of CCEP, FE, and Combined Estimators, $n = 100$, $T = 8$, $q = 3$, $r = 3$, $\{\tau, 2\tau\}$.

Table 1. Economics of Real House Prices: Correlated Common Effects.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
FE	0.5804	1.3286	−0.5088
	(0.3013)	(1.8132)	(0.1546)
CCEP	1.2705	1.6367	−0.1781
	(0.2990)	(1.5217)	(0.1541)
Combined	1.2151	1.6120	−0.2047
	(0.2986)	(1.4237)	(0.1530)

Notes: 49 US States (1975−2011). Reported are parameter estimates with the standard errors in parentheses.

change of population and the net cost of borrowing are 1.6120 and −0.2047, respectively, for the combined estimator. We find a significant positive effect for population growth and a significant negative effect associated with net cost of borrowing, which are in agreement with the results of HPY. The other two rows report the FE and CCEP estimates. The estimates of the combined estimator lie quite close to that of the CCEP estimator. We bootstrap the data 5000 times by resampling across individuals and keep the time series structure for each individual unchanged. The bootstrap MSE and the standard errors for the above estimates, then, can be calculated based on the estimates of the coefficients for each bootstrap data. The MSE for FE, CCEP, and combined estimators are 3.4250, 2.4288, and 2.1425, respectively. Among these three estimators, the combined estimator has the smallest MSE. The Hausman statistic is 24.9018. Thus, the exogeneity assumption is rejected at the one percent level of significance, which also indicates that the CCEP estimator is more reliable.

6. CONCLUSIONS

This chapter extends the study by Hansen (2017) for the combined (model averaging) estimation of the parametric panel data model with weak endogeneity (i.e., local to exogeneity) from common correlated effects. We introduce a combined estimation of the FE and CCEP estimators for the panel data models when the FE estimator suffers from inconsistency due to endogeneity arising from the correlated common effects. This can be viewed as the panel data model version of the shrinkage estimator combining the 2SLS estimator (CCEP) and the OLS estimator (FE) because the CCEP estimator is a control function estimator to remove the endogeneity from the correlated common effects.

The use of the combined estimation allows applied researchers to implement efficient estimation under the presence of weak endogeneity. The combined estimation would work even when there is no endogeneity or when there is strong endogeneity, without having to select a consistent estimator or an efficient estimator since the weights in the combined estimator will then be 1 or 0. Hence, the combined estimator is an omnibus estimator across all degrees of endogeneity, particularly useful when it is not clear which estimator to choose when endogeneity is weak.

ACKNOWLEDGMENT

We thank the participants at the AIE40 Conference at UCI, the co-editor, and an anonymous referee for many valuable inputs that have helped improving the chapter.

REFERENCES

Baltagi, B. H., & Li, J. (2014). Further evidence on the spatio-temporal model of house prices in the United States. *Journal of Applied Econometrics*, *29*(3), 515–522.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B*, *57*, 45–70.

Hansen, B. E. (2007). Least squares model averaging. *Econometrica*, *75*(4), 1175–1189.

Hansen, B. E. (2017). A Stein-like 2SLS estimator. *Econometric Reviews*, *36*, 840–852.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning, data mining, inference, and prediction*. (2nd ed.). New York: Springer.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*, 1251–1271.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian Model averaging: A tutorial. *Statistical Science*, *14*, 382–417.

Holly, S., Pesaran, M. H., & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, *158*(1), 160–173.

Jeffreys, H. (1961). *Theory of probability*. (3rd ed.). Oxford: Clarendon Press.

Judge, G. G., & Bock, M. E. (1978). *Statistical implications of pre-test and Stein rule estimators in econometrics*. Amsterdam: North Holland.

Koop, G., & Poirier, D. J. (1995). An empirical investigation of Wagner's hypothesis by using a model occurrence framework. *Journal of the Royal Statistical Society Series A*, *158*, 123–141.

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

Poirier, D. J. (1988). Causal relationships and replicability. *Journal of Econometrics*, *39*, 213–234.

Poirier, D. J. (1989). Errata. *Journal of Econometrics*, *42*, 381.

Poirier, D. J. (1995). *Intermediate statistics and econometrics: A comparative approach*. Cambridge, MA: MIT Press.

Poirier, D. J., & Klepper, S. (1981). Model occurrence and model selection in panel data sets. *Journal of Econometrics*, *17*, 333–350.

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for regression models. *Journal of the American Statistical Association*, *92*, 179–191.

Stein, C. (1956). *Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution*. 3rd Berkeley Symposium on Probability and Statistics (Vol. 1, pp. 197–206). Berkeley, CA: University of California Press.

APPENDIX

Proof of Theorem 1. Here, we derive only the asymptotic distribution of the FE estimator for the parametric panel data model with common correlated effects as specified in Eqs. (3) and (4). The asymptotic distribution of the CCEP estimator for this model is shown in the study by Pesaran (2006). For the joint asymptotic distribution of the FE and the CCEP estimators, the asymptoric covariance V_{12} is complicated and thus we will use bootstrap to estimate V_{12}.

To derive the asymptotic distribution of the FE estimator, we use the notation $h_1 = G_1 h$ and $\xi_1 = G_1 \xi$ with $G_1 = (I0)'$. Now, write the FE estimator as

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i' Q_T y_i\right),$$

$$\hat{\beta}_{FE} - \beta = \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i' Q_T u_i\right).$$

Given that $u_i = F\gamma_i + \varepsilon_i$,

$$\hat{\beta}_{FE} - \beta = \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} \left(\sum_{i=1}^{n} \left(X_i' Q_T F \gamma_i + X_i' Q_T \varepsilon_i\right)\right).$$

Since $\gamma_i = \frac{1}{\sqrt{n}} \tilde{\gamma}_i \delta$, we have

$$\hat{\beta}_{FE} - \beta = \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} \left(\sum_{i=1}^{n} \left(X_i' Q_T F \frac{\tilde{\gamma}_i \delta}{\sqrt{n}} + X_i' Q_T \varepsilon_i\right)\right)$$

$$\sqrt{n}\left(\hat{\beta}_{FE} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i' Q_T F \tilde{\gamma}_i \delta\right) + \left(\frac{1}{n}\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i' Q_T \varepsilon_i\right) \xrightarrow{d} h_1 + \xi_1,$$

where

$$h_1 = \left(\text{plim} \frac{1}{n}\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} \left(\text{plim} \frac{1}{n}\sum_{i=1}^{n} X_i' Q_T F \tilde{\gamma}_i \delta\right) = \sigma_\varepsilon^{-2} V_1 \Sigma \delta,$$

with $\sum \equiv \text{plim} \frac{1}{n} \sum_{i=1}^{n} X_i' Q_T F \tilde{\gamma}_i$, and

$$\xi_1 \sim \left(\text{plim}\frac{1}{n}\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1} Z,$$

with

$$Z = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i' Q_T \varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\left(\text{plim}\frac{1}{n}\sum_{i=1}^{n} X_i' Q_T X_i\right)\right).$$

Hence,

$$\xi_1 \sim N\left(0, \sigma_\varepsilon^2\left(\text{plim}\frac{1}{n}\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1}\right) = N(0, V_1),$$

and

$$\sqrt{n}\left(\hat{\beta}_{FE} - \beta\right) \overset{d}{\to} N(h_1, V_1),$$

where

$$V_1 = \sigma_\varepsilon^2\left(\text{plim}\frac{1}{n}\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1},$$

as defined in Assumption 8 with $\sigma_\varepsilon^2 = E\left(\varepsilon_{it}^2\right)$ is the variance of the idiosyncratic error in Eq. (3).

Proof of Theorem 2:

Noting that $\sqrt{n}\left(\hat{\beta}_{CCEP} - \beta\right) \to_d G_2'\xi \sim N(0, V_2), \sqrt{a^2 + b^2}$ then

$$R\left(\hat{\beta}_{CCEP}\right) = E\left(\xi' G_2' W G_2' \xi\right) = \text{tr}(W V_2).$$

Define Ψ^* as a random variable without positive part trimming

$$\Psi^* = G_2'\xi - \left(\frac{\tau}{(h+\xi)' B (h+\xi)}\right) G'(h+\xi).$$

Then using the fact that the pointwise quadric risk of Ψ is strictly smaller than that of Ψ^*

$$R\left(\hat{\beta}_c\right) = E(\Psi' W \Psi) < E\left(\Psi^{*'} W \Psi^*\right),$$

we can calculate that

$$E\left(\Psi^{*'}W\Psi^{*}\right) = R\left(\hat{\beta}_{CCEP}\right) + \tau^2 E\left(\frac{(h+\xi)'GWG'(h+\xi)}{\left((h+\xi)'B(h+\xi)\right)^2}\right) - 2\tau E\left(\frac{(h+\xi)'GWG_2'\xi}{(h+\xi)'B(h+\xi)}\right).$$

By Stein's Lemma: If $Z \sim N(0,V)$ is $q\times 1$, K is $q\times q$, and $\eta(x){:}R^q \to R^q$ is absolutely continuous, then

$$E\left(\eta(Z+h)'KZ\right) = E\mathrm{tr}\left(\frac{\partial}{\partial x}\eta(Z+h)'KV\right),$$

$\eta(x) = x/(x'Bx)$, and

$$\frac{\partial}{\partial x}\eta(x) = \frac{1}{x'Bx}I - \frac{2}{(x'Bx)^2}Bxx'.$$

Therefore,

$$\begin{aligned} E\left(\frac{(h+\xi)'GWG_2'\xi}{(h+\xi)'B(h+\xi)}\right) &= E\mathrm{tr}\left(\frac{GWG_2'V}{(h+\xi)'B(h+\xi)} - \frac{2GWG_2'V}{\left((h+\xi)'B(h+\xi)\right)^2}B(h+\xi)(h+\xi)'\right) \\ &= E\left(\frac{\mathrm{tr}\left(GWG_2'V\right)}{(h+\xi)'B(h+\xi)}\right) - 2E\mathrm{tr}\left(\frac{GWG_2'V}{\left((h+\xi)'B(h+\xi)\right)^2}B(h+\xi)(h+\xi)'\right). \end{aligned}$$

Since

$$GWG_2'V = WG_2'VG = W(V_2 - V_{21}),$$

and

$$\begin{aligned} GWG_2'VB &= GWG_2'VG\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-1}G' \\ &= GW(V_2 - V_{21})\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-1}G', \end{aligned}$$

set $W(V_2 - V_{21})\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-1} = C$, then

$$E\mathrm{tr}\left(\frac{GWG_2'V}{\left((h+\xi)'B(h+\xi)\right)^2}B(h+\xi)(h+\xi)'\right) = E\mathrm{tr}\left(\frac{(h+\xi)'GCG'(h+\xi)}{\left((h+\xi)'B(h+\xi)\right)^2}\right).$$

Thus,

$$E\left(\Psi^{*\prime}W\Psi^{*}\right) = R\left(\hat{\beta}_{CCEP}\right) + \tau^2 E\left(\frac{(h+\xi)'GWG'(h+\xi)}{((h+\xi)'B(h+\xi))^2}\right) + 4\tau E\mathrm{tr}\left(\frac{(h+\xi)'GCG'(h+\xi)}{\left((h+\xi)'B(h+\xi)\right)^2}\right) - 2\tau E\mathrm{tr}\left(\frac{(W(V_2-V_{21}))}{(h+\xi)'B(h+\xi)}\right). \tag{A.1}$$

Define $B_1 = \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-\frac{1}{2}}G'$ and $A = \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}}$ $C\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}}$

Note that $GWG_2'VB = GCG' = B_1'AB_1$, $B_1'B_1 = B$. Using the inequality $b'ab \leq (b'b)\lambda_{\max}(a)$ for symmetric a, and let

$$\lambda_{\max}(a) = \lambda_{\max}\left(\frac{A+A'}{2}\right) = \lambda_1.$$

Then,

$$\begin{aligned}\mathrm{tr}\left(B(h+\xi)(h+\xi)'GWG_2'V\right) &= \frac{(h+\xi)'B_1'(A+A')B_1(h+\xi)}{2} \\ &\leq (h+\xi)'B(h+\xi)\lambda_1.\end{aligned} \tag{A.2}$$

Define $A^* = \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}}W\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}}$. Note that $GWG' = B_1'A^*B_1$, $B_1' = B$, and let

$$\lambda_{\max}(a) = \lambda_{\max}\left(\frac{A^* + A^{*\prime}}{2}\right) = \lambda_1^*,$$

we have

$$\begin{aligned}\mathrm{tr}\left((h+\xi)'GWG'(h+\xi)\right) &= \frac{(h+\xi)'B_1'\left(A^* + A^{*\prime}\right)B_1(h+\xi)}{2} \\ &\leq (h+\xi)'B(h+\xi)\lambda_1^*.\end{aligned} \tag{A.3}$$

Plug Eqs. (A.2) and (A.3) into Eq. (A.1) and use Jensen's inequality, then we have

$$\begin{aligned} E\left(\Psi^{*'} W \Psi^{*}\right) & \leq R\left(\hat{\beta}_{CCEP}\right) + \tau^2 E\left(\frac{\lambda_1^*}{(h+\xi)' B (h+\xi)}\right) + 4\tau E\left(\frac{\lambda_1}{(h+\xi)' B (h+\xi)}\right) \\ & \quad -2\tau E \operatorname{tr}\left(\frac{(W(V_2 - V_{12}))}{(h+\xi)' B (h+\xi)}\right) \\ & = R\left(\hat{\beta}_{CCEP}\right) - E\left(\frac{\tau\left(2(\operatorname{tr} W(V_2 - V_{21}) - 2\lambda_1) - \lambda_1^* \tau\right)}{(h+\xi)' B (h+\xi)}\right) \\ & \leq R\left(\hat{\beta}_{CCEP}\right) - \frac{\tau\left(2(\operatorname{tr} W(V_2 - V_{21}) - 2\lambda_1) - \lambda_1^* \tau\right)}{E\left((h+\xi)' B (h+\xi)\right)}. \end{aligned} \tag{A.4}$$

Since $\operatorname{tr}(BV) = \operatorname{tr}\left(G\left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-1} G' V\right) = q$. We have

$$\begin{aligned} E\left((h+\xi)' B (h+\xi)\right) & = h' B h + \operatorname{tr}(BV) \\ & = \sigma_1^{-4} \delta' \Sigma' V_1 \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-1} V_1 \Sigma \delta + q. \end{aligned}$$

Substitute into Eq. (A.4), then we have

$$R\left(\hat{\beta}_c\right) \leq R\left(\hat{\beta}_{CCEP}\right) - \frac{\tau\left(2\lambda_1(d-2) - \lambda_1^* \tau\right)}{\sigma^{-4} \delta \Sigma' V_1 \left[V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right]^{-1} V_1 \Sigma \delta + q}.$$