Journal of Management Science and Engineering 4 (2019) 28–44

Contents lists available at ScienceDirect

Journal of Management Science and Engineering

journal homepage: www.keaipublishing.com/en/journals/journal-of-management-science-and-engineering/

A combined random effect and fixed effect forecast for panel data models

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

[a] School of Statistics and Mathematics, Central University of Finance & Economics, Beijing, China
[b] Department of Economics, University of California, Riverside, CA, Unites States

ARTICLE INFO

Article history:
Available online 25 March 2019

JEL classification:
C13
C33
C52

Keywords:
Endogeneity
Panel data
Fixed effect
Random effect
Hausman test
Combined estimator
Combined forecast

ABSTRACT

When some of the regressors in a panel data model are correlated with the random individual effects, the random effect (RE) estimator becomes inconsistent while the fixed effect (FE) estimator is consistent. Depending on the various degree of such correlation, we can combine the RE estimator and FE estimator to form a combined estimator which can be better than each of the FE and RE estimators. In this paper, we are interested in whether the combined estimator may be used to form a combined forecast to improve upon the RE forecast (forecast made using the RE estimator) and the FE forecast (forecast using the FE estimator) in out-of-sample forecasting. Our simulation experiment shows that the combined forecast does dominate the FE forecast for all degrees of endogeneity in terms of mean squared forecast errors (MSFE), demonstrating that the theoretical results of the risk dominance for the in-sample estimation carry over to the out-of-sample forecasting. It also shows that the combined forecast can reduce MSFE relative to the RE forecast for moderate to large degrees of endogeneity and for large degrees of heterogeneity in individual effects.

1. Introduction

This paper investigates the forecast combination in the panel data model. Despite the scarcity of panel data studies on the combined forecasts, there has been panel data research on forecast focusing on the pooling of information; see Stock and Watson (1999, 2002a,b) and Forni, Hallim, Lippi, and Reichlin (2000, 2005). Nevertheless, there is little research on pooling forecasts in the context of the forecast combination of Bates and Granger (1969). We consider a panel data regression model

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

where $i = 1, \ldots, n$ and $t = 1, \ldots, T$, β is $q \times 1$, x_{it} is the ith observation on q explanatory variables, α_i is the individual effect, and u_{it} is the random error. The individual effect terms can be modeled as either random or fixed effects.

When estimating a panel data model, we need to decide whether we should use fixed effects (FE) or random effects (RE) estimator. The FE and RE estimators and their combination are considered by Huang (2015) and Wang, Zhang, and Zhou

* Corresponding author. School of Statistics and Mathematics, Central University of Finance & Economics, Beijing, China.
E-mail addresses: huangbai@cufe.edu.cn (B. Huang), tae.lee@ucr.edu (T.-H. Lee), aman.ullah@ucr.edu (A. Ullah).

https://doi.org/10.1016/j.jmse.2019.03.004

(2016), who independently derive their asymptotic distributions using a local-to-exogeneity condition and calculate the asymptotic risk of the estimators based on Hansen (2017). If the individual effects are correlated with the other regressors in the model, the FE model is consistent and the RE model is inconsistent. The RE estimator becomes inconsistent since the regressors are correlated with the individual effects and thus become endogenous. On the other hand, if the individual effects are not correlated with the other regressors in the model, both RE and FE estimators are consistent and the RE estimator is more efficient. Therefore, there is a trade-off between inefficient FE estimation and biased RE estimation.

In this paper, we consider the combined forecast approach to the combined estimation results for a panel data model. We examine whether the FE and RE forecasts can be combined to produce a better forecast when the regressors (predictors) are endogenous, and specifically we wish to see if the forecast combining the FE and RE models can outperform the FE model forecast in terms of mean squared forecast error (MSFE).[1] Our simulation experiment shows that the combined forecast can uniformly dominate the FE forecast for all degrees of endogeneity, demonstrating that the in-sample estimation result carries over to the out-of-sample forecasting. It also shows that the combined forecast can reduce MSFE relative to the RE forecast for moderate to large degrees of endogeneity of the regressors and heterogeneity of the individual effects.

We illustrate this method with an application to forecasting electricity and natural-gas demands for 51 US states. Since electricity and gasoline demand has been studied extensively, strong priors exist as to the plausibility of price and income effects, providing a useful plausibility check to the results of the study. In this literature, Maddala, Trost, Li, and Joutz (1997) obtained short-run and long-run elasticities of energy demand for each of 49 US states over the period 1970–1990. They showed that heterogeneous time series estimates for each state yield inaccurate signs for the coefficients, while panel data estimates are not valid because the hypothesis of homogeneity of the coefficients was rejected. Baltagi, Bresson, and Pirotte (2002) compared the out-of-sample forecast performance of ten homogeneous and nine heterogeneous estimators including the shrinkage estimators applying them to the same data set. They showed that the homogeneous panel data estimates give the best out-of-sample forecasts. Our objective here is to compare the out-of-sample forecast performance of the FE forecast, RE forecast, and the proposed combined forecasting procedures, by applying them to the updated electricity and natural-gas panel data across 51 states (including Washington DC) over the period 1997–2012. We find that the combined forecast outperforms.

The rest of this paper is organized as follows. We begin with Section 2 where the combined estimation and its asymptotic results are presented. Sections 3 presents the combined forecasting approach. Section 4 gives Monte Carlo simulation. An empirical application is given in Section 5. Section 6 concludes.

2. Stein-like combined estimation for panel data models

First, we consider estimation using a panel data regression model with the random effects

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

where $i = 1, \ldots, n$, $t = 1, \ldots, T$, x_{it} is the q explanatory variables, β is a $q \times 1$ unknown parameter, α_i is the individual effect, and u_{it} is the regression error. The RE model assumes that $\alpha_i \sim$ i.i.d. $(0,\ \sigma_\alpha^2)$, $u_{it} \sim$ i.i.d. $(0,\ \sigma_u^2)$, and α_i are independent of the u_{it}. In addition, the regressors x_{it} are independent of the α_i and u_{it} for all i and t. Under these assumptions, we can write

$$y_{it} = x_{it}{}'\beta + v_{it}, \qquad E(v_{it}|x_i) = 0, \tag{3}$$

where $v_{it} = \alpha_i + u_{it}$. Write the model (3) in matrix form

$$y = X\beta + v, \tag{4}$$

where $y = (y_{11}, \ldots, y_{1T}, \ldots, y_{n1}, \ldots, y_{nT})'$ is $nT \times 1$, $X = (x_{11}, \ldots, x_{1T}, \ldots, x_{n1}, \ldots, x_{nT})'$ is $nT \times q$, $v = D\alpha + u$ with $D = I_n \otimes \iota_T$. Let ι_T be a vector of ones of dimension T, $J_T = \iota_T \iota_T{}'$, and $P = I_n \otimes \bar{J}_T$ where $\bar{J}_T = J_T / T$. Let $Q = I_{nT} - P$ be a matrix which obtains the deviations from individual means. The variance-covariance matrix of v is given by

$$\Omega = \sigma_\alpha^2 (I_n \otimes J_T) + \sigma_u^2 (I_n \otimes I_T) = \sigma_1^2 P + \sigma_u^2 Q, \tag{5}$$

where $\sigma_1^2 = T\sigma_\alpha^2 + \sigma_u^2$. The feasible estimator of $\widehat{\Omega}$ of Ω can be obtained by first running the OLS regression y on X to get $\widehat{v}_{it} = y_{it} - x_{it}\widehat{\beta}_{OLS}$ as the OLS residual and $\widehat{\beta}_{OLS} = (X'X)^{-1}X'y$. This gives

$$\widehat{\sigma}_u^2 = \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\widehat{v}_{it} - \overline{\widehat{v}}_i \right)^2. \tag{6}$$

[1] This paper considers a static model. If one wants to perform a better forecasting a dynamic panel data model with a lagged dependent variable would be more appropriate. However, for a dynamic panel data model, both the FE and RE estimators become inconsistent and thus we need other robust estimation methods such as IV and GMM estimators. We can certainly extend the current paper to the dynamic models by combining RE with GMM estimators for instance, but it would be beyond the scope of the current paper and thus we leave this for a future work.

Similarly, doing the OLS regression of $\bar{y}_i = \bar{x}_i\beta + \bar{v}_i$, where $V(\bar{v}_i) = T\sigma_\alpha^2 + \sigma_u^2/T = \sigma_1^2/T$ and $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, we get

$$\hat{\sigma}_1^2 = \frac{T}{n}\sum_{i=1}^{n}\hat{\bar{v}}_i^2. \tag{7}$$

Note that $\hat{\sigma}_\alpha^2 = \frac{1}{T}(\hat{\sigma}_1^2 - \hat{\sigma}_u^2)$. With these estimates, one can obtain the generalized least squares (GLS) of β based on (4) is

$$\hat{\beta}_{RE} = \left(X'\hat{\Omega}^{-1}X\right)^{-1}X'\hat{\Omega}^{-1}y, \tag{8}$$

and $\hat{\beta}_{RE}$ has the asymptotic distribution as

$$\sqrt{n}\left(\hat{\beta}_{RE} - \beta\right) \to N(0, V_{RE}), \tag{9}$$

where

$$V_{RE} = \left(\text{plim}\frac{X'\Omega^{-1}X}{n}\right)^{-1}. \tag{10}$$

Remark 1. Except for Nerlove's (1971) method, there is no guarantee that the estimate of $\hat{\sigma}_\alpha^2$ would be nonnegative regardless of the existence of the endogeneity problem. One solution suggested by Maddala and Mount (1973) is to replace these negative estimates by zero. They find that the negative estimates occurred only when the true σ_α^2 was small and close to zero, in which case OLS is still viable, and therefore the problem is dismissed as not being serious. In the parametric models, one may consider some other positive estimators of two unknown error variances in the random effect covariance matrix. In the nonparametric model, one of them is as considered in Henderson and Ullah (2005). A detailed study on comparing estimators under various estimates of variance parameters in the RE covariance matrix will be subject to a future study.

Second, we consider estimation using a panel data regression model with the fixed effects, for which the α_i are assumed to be fixed parameters to be estimated. Pre-multiplying the model by Q and performing the OLS on the resulting transformed model

$$Qy = QX\beta + Qu, \tag{11}$$

we obtain the OLS estimator

$$\hat{\beta}_{FE} = (X'QX)^{-1}X'Qy, \tag{12}$$

$$\hat{\alpha}_{FE} = (D'D)^{-1}D'\left(y - X\hat{\beta}_{FE}\right). \tag{13}$$

The asymptotic distribution of $\hat{\beta}_{FE}$ follows

$$\sqrt{n}\left(\hat{\beta}_{FE} - \beta\right) \xrightarrow{d} N(0, V_{FE}), \tag{14}$$

where

$$V_{FE} = \sigma_u^2\left(\text{plim}\frac{X'QX}{n}\right)^{-1}. \tag{15}$$

Under the random effects specification, $\hat{\beta}_{RE}$ is the asymptotically efficient estimator while $\hat{\beta}_{FE}$ is unbiased and consistent but not efficient. If $E(\alpha_i x_{it}) \neq 0$, $\hat{\beta}_{RE}$ is biased and inconsistent while $\hat{\beta}_{FE}$ is not affected.

Third, we consider the following combined estimator of β

$$\hat{\beta}_c = w\hat{\beta}_{RE} + (1 - w)\hat{\beta}_{FE}, \tag{16}$$

where

$$w = \begin{cases} \frac{\tau}{H_n} & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases}, \tag{17}$$

$$H_n = n\left(\widehat{\beta}_{FE} - \widehat{\beta}_{RE}\right)'\left[\widehat{V}_{FE} - \widehat{V}_{RE}\right]^{-1}\left(\widehat{\beta}_{FE} - \widehat{\beta}_{RE}\right), \tag{18}$$

and τ is a shrinkage parameter, H_n is the Hausman (1978) statistic. We set $\tau = q-2$ when $q>2$. The degree of shrinkage depends on the ratio τ/H_n. When $H_n<\tau$ then $\widehat{\beta}_c = \widehat{\beta}_{RE}$, When $H_n \geq \tau$ then $\widehat{\beta}_c$ is a weighted average of $\widehat{\beta}_{RE}$ and $\widehat{\beta}_{FE}$, with more weight on $\widehat{\beta}_{FE}$ when H_n is larger. The combined estimator can alternatively be written as a positive-part James-Stein estimator

$$\widehat{\beta}_c = \widehat{\beta}_{RE} + \left(1 - \frac{\tau}{H_n}\right)^+ \left(\widehat{\beta}_{FE} - \widehat{\beta}_{RE}\right) \tag{19}$$

where $(b)^+ = \max(b, 0)$.

Next, to examine the asymptotic properties of $\widehat{\beta}_c$, we use the local asymptotic approach based on the Mundlak's (1978) projection, where we write α_i as a linear function of $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$

$$\alpha_i = \overline{x}_i{}'\rho + \varepsilon_i, \tag{20}$$

with $E(\overline{x}_i\varepsilon_i) = 0$. The variable x_{it} are exogenous if α_i and x_{it} are uncorrelated (when the coefficient ρ is zero). For fixed T, ρ is local to zero

$$\rho = \frac{1}{\sqrt{n}}\delta, \tag{21}$$

where the $q \times 1$ parameter δ is a localizing parameter, which is the degree of correlation between x_{it} and α_i. If $\delta \neq 0$, then x_i are endogenous and the FE estimator is preferred. If $\delta = 0$, x_{it} are exogenous and the RE estimator is preferred. Note that $\overline{x}_i{}'$ is $1 \times q$ and is the ith row of the $n \times q$ matrix $\overline{X}$.

Remark 2. We focus on the Mundlak's (1978) projection, which projects the unobserved effect α_i onto the average of x_{it} all across all T time period. As a referee pointed out, Chamberlain's (1984) approach can be used instead. Chamberlain's method is a generalization of Mundlak's model, but rather to replace α_i with the linear projection of it onto the explanatory variables in all time periods. Specifically, Chamberlain's method leads to the following equation

$$\alpha_i = c + x_{i1}\rho_1 + x_{i2}\rho_2 + \ldots + x_{iT}\rho_T + \varepsilon_i$$

We leave this for a future work.

Now, we make the following assumptions:

Assumption 1. $\{x_{it},\ \alpha_i,\ u_{it}\}$ are i.i.d. over i, u_{it} is i.i.d. over t, $E(u_{it}|x_{it},\alpha_i) = 0$, $E(u_{it}^4|x_{it},\alpha_i) < \infty$..

Assumption 2. $E||x_{it}||^{2+k} < \infty$ and $E\left|u_{it}\right|^{2+k} < \infty$ for some $k>0$.

Assumption 3. $\widehat{\sigma}_u^2 = \sigma_u^2 + o_p(1)$ and $\widehat{\sigma}_\alpha^2 = \sigma_\alpha^2 + o_p(1)$.

Assumptions 1 and 2 specify that the variables have finite fourth moments (so that the conventional central limit theory applies) and that the error is conditionally homoskedastic given the regressors, which is used to simplify the asymptotic covariance expressions. We have the following asymptotic results, extending Hansen (2017) for the panel data models.

Theorem 1. Under Assumptions 1-3,

$$\sqrt{n}\begin{pmatrix} \widehat{\beta}_{RE} - \beta \\ \widehat{\beta}_{FE} - \beta \end{pmatrix} \xrightarrow{d} h + \xi, \tag{22}$$

where

$$h = \begin{pmatrix} \sigma_1^{-2} V_{RE}\Sigma\delta \\ 0 \end{pmatrix}, \qquad \text{with } \Sigma = \text{plim}\frac{\overline{X}'\overline{X}}{n}, \tag{23}$$

and

$$\xi \sim N(0, V), \qquad \text{with } V = \begin{pmatrix} V_{RE} & V_{RE} \\ V_{RE} & V_{FE} \end{pmatrix}. \tag{24}$$

Furthermore,

$$H_n \xrightarrow{d} (h+\xi)' B(h+\xi), \tag{25}$$

and

$$\sqrt{n}\left(\widehat{\beta}_c - \beta\right) \xrightarrow{d} \Psi \equiv G_2{}'\xi - \left(\frac{\tau}{(h+\xi)' B(h+\xi)}\right)_1 G'(h+\xi), \tag{26}$$

Where $\overline{X}$ is $n \times q$ with $\overline{x}_i{}'$ in its ith row, $B = G(V_{FE} - V_{RE})^{-1}G'$, $G = (-I \quad I)'$, $G_2 = (0 \quad I)'$, and $(a)_1 = \min[1, a]$.

Proof 1: See Appendix.

Theorem 1 presents the joint asymptotic distribution of $\widehat{\beta}_{RE}$ and $\widehat{\beta}_{FE}$, the Hausman statistic, and $\widehat{\beta}_c$ under the local endogeneity setup in (21). The joint asymptotic distribution of $\widehat{\beta}_{RE}$ and $\widehat{\beta}_{FE}$ is normal. $\widehat{\beta}_{RE}$ has an asymptotic bias when $\delta \neq 0$ but $\widehat{\beta}_{FE}$ is consistent. The Hausman statistic has an asymptotic non-central chi-square distribution, with non-centrality parameter h depending on the local endogeneity parameter δ. The asymptotic distribution of $\widehat{\beta}_c$ is a nonlinear function of the normal random vector ξ and a function of the noncentrality parameter h.

Finally, we compare $\widehat{\beta}_{RE}, \widehat{\beta}_{FE}, \widehat{\beta}_c$ in the asymptotic risk. The asymptotic risk of any sequence of estimators β_n of β is defined as

$$R(\beta_n, \beta, W) = \lim_{n\to\infty} E\left[n(\beta_n - \beta)' W(\beta_n - \beta)\right]. \tag{27}$$

Denote $R(\beta_n) = R(\beta_n, \beta, W)$ for notational brevity. So long as the estimator has an asymptotic distribution

$$\sqrt{n}(\beta_n - \beta) \xrightarrow{d} \psi,$$

for some random variable ψ, the asymptotic risk can be calculated using

$$R(\beta_n) = E(\psi' W \psi) = \operatorname{tr}(WE(\psi\psi')). \tag{28}$$

For example,

$$R\left(\widehat{\beta}_{FE}\right) = \operatorname{tr}(WV_{FE}) \tag{29}$$

from (14).

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q$ denote the ordered eigenvalues of $W(V_{FE} - V_{RE})$. Denote the ratio

$$d = \frac{\operatorname{tr}(W(V_{FE} - V_{RE}))}{\lambda_1}. \tag{30}$$

Notice that (30) satisfies $1 \leq d \leq q$. In the case $W = (V_{FE} - V_{RE})^{-1}$, $\lambda_1 = 1$ and we have the simplification $d = q$.

Theorem 2. Under Assumption 1-3, if$d > 2$and

$$0 < \tau \leq 2(d-2), \tag{31}$$

then

$$R\left(\widehat{\beta}_c\right) < R\left(\widehat{\beta}_{FE}\right) - \frac{\tau\lambda_1[2(d-2) - \tau]}{\sigma_1^{-4}\delta'\Sigma V_{RE}(V_{FE} - V_{RE})^{-1} V_{RE}\Sigma\delta + q}. \tag{32}$$

□

Proof 3: See Appendix.

Remark 3. Equation (32) shows that the asymptotic risk of $\widehat{\beta}_c$ is strictly less than that of $\widehat{\beta}_{FE}$, so long as τ satisfies the condition (31). The assumption $d > 2$ is the critical condition needed to ensure that $\widehat{\beta}_c$ can have smaller asymptotic risk than that of $\widehat{\beta}_{FE}$. It is necessary in order for the right-hand-side of (31) to be positive, which is necessary for the existence of τ. τ appears in the risk bound (32) as a quadratic expression, so there is an optimal choice

$$\tau^* = d - 2, \tag{33}$$

which minimizes this bound. Note that $d>2$ is equivalent to $\lambda_2+\cdots+\lambda_q>\lambda_1$. This is violated only if λ_1 is much larger than the other eigenvalues. (31) is equivalent to $0<\tau\leq 2\left(\sum_{i=1}^{q}\frac{\lambda_i}{\lambda_1}-2\right)$ For the practical implementation we replace the maximum eigenvalue λ_1 with the average $\frac{\mathrm{tr}(W(V_{FE}-V_{RE}))}{q}=\frac{1}{q}\sum_1^q\lambda_i$. See Hansen (2016).

Corollary 1. $R(\widehat{\beta}_c)-R(\widehat{\beta}_{FE})<0$, for $d>2$ and $0<\tau\leq 2(d-2)$. When $W=(V_{FE}-V_{RE})^{-1}$, the condition simplifies to $q>2$ and $0<\tau\leq 2(q-2)$, which is Stein's (1956) classic condition for shrinkage.□

The following two corollaries are obtained with $W=(V_{FE}-V_{RE})^{-1}$.

Corollary 2. *$R(\widehat{\beta}_{RE})=tr(WV_{RE})+\sigma_1^{-4}\delta'\Sigma V_{RE}WV_{RE}\Sigma\delta$; tr $R(\widehat{\beta}_{RE})\leq R(\widehat{\beta}_{FE})$ when $\sigma_1^{-4}\delta'\Sigma V_{RE}WV_{RE}\Sigma\delta\leq q$, and $R(\widehat{\beta}_{RE})>R(\widehat{\beta}_{FE})$ otherwise.*□

Proof 4: See Appendix.

Corollary 3. *$R(\widehat{\beta}_c)-R(\widehat{\beta}_{RE})<0$, for $q<\sigma_1^{-4}\delta'\Sigma V_{RE}WV_{RE}\Sigma\delta$, $d>2$, and $0<\tau\leq 2(d-2)$.*□

Remark 4. Corollary 1 shows, as in Stein's (1956), that $q>2$ is necessary in order for the Stein estimator to achieve global reductions in risk relative to the usual estimator. $d>2$ is the generalization to allow for general weight matrices. Corollary 2 indicates that when endogeneity is weak (ρ and hence δ is close to zero) the random effects estimator may perform better than the fixed effects estimator. Corollary 3 indicates that when endogeneity is strong, $d>2$, and $0<\tau\leq 2(d-2)$, the combined estimator is better than the RE estimator.□

3. Stein-like combined forecast for panel data models

First, we consider forecasting using a panel data regression model with the random effects. Suppose we want to predict s periods ahead for the ith individual. By minimizing

$$\frac{\sum_i\sum_t(y_{it}-x_{it}'\beta-\alpha_i)^2}{\sigma_u^2}+\frac{\sum_i\alpha_i^2}{\sigma_\alpha^2}, \tag{34}$$

we can obtain

$$\widehat{\alpha}_i=\frac{\widehat{\sigma}_\alpha^2}{\widehat{\sigma}_1^2}T\overline{\widehat{v}}_{i(RE)}, \tag{35}$$

where $\overline{\widehat{v}}_{i(RE)}=\frac{1}{T}\sum_t\widehat{v}_{it(RE)}$. Then the s period ahead forecast for the ith individual is

$$\widehat{y}_{i,T+s,RE}=x_{i,T+s}'\widehat{\beta}_{RE}+\frac{\widehat{\sigma}_\alpha^2}{\widehat{\sigma}_1^2}T\overline{\widehat{v}}_{i(RE)}, \tag{36}$$

where $\frac{\widehat{\sigma}_\alpha^2}{\widehat{\sigma}_1^2}\sum_t\widehat{v}_{it(RE)}$ can be treated as $\widehat{\alpha}_{i,RE}$. Baltagi (2008) showed that Goldberger (1962) gave the best linear unbiased predictor (BLUP) of $y_{i,T+s}$ as following

$$\widehat{y}_{i,T+s,RE}=x_{i,T+s}'\widehat{\beta}_{RE}+\varpi'\Omega^{-1}\widehat{v}_{RE}, \tag{37}$$

where $\widehat{v}_{RE}=y-X\widehat{\beta}_{RE}$ and $\varpi=E(v_{i,T+1}v)$. Note that for period $T+s$

$$v_{i,T+s}=\alpha_i+u_{i,T+s}, \tag{38}$$

and $\varpi=\sigma_\alpha^2(l_i\otimes\iota_T)$ where l_i is ith column of I_N, i.e. l_i is a vector that has 1 in the ith position and zero elsewhere. In this case

$$\varpi'\Omega^{-1}=\sigma_\alpha^2(l_i'\otimes\iota_T')\left[\frac{1}{\sigma_1^2}P+\frac{1}{\sigma_u^2}Q\right]=\frac{\sigma_\alpha^2}{\sigma_1^2}l_i'\otimes\iota_T', \tag{39}$$

since $(l_i'\otimes\iota_T')P=(l_i'\otimes\iota_T')$ and $(l_i'\otimes\iota_T')Q=0$. The typical element of $\varpi'\Omega^{-1}\widehat{v}_{RE}$ becomes $\left(\frac{T\widehat{\sigma}_\alpha^2}{\widehat{\sigma}_1^2}\overline{\widehat{v}}_{i(RE)}\right)$ Therefore, this BLUP for $\widehat{y}_{i,T+s}$ corrects the RE prediction by a fraction of the mean of the RE residuals corresponding to the ith individual.

Next, consider forecasting using a panel data regression model with the fixed effects, from (13), we know that for the ith individual, $\widehat{\alpha}_{i,FE}=\overline{y}_i-\overline{x}_i\widehat{\beta}_{FE}$. Thus, the s period ahead forecast for the ith individual is

$$\widehat{y}_{i,T+s,FE} = \overline{y}_i + \left(x_{i,T+s} - \overline{x}_i\right)' \widehat{\beta}_{FE}, \tag{40}$$

or alternatively,

$$\widehat{y}_{i,T+s,FE} = x_{i,T+s}{}' \, \widehat{\beta}_{FE} + \overline{\widehat{u}}_{i(FE)}. \tag{41}$$

Now, we consider the combined forecast. In Section 2, we have shown that the combined estimator is better than the FE estimator in asymptotic risk (Theorem 2), and also better than the RE estimator when the endogeneity is strong (Corollary 3). To see if this dominance in in-sample estimation holds true in out-of-sample forecasting, we combine $\widehat{y}_{i,T+s,RE}$ and $\widehat{y}_{i,T+s,FE}$ using the weight in (17), then

$$\widehat{y}_{i,T+s,c} = x_{i,T+s}{}'\widehat{\beta}_c + w\varpi'\Omega^{-1}\widehat{v}_{RE} + (1-w)\overline{\widehat{u}}_{i(FE)} = w\,\widehat{y}_{i,T+s,RE} + (1-w)\widehat{y}_{i,T+s,FE}, \tag{42}$$

In the following two sections, we conduct the comparison of the combined forecast $\widehat{y}_{i,T+s,c}$ (the forecast using $\widehat{\beta}_c$) with the RE forecast $\widehat{y}_{i,T+s,RE}$ (the forecast using $\widehat{\beta}_{RE}$) and the FE forecast $\widehat{y}_{i,T+s,FE}$ (the forecast using $\widehat{\beta}_{FE}$) based on Monte Carlo and an application.

4. Monte Carlo

We consider the following data generating process

$$y_{it} = x_{it}{}'\beta + \alpha_i + u_{it}, \tag{43}$$

$$\alpha_i = \rho\sqrt{T}\overline{x}_i{}'\,\iota\,/q + \sqrt{1-\rho^2}\,\varepsilon_i, \tag{44}$$

where x_{it} is $q \times 1$ i.i.d. $N(0,\ I_q)$, ι is a $q \times 1$ vector of ones, u_{it} are i.i.d. $N(0,\ \sigma_u^2)$ across i,t. ε_i are i.i.d. $N(0,\ 1)$ independent of $\{x_{it},\ u_{it}\}$. $\sigma_u \in \{.6,\ .8,\ 1\}$, $\mathrm{Var}(\alpha_i) = 1$. We set $\sqrt{\theta} = \frac{\sigma_\alpha}{\sigma_u} \in \left\{\frac{5}{3},\ \frac{5}{4},\ 1\right\}$, so that we have $\rho^* = \frac{\theta}{1+\theta} = \{74,\ .61,\ .50\}$ ρ^* controls the degree of heterogeneity which is the temporal correlation between $\alpha_i + u_{it}$ and $\alpha_i + u_{it'}$. Then α_i and x_{it} have correlation $\frac{\rho}{\sqrt{q}}$, which controls the degree of endogeneity. We allow ρ to vary in $(-1,\ 1)$. The distribution are invariant to β so we set it to zero, $\beta = 0$.

First we generated 300,000 samples on each calculated $\widehat{\beta}_{RE},\ \widehat{\beta}_{FE},\ \widehat{\beta}_c$. We also calculated the Hausman pre-test (PT) estimator

$$\widehat{\beta}_{PT} = \widehat{\beta}_{RE} 1\left(H_n < \chi^2_{q,0.05}\right) + \widehat{\beta}_{FE} 1\left(H_n \geq \chi^2_{q,0.05}\right) \tag{45}$$

where $\chi^2_{q,0.05}$ is the 5% critical value from the χ^2_q distribution.

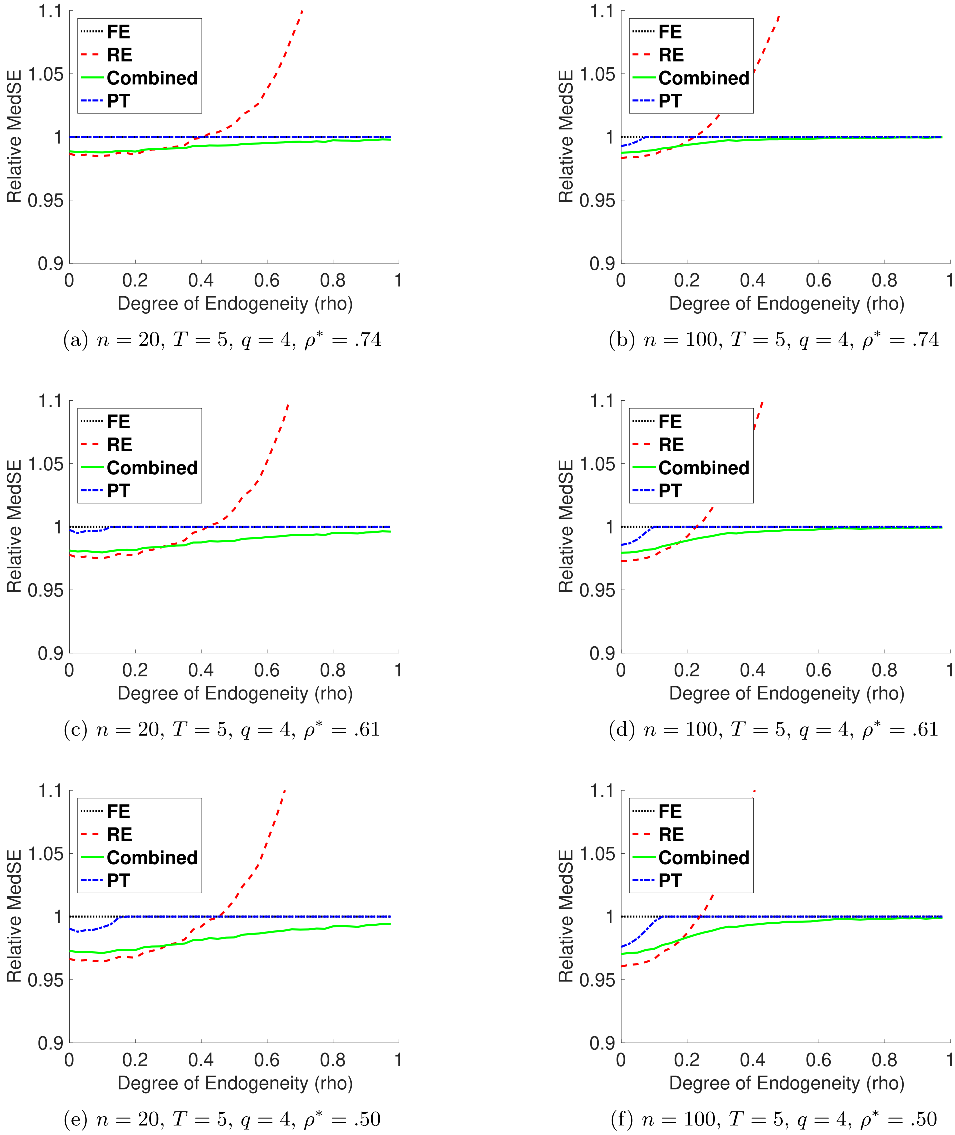
To compare the in sample fit of these estimators, calculate the median squared error (MedSE) of each estimator and plot the relative MedSE relative to that of the robust estimator FE under endogeneity, that is

$$\frac{\text{median}\left[\left(\widehat{\beta}-\beta\right)'\left(\widehat{\beta}-\beta\right)\right]}{\text{median}\left[\left(\widehat{\beta}_{FE}-\beta\right)'\left(\widehat{\beta}_{FE}-\beta\right)\right]}. \tag{46}$$

Thus a value less than one indicates improved precision relative to the FE estimator, and a value greater than one indicates worse performance than the FE estimator, with larger MedSE than FE estimator. The MedSE is symmetric with respect to ρ, so we only report the results with ρ between 0 and 1.

We use a portion of the available data for forecasting and use the other portion of the data for estimating the model as follows: Use the observations over $t = 1, \ldots, T-s$ to estimate the forecasting models. Compute the s-step error on the forecast for time T. Compute the forecast accuracy measures based on the forecast errors obtained. To compare the prediction procedures, we calculate the s-step ahead out-of-sample mean squared forecast error (MSFE) of each approach. The forecast error is defined as $e_{T+s} = y_{T+s} - \widehat{y}_{T+s}$, and its MSFE is $MSFE(e_{T+s}) = E(e_{T+s}{}'e_{T+s})$. Generate 10,000 samples on the FE forecast $\widehat{y}_{T+s,FE}$, the RE forecast $\widehat{y}_{T+s,RE}$, the Hausman pre-test forecast $\widehat{y}_{T+s,PT}$, and the combined forecast $\widehat{y}_{T+s,c}$. Plot the MSFEs of the RE, FE, combined forecasts relative to the MSFE of the FE forecast. Thus values less than one indicate improved precision relative to the FE forecast, and values greater than one indicate worse performance than the FE forecast. We set $T = 5$, $s = 1$, $q = 4$, and $n \in \{20, 100\}$.

Figs. 1 and 3 present the in-sample estimation results. In Fig. 1: (a), (c) and (e) plot the relative MedSE for $n = 20$; (b), (d) and (f) plot the relative MedSE for $n = 100$. We see that the region of dominance for the combined estimator over RE



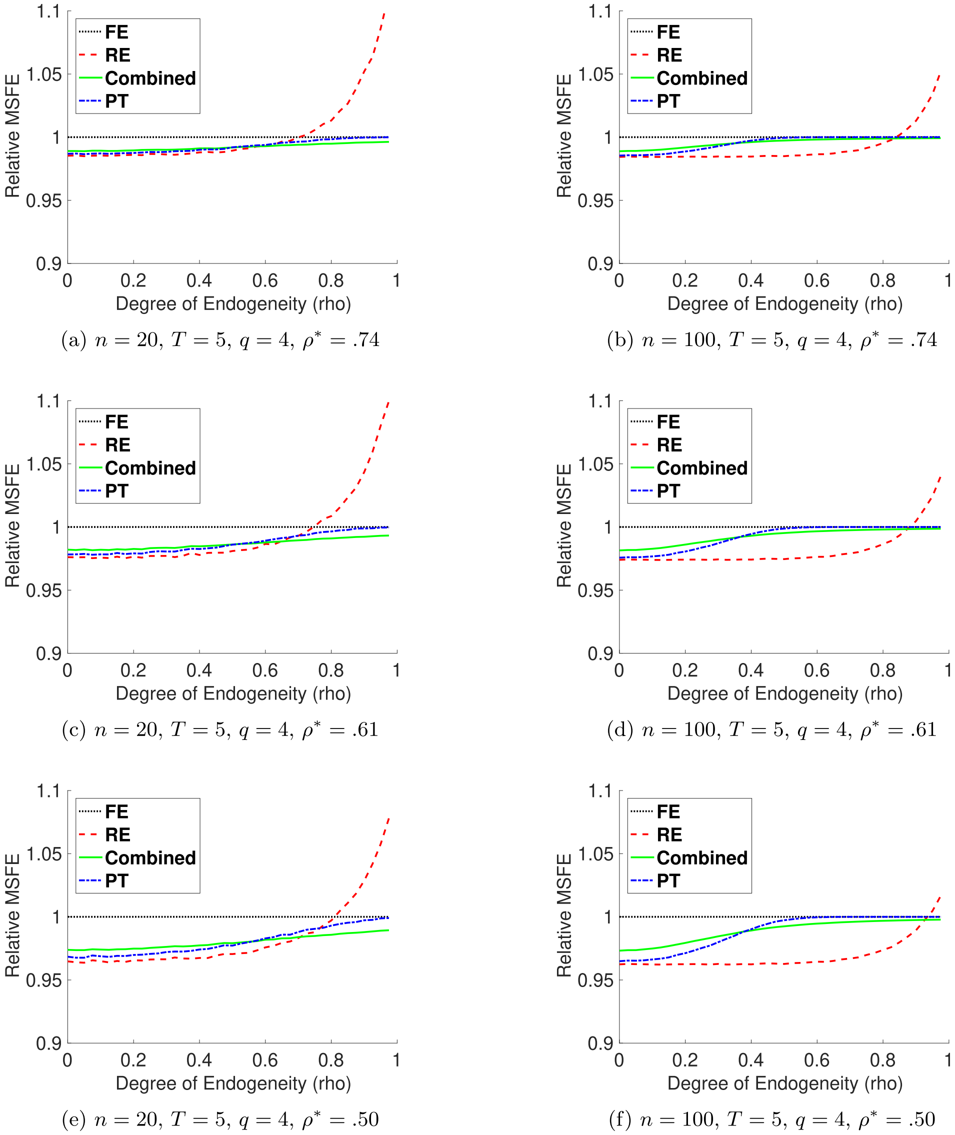
(a) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .74$ (b) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .74$

(c) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .61$ (d) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .61$

(e) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .50$ (f) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .50$

Fig. 1. Relative MedSE of FE, RE, combined and pretest estimators, $n = \{20, 100\}, T = 5, q = 4, \rho^* = \{.74, .61, .50\}$.

estimator is greater for small n. Fig. 1 plots the relative MedSE for $\rho^* \in \{74,\ .61,\ .50\}$ and $\tau = \tau^*$. This is the case of moderate to large degree of heterogeneity. Fig. 1(e) and (f) are the cases, $\rho^* = .50$. We see that the gains from the combined estimator are strong for small ρ, with the MedSE converging to that of FE as ρ increases towards 1. This is consistent with Theorem 2, which shows that the improvements are asymptotically local to $\rho = 0$. The RE estimator has lower MedSE than the combined estimator, but the ranking is reversed for larger values of ρ. Fig. 1(a), (b), 1(c), and 1(d) are the cases, $\rho^* \in \{.74,\ .61\}$. The general nature of the plot is the same, except that the gains are not as strong as in the case $\rho^* = .50$. Still the combined estimator has uniformly smaller MedSE than that of FE. The FE and the combined estimators are getting closer as ρ^* increases. RE has similar MedSE to FE and combined estimators for small ρ, but the MedSE of the RE estimator increases dramatically after intermediate values of ρ. It is also instructive to examine the performance of the pre-test estimator. The MedSE of the pre-test estimator is generally similar to FE for moderate and higher values of ρ. In summary, for moderate ρ^* and higher ρ, or moderate ρ and higher ρ^*, the combined estimator is better than RE estimator. For very large ρ^* and very low ρ, the combined estimator is close to RE estimator. The estimation simulation results provide strong finite sample confirmation of Theorem 2. Fig. 2 plots the relative MedSE for $\tau = 2\tau$, which still satisfies the classic James-Stein conditions in equations (30) and (31). The region of dominance for the RE and combined estimators over FE is greater for the large value of τ. The MSE of the combined estimator is closer to that of FE when the degree of endogeneity is small for large τ. This can be seen by contrasting Figs. 1 and 2. Hence, this indicates that the optimal choice of $\tau = \tau^*$ in (33) obtained from minimizing the “bound” of the risk of $R(\widehat{\beta}_c)$ is



(a) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .74$ (b) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .74$

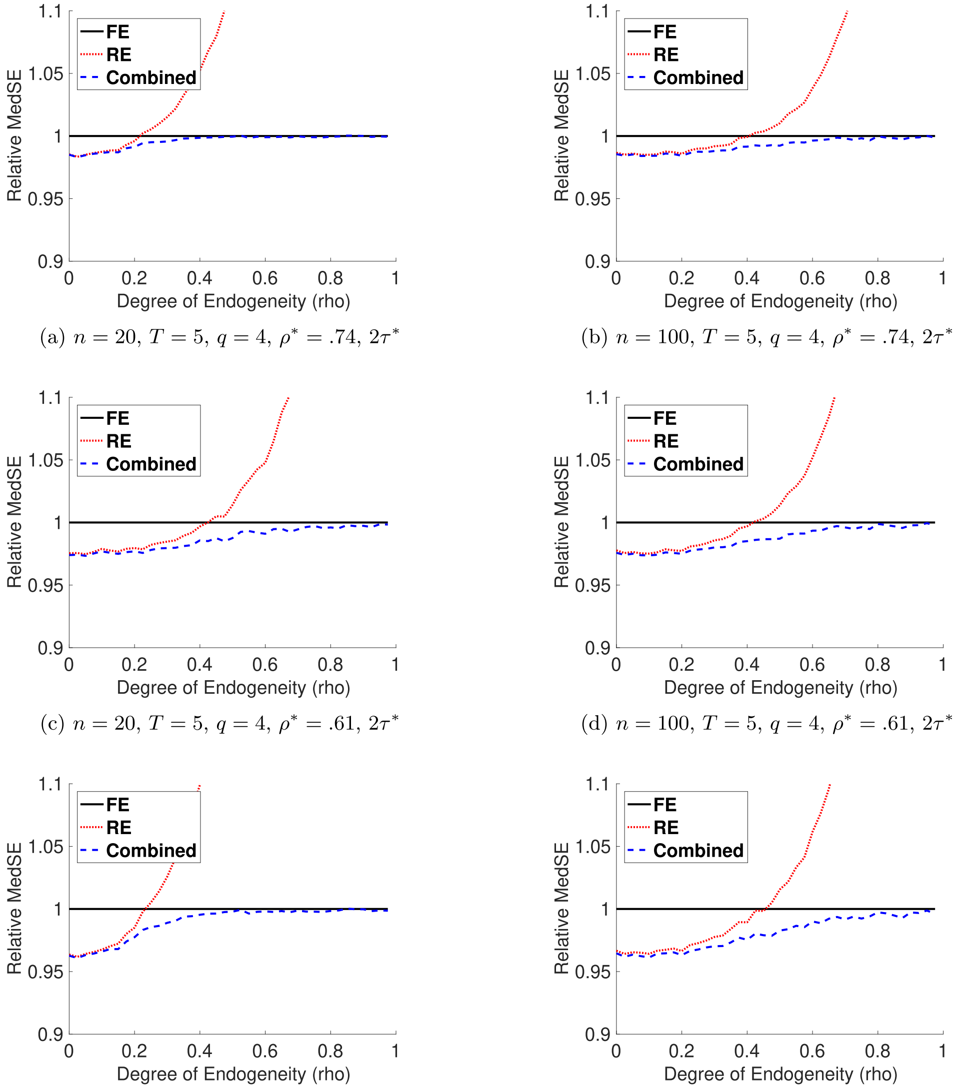
(c) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .61$ (d) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .61$

(e) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .50$ (f) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .50$

Fig. 2. Relative MSFE of the FE, RE combined and pretest estimators, one-step forecast, $s = 1, n = \{20, 100\}, T = 5, q = 4, \rho^* = \{.74, .61, .50\}$.

not “optimal” in the sense of minimizing the risk of $R(\widehat{\beta}_c)$ itself. To our knowledge, there is no result on this yet and thus we leave this for a future work.

Figs. 2 and 4 present the out-of-sample forecasting results. Fig. 2 shows the relative 1-step ahead out-of-sample MSFE of each approach with $\rho^* \in \{74,\ .61,\ .50\}$. By contrasting Fig. 2 with Fig. 1, we see that the general nature of the plots is the same. In Fig. 2: (a), (c) and (e) plot the MSFE for $n = 20$; (b), (d) and (f) plot the MSFE for $n = 100$. We see again that the region of dominance for the combined forecast over the FE forecast is greater for small n. Fig. 2(e) and (f) are the cases for $\rho^* = .50$. This is similar to Fig. 1(e) and (f) that the combined forecast has much lower MSFE than the FE forecast, regardless of the degree of endogeneity. For small ρ, the RE forecast has lower MSFE than the combined forecast, but the ranking is reversed for larger values of ρ. Figs. 2(a), (b), (c), and (d) are the cases for larger $\rho^* = \{.74,\ .61\}$, for which the combined forecast has lower MSFE than the FE forecast for small ρ but the reverse holds for large ρ. The combined and the pre-test forecasts have much smaller MSFE than FE for small values of ρ, but the ranking is reversed for large values of ρ. For large values of ρ, the MSFE of the pre-test forecast is typically larger than the combined forecast. In all the cases, the combined forecast uniformly dominates the FE forecast (which is the same as for the in-sample estimation results in Theorem 2 and in Fig. 1), demonstrating that the in-sample estimation results (Theorem 2) can hold true for the out-of-sample forecasting. Fig. 3 shows the relative 1-step ahead out-of-sample MSFE of each approach with $\tau = \tau^*$. Fig. 4 shows the relative 1-step ahead out-of-sample MSFE of



(a) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .74,\ 2\tau^*$ (b) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .74,\ 2\tau^*$

(c) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .61,\ 2\tau^*$ (d) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .61,\ 2\tau^*$

e) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .50,\ 2\tau^*$ (f) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .50,\ 2\tau^*$

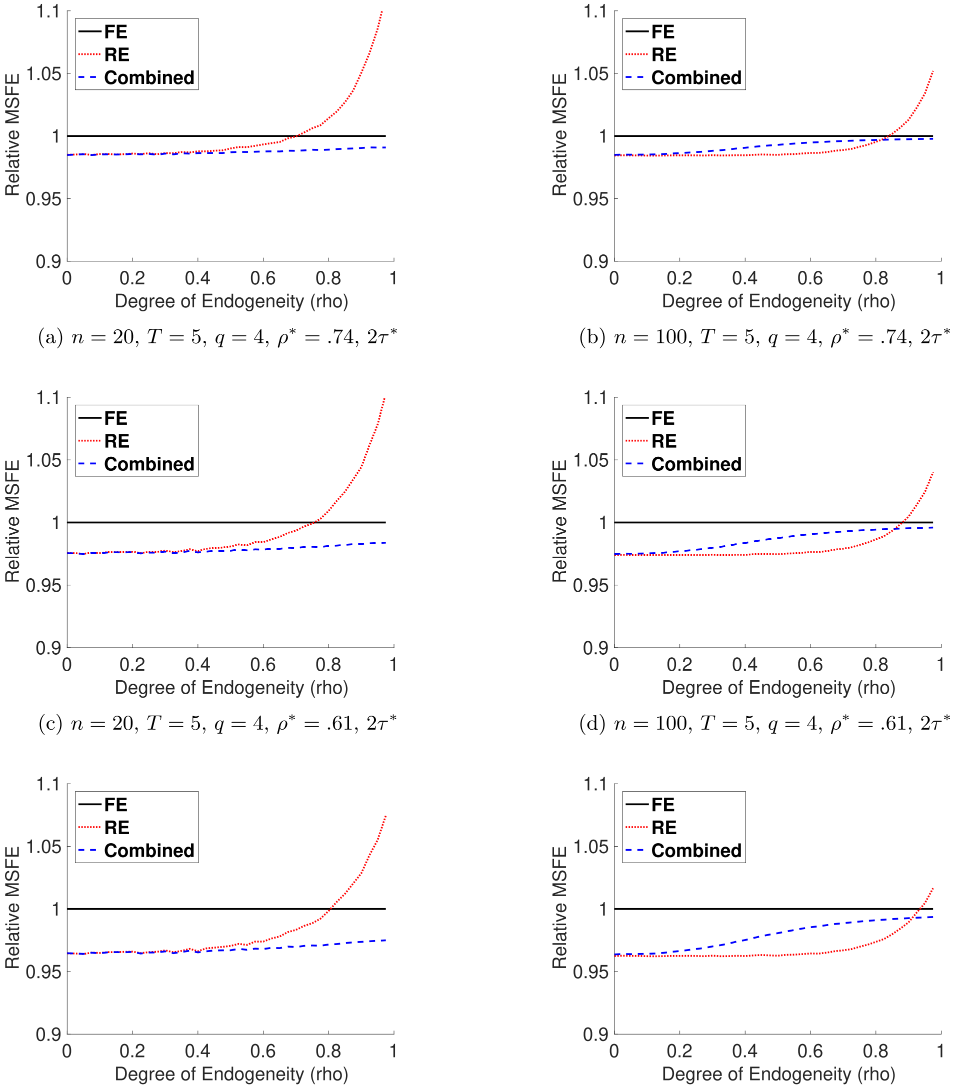
Fig. 3. Relative MedSE of FE, RE and combined estimators, $n = \{20, 100\}, T = 5, q = 4, \rho^* = \{.74, .61, .50\}, 2\tau^*$.

each approach with $\tau = 2\tau^*$. By contrasting Figs. 2 and 4, we see again that the region of dominance for the combined forecast over FE is greater for the large value of τ.

Fig. 5 plots the relative 1-step ahead out-of-sample MSFE of each approach with fixed δ on the interval [0, 5] on the horizontal abscissa for $n \in \{25, 100, 400\}$, which correspond to varying ranges of ρ on [0, 1], [0, .5], [0, .25] for the different sample size $n \in \{25, 100, 400\}$, respectively. These are the cases where the degree of endogeneity does not depend on the sample size. In this plot, we see again that the gain form the combined forecast has uniformly smaller MSFE than FE, with the MSFE converging to that of FE as the degree of endogeneity increases. This is consistent with holds, which shows that the improvements are asymptotic local to zero. This shows numerically that the improvements in Theorem 2 can be expected to hold broadly in the parameter space.

In summary, the simulation evidence provides strong finite sample confirmation of the predictions from the large sample theory on the estimation (Theorems 1 and 2). It also shows that the finite sample properties of the in-sample combined estimation is carried over to the out-of-sample combined forecasting. The improvement in the combined forecast over the FE forecast is greater for smaller heterogeneity ρ^*. For moderate to large ρ^* and higher ρ, or moderate to large ρ and higher ρ^*, the combined forecast is better than the RE forecast. For very large ρ^* and low ρ, the combined forecast is close to the RE forecast.

Remark 5. In constructing the Stein-like combined estimator we focus on the shrinkage parameter τ that makes the Stein estimator dominate FE in the asymptotic risk (Theorem 2), rather than making it dominate the RE estimator. The cost of



(a) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .74,\ 2\tau^*$

(b) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .74,\ 2\tau^*$

(c) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .61,\ 2\tau^*$

(d) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .61,\ 2\tau^*$

(e) $n = 20,\ T = 5,\ q = 4,\ \rho^* = .50,\ 2\tau^*$

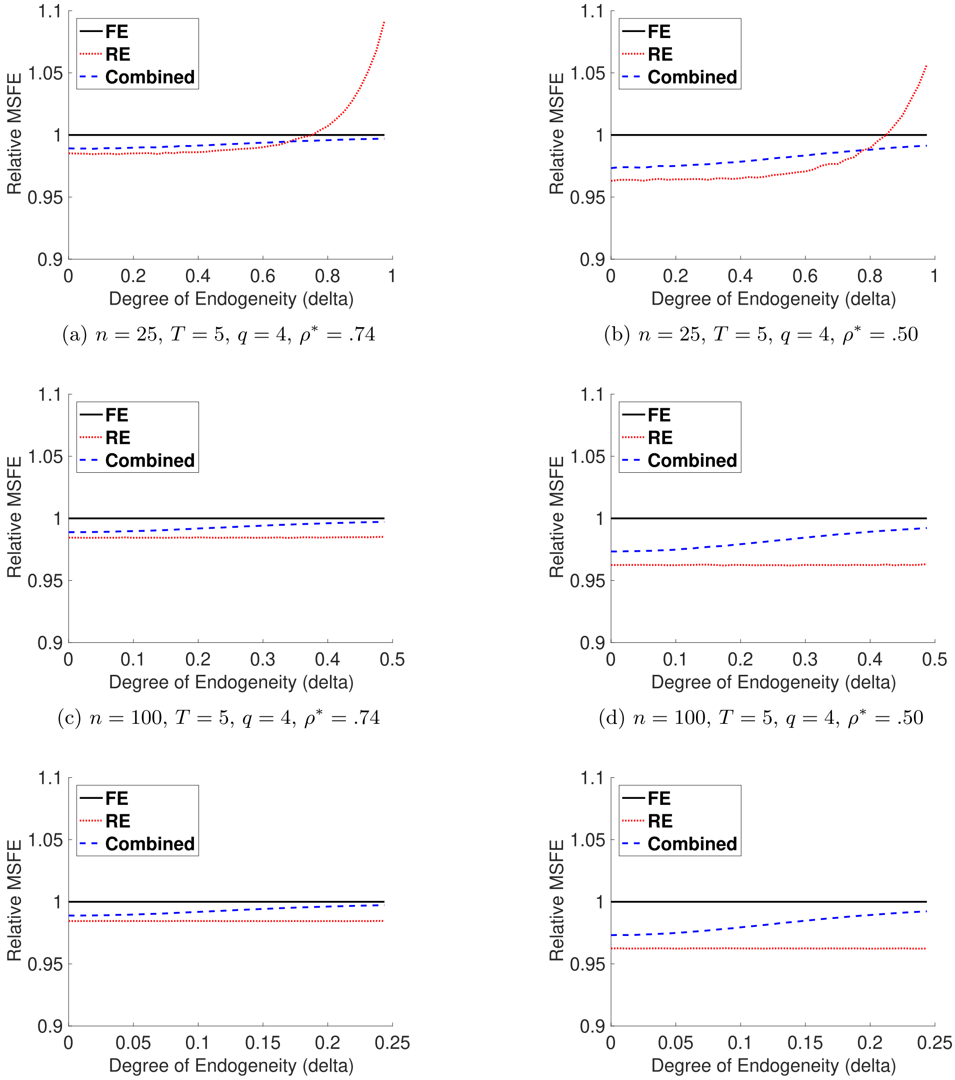
(f) $n = 100,\ T = 5,\ q = 4,\ \rho^* = .50,\ 2\tau^*$

Fig. 4. Relative MSFE of the FE, RE and combined forecasts, one-step forecast, $s = 1, n = \{20, 100\}, T = 5, q = 4, \rho^* = \{.74, .61, .50\}, 2\tau^*$.

winning over the FE is to increase the probability of losing to the RE when the endogeneity is weak (when ρ is small) because τ^* is too small for the Hausman statistic H_n to go below τ^*. Therefore, we also consider a two-step approach, the pretesting (PT) estimator in (45), using the Hausman statistic H_n to construct the PT estimator in (45) as done also in Hansen (2017). For the PT estimator based on the Hausman statistic, we use the 5% critical value $\chi^2_{q,0.95}$. Under the null of no endogeneity ($\rho = 0$), $\Pr(H_n < \chi^2_{q,0.95}) = \Pr(H_n < 9.49) = 0.95$ as $\chi^2_{q,0.95} = 9.49$ with d.f. $= q = 4$. Thus, PT will have a 95% chance to have the weight $w = 1$ to completely shrink FE towards RE, while the Stein combined estimator has only 5% chance to do that with $w = 1$.

Remark 6. The optimal choice of τ, τ^* in (33), is obtained to minimize the "bound" of the risk of $R(\widehat{\beta}_c)$ in Theorem 2. The bound is the RHS term in equation (32). However, as discussed in Remark 5 above, the optimal choice of $\tau = \tau^*$ is too small when ρ is small because the probability that the Hausman statistic is smaller than $\tau = \tau^*$ will be too small. Hence we have increased it to $2\tau^*$ as this choice still satisfies the classic James-Stein conditions in equations (30) and (31). This makes the MSE of the combined estimator closer to the MSE of FE when the degree of endogeneity is small. The results are reported in Figs. 3 and 4, where $\tau = 2\tau^*$ are considered. Hence, this indicates that the optimal choice of $\tau = \tau^*$ in (33) obtained from minimizing the "bound" of the risk of $R(\widehat{\beta}_c)$ is not "optimal" in the sense of minimizing the risk of $R(\widehat{\beta}_c)$ itself. To our knowledge, there is no result on this yet and thus we leave this for a future work.

Remark 7. The risk function (27) with a general weight matrix W includes many special cases. For example, the unweighted MSE is obtained by setting $W = I_q$, in which case the coefficients are of equal importance. The canonical case is motivated by



(a) $n = 25$, $T = 5$, $q = 4$, $\rho^* = .74$ (b) $n = 25$, $T = 5$, $q = 4$, $\rho^* = .50$

(c) $n = 100$, $T = 5$, $q = 4$, $\rho^* = .74$ (d) $n = 100$, $T = 5$, $q = 4$, $\rho^* = .50$

(e) $n = 400$, $T = 5$, $q = 4$, $\rho^* = .74$ (f) $n = 400$, $T = 5$, $q = 4$, $\rho^* = .50$

Fig. 5. Relative MSFE of the FE, RE and Combined Forecasts, One-Step Forecast, $s = 1, n = \{25, 100, 400\}, T = 5, q = 4, \rho^* = \{.74, .50\}, \delta$.

ease of use and simplicity, which is obtained by setting $W = (V_{FE} - V_{RE})^{-1}$. This choice simplifies many formulae, e.g., equation (30) has the simplification $d = q$, and the optimal choice of τ, τ^* in (33) is $q - 2$. Following Hansen (2017), we set $\tau = q - 2$, which would be the same as the optimal choice $\tau^* = d - 2$ when $W = (V_{FE} - V_{RE})^{-1}$ is used. See Corollary 1. See also Remark 3 and equation (33). However, when $W = I_q$ is used instead of $W = (V_{FE} - V_{RE})^{-1}$, it is possible that the condition $0 < \tau < 2(d - 2)$ in equation (31) may not hold if $\tau = q - 2$ is used, especially when the dimension q of X is large. In that case, we should use the theoretical optimal $\tau = d - 2 = \frac{\mathrm{tr}(W(V_{FE} - V_{RE}))}{\lambda_1} - 2$.

Remark 8. If τ is small, then w will be small towards zero. The Stein-like combined estimator puts more weights on FE, resulting in less bias and more variance; If τ is large, then w will be large towards one. The Stein combined estimator puts more weights on RE, resulting in more bias and less variance.

5. Application

There have been numerous studies on the price and income elasticities of residential natural-gas and electricity demand. Maddala et al. (1997) applied classical, empirical Bayes, and Bayesian procedures to the problem of estimating short-run and long-run elasticities of residential demand for electricity and natural gas in the US for each of 49 states over the period 1970–1990. They found that shrinkage Bayesian type estimators are superior to either the individual heterogeneous estimates or the homogeneous estimates, especially for prediction purpose, through shrinking the individual estimates towards the pooled estimate using weights depending on their corresponding variance-covariance matrices.

Using the Maddala et al. (1997) specification and data sets, Baltagi et al. (2002) compare the out-of-sample forecast performance of homogeneous and heterogeneous estimators applying them to electricity and natural-gas. In this section, we compare the performances of the residential gas and electricity demand forecast using panel data across 51 states (including Washington DC) over the period 1997–2012. The annual state residential electricity and gas price data used in this study were obtained from The State Energy Price and Expenditure System of the U.S. Energy Information Administration. Annual personal income per capita by state were drawn from the Bureau of Business and Economic Research, and the annual Consumer Price Index for the United States was from CITIBASE.

Following Baltagi et al. (2002), we consider the following panel data model:

$$\widehat{y}_{i,T+s} = \widehat{\beta}_0 + \widehat{\beta}_1 x_{1i,T} + \widehat{\beta}_2 x_{2i,T} + \widehat{\beta}_3 x_{3i,T} + \widehat{\alpha}_i, \tag{47}$$

where $i = 1, \ldots, 51, t = 1, \ldots, 14$. The LHS variable of the equation is $y = \log$ (residential electricity per capita consumption) for the electricity demand equation or $y = \log$ (residential natural-gas per capita consumption) for the natural gas demand equation. The RHS variables for the electricity demand equation are $x_1 = \log$ (real per capita personal income), $x_2 = \log$ (real residential electricity price), and $x_3 = \log$ (real residential natural-gas price). The RHS variables for the natural-gas demand equation are $x_1 = \log$ (real per capita personal income), $x_2 = \log$ (real residential natural-gas price), and $x_3 = \log$ (real residential electricity price).

We compare alternative estimators in the prediction performance. Given the large data set of $N = 51$ states over $T = 14$ years, we estimate our model using a truncated data set (i.e. without the last 3 years of data) and then apply each estimator to an out-of-sample forecast period. Panel A of Table 1 gives a comparison of forecasts using the root-mean-squared-forecast-errors (RMSFE) for residential electricity demand and similarly Panel B for residential natural-gas demand. Because the ability of an estimator to characterize the long-run as well as the short-run response is at issue, the RMSFE is calculated across the 51 states at different forecast horizons. The RMSFEs are reported in Table 1 for forecast horizons $s = 1, 3$ years. Both for the electricity demand (Panel A) and for the natural-gas demand (Panel B), the combined estimator dominates FE forecast whether it is for the 1-year ahead or 3-years ahead forecasts, confirming that the gains in estimation by the combined estimator (Theorem 2) can benefit the combined forecast in out-of-sample forecasting. The combined forecast using the combined estimator performs better than both FE and RE forecasts.

The overall forecast ranking in RMSFE offers a clear and strong endorsement for the combined forecast which is constructed using the Stein-like combined estimator $\widehat{\beta}_c$. The "Stein-like combined forecast" is superior to both the FE forecast and the RE forecast in out-of-sample prediction.

6. Conclusions

The goal of this paper is to examine if the theoretical results on the combined estimation for the parametric panel data model with weak endogeneity (i.e., local to exogeneity) may be useful to form the combined forecasting in the panel data model such that the gains in asymptotic risk in the combined estimator may be carried over to the gains in mean squared forecast errors. We examine this by asymptotic theory, by Monte Carlo simulation, and empirical applications. The FE and RE forecasts are combined when the RE estimator suffers from various degrees of endogeneity to produce a combined forecast. Specifically we show that the forecast combining the FE and RE models can outperform the FE model forecast in terms of mean squared forecast error. Our simulation experiment shows that the combined forecast can uniformly dominate the FE forecast for all degrees of endogeneity, demonstrating that the in-sample estimation result is carried over to the out-of-sample forecasting. It also shows that the combined forecast can reduce MSFE relative to the RE forecast for moderate to large degrees of endogeneity and heterogeneity in the individual effects.

The use of the combined forecasting approach allows applied researchers to implement efficient forecasting under the presence of weak endogeneity. Even when there is no endogeneity or when there is strong endogeneity, without having to select a consistent forecast or an efficient forecast, the weights in the combined estimator will be 1 or 0. Hence, the combined forecast is particularly useful when the degree of endogeneity is weak or when it is not clear which of the RE or FE panel data models to choose.

Table 1
Combined forecast from the combined estimator.

	A. Electricity Demand		B. Natural Gas Demand	
	$s = 1$	$s = 3$	$s = 1$	$s = 3$
FE	5.5261	6.4135	6.1593	8.1922
RE	6.6025	7.9441	7.1044	9.0036
Combined	5.2837	6.1467	5.7310	7.4145

51 U.S. States including Washington DC. Reported are RMSFE for forecast horizons $s = 1, 3$.

Conflicts of interest

The authors declare no conflict of interest.

Appendix

Proof of Theorem 1

Let $h_1 = G_1 h$ and $\xi_1 = G_1 \xi$ with $G_1 = (I \ \ 0)'$, and let $h_2 = G_2 h$ and $\xi_2 = G_2 \xi$ with $G_2 = (0 \ \ I)'$. Here we derive the asymptotic distribution of the RE estimator for the parametric panel data model. Note that

$$\widehat{\beta}_{RE} = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y$$
$$= \left(X'\left(\frac{P}{\sigma_1^2}+\frac{Q}{\sigma_u^2}\right)X\right)^{-1}X'\left(\frac{P}{\sigma_1^2}+\frac{Q}{\sigma_u^2}\right)y$$
$$= \left(X'(\lambda P+Q)X)^{-1}X'\lambda P+Q\right)y \equiv Ay,$$

where $\lambda = \frac{\sigma_u^2}{\sigma_1^2}$, $\Omega^{-1} = \frac{1}{\sigma_u^2}(\lambda P+Q) = \frac{P}{\sigma_1^2}+\frac{Q}{\sigma_u^2}$. We then have

$$\sqrt{n}\left(\widehat{\beta}_{RE}-\beta\right) = AD\overline{X}\delta + \left(\frac{1}{n}X'(\lambda P+Q)X\right)^{-1}\frac{1}{\sqrt{n}}X'(\lambda P+Q)(D\varepsilon+u) \rightarrow h_1+\xi_1,$$

where

$$\xi_1 \sim \left(\text{plim}\frac{1}{n}X'(\lambda P+Q)X\right)^{-1}Z,$$

with

$$Z = \frac{1}{\sqrt{n}}X'(\lambda P+Q)(D\varepsilon+u) \sim N\left(0,\ \sigma_u^2\left(\text{plim}\frac{1}{n}X'(\lambda P+Q)X\right)\right).$$

Hence,

$$\xi_1 \sim N\left(0,\ \sigma_u^2\left(\text{plim}\left(\frac{1}{n}X'(\lambda P+Q)X\right)\right)^{-1}\right)$$
$$= N\left(0,\ \left(\text{plim}\frac{1}{n}X'\Omega^{-1}X\right)^{-1}\right) = N(0,\ V_1).$$

The asymptotic bias h_1 is

$$h_1 = \left(\text{plim}\frac{1}{n}X'(\lambda P+Q)X\right)^{-1}\left(\text{plim}\frac{1}{n}X'(\lambda P+Q)D\overline{X}\right)\delta$$
$$= \sigma_u^{-2}\left(\text{plim}\frac{1}{n}X'\Omega^{-1}X\right)^{-1}\left(\text{plim}\lambda\frac{1}{n}\overline{XX}\right)\delta$$
$$= \sigma_1^{-2}V_1\Sigma\delta,$$

where the second equality follows from noting that $QD = 0$, $PD = D$, and $X'D = \overline{X}$, and the last equality follows from denoting $\Sigma \equiv \text{plim}\ \frac{1}{n}\overline{XX}$.

Proof of Theorem 2

First, we derive the asymptotic risk of the RE estimator $R(\widehat{\beta}_{RE})$ for the parametric panel data model. From Theorem 1.1,

$$\sqrt{n}\left(\widehat{\beta}_{RE}-\beta\right) \rightarrow h_1+\xi_1,$$

with $\xi_1 \sim N(0, V_1)$. Hence,

$$\begin{aligned} R\left(\widehat{\beta}_{RE}\right) &= \mathrm{tr}\left[E(\xi_1 W \xi_1{}') + h_1 W h_1{}'\right] \\ &= \mathrm{tr}\, E(\xi_1 W \xi_1{}') + \mathrm{tr}(h_1 W h_1{}') \\ &= \mathrm{tr}\,(WV_1) + \sigma_1^{-4}\delta'\Sigma V_1 W V_1 \Sigma\delta. \end{aligned}$$

For the asymptotic risk of the FE estimator, note that $\sqrt{n}(\widehat{\beta}_{FE} - \beta) \to \xi_2 \sim N(0, V_2)$, and thus

$$R\left(\widehat{\beta}_{FE}\right) = E(\xi_2 W \xi_2{}') = \mathrm{tr}(WV_2).$$

The rest of the proof is based on Theorem 2 of Hansen (2017). 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^*.

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

we can calculate that

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

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

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

Since

$$GWG_2{}'V = WG_2{}'VG = W(V_2 - V_1)$$

and

$$GWG_2{}'VB = GWG_2{}'VG(V_2 - V_1)^{-1}G' = GWG',$$

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

Thus

$$E\left(\psi^{*'}W\psi^{*}\right)=R\left(\widehat{\beta}_{FE}\right)+\tau^{2}E\left(\frac{(h+\xi)^{'}GWG'(h+\xi)}{\left((h+\xi)^{'}B(h+\xi)\right)^{2}}\right)+4\tau E\mathrm{tr}\left(\frac{(h+\xi)'GWG'(h+\xi)}{\left((h+\xi)^{'}B(h+\xi)\right)^{2}}\right)$$
$$-2\tau E\mathrm{tr}\left(\frac{(W(V_2-V_1))}{(h+\xi)^{'}B(h+\xi)}\right).$$

Define $B_1=(V_2-V_1)^{-\frac{1}{2}}G'$ and $A^{*}=(V_2-V_1)^{\frac{1}{2}}W(V_2-V_1)^{\frac{1}{2}}$. Note that $GWG_2{}'VP=GWG'=B_1{}'A^{*}B_1$, $B_1{}'B_1=B$. Using the inequality $b'ab\leq(b'b)\lambda_{\max}(a)$ for symmetric a, and let

$$\lambda_{\max}(a)=\lambda_{\max}(W(V_2-V_1))=\lambda_1.$$

Then

$$\begin{aligned}&\mathrm{tr}\left(B(h+\xi)(h+\xi)'GWG_2{}'V\right)=(h+\xi)'B_1{}'A^{*}B_1(h+\xi)\\&\leq(h+\xi)'B(h+\xi)\lambda_1.\end{aligned}\tag{48}$$

Using equation (48) and Jensen's inequality, we have

$$\begin{aligned}E\left(\psi^{*'}W\psi^{*}\right)&\leq R\left(\widehat{\beta}_{FE}\right)+\left(\tau^{2}+4\tau\right)E\left(\frac{\lambda_1}{(h+\xi)^{'}B(h+\xi)}\right)-2\tau E\mathrm{tr}\left(\frac{(W(V_2-V_1))}{(h+\xi)^{'}B(h+\xi)}\right)\\&=R\left(\widehat{\beta}_{FE}\right)-E\left(\frac{\tau(2(\mathrm{tr}(W(V_2-V_1))-2\lambda_1)-\tau\lambda_1)}{(h+\xi)^{'}B(h+\xi)}\right)\\&\leq R\left(\widehat{\beta}_{FE}\right)-\frac{\tau(2(\mathrm{tr}(W(V_2-V_1))-2\lambda_1)-\tau\lambda_1)}{E\left((h+\xi)^{'}B(h+\xi)\right)}.\end{aligned}\tag{49}$$

Since $\mathrm{tr}(BV)=\mathrm{tr}(G(V_2-V_1)^{-1}G'V)=q$. We have

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

Substituted into (49) we have

$$R\left(\widehat{\beta}_{c}\right)<R\left(\widehat{\beta}_{FE}\right)-\frac{\tau(2(\mathrm{tr}(W(V_2-V_1))-2\lambda_1)-\tau\lambda_1)}{\sigma_1^{-4}\delta'\Sigma V_1(V_2-V_1)^{-1}V_1\Sigma\delta+q}$$

with $0<\tau\leq 2\left(\frac{\mathrm{tr}(W(V_2-V_1))}{\lambda_1}-2\right)$

Proof of Corollary 2

$R(\widehat{\beta}_{RE})\leq R(\widehat{\beta}_{FE})$ when $\sigma_1^{-4}\delta'\Sigma V_1WV_1\Sigma\delta\leq q$; and $R(\widehat{\beta}_{RE})>R(\widehat{\beta}_{FE})$ otherwise.

$$R\left(\widehat{\beta}_{RE}\right)=\mathrm{tr}(WV_1)+\sigma_1^{-4}\delta'\Sigma V_1WV_1\Sigma\delta$$
$$R\left(\widehat{\beta}_{FE}\right)=\mathrm{tr}(WV_2)$$
$$\mathrm{tr}(W(V_2-V_1))=q\text{ if }W=(V_2-V_1)^{-1}$$
$$R\left(\widehat{\beta}_{FE}\right)-R\left(\widehat{\beta}_{RE}\right)=\mathrm{tr}(W(V_2-V_1))-\sigma_1^{-4}\delta'\Sigma V_1WV_1\Sigma\delta=q-\sigma_1^{-4}\delta'\Sigma V_1WV_1\Sigma\delta.$$

References

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of Forecasting, 27*(2), 153–173.
Baltagi, B. H., Bresson, G., & Pirotte, A. (2002). Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: Some empirical evidence from US electricity and natural-gas consumption. *Economics Letters, 76*(3), 375–382.
Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society, 20*(4), 451–468.
Chamberlain, G. (1984). Panel data. *Handbook of Econometrics, 2*, 1247–1318.

Forni, M., Hallim, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics, 82*, 540–554.
Forni, M., Hallim, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association, 100*, 830–840.
Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association, 57*(298), 369–375.
Hansen, B. E. (2016). Efficient shrinkage in parametric models. *Journal of Econometrics, 190*(1), 115–132.
Hansen, B. E. (2017). A stein-like 2SLS estimator. *Econometric Reviews, 36*, 840–852.
Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica, 46*, 1251–1271.
Henderson, D. J., & Ullah, A. (2005). A nonparametric random effects estimator. *Economics Letters, 88*(3), 403–407.
Huang, B. (2015). *A combined fixed and random effects estimator for parametric panel model*. Riverside: University of California.
Maddala, G. S., & Mount, T. D. (1973). A comparative study of alternative estimators for variance components models used in econometrics applications. *Journal of the American Statistical Association, 68*, 324–328.
Maddala, G. S., Trost, R. P., Li, H., & Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business & Economic Statistics, 15*, 90–100.
Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica, 46*, 69–85.
Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time-series of cross-sections. *Econometrica, 72*, 214–253.
Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1*, 399.
Stock, J., & Watson, M. (1999). Forecasting inflation. *Journal of Monetary Economics, 44*, 293–335.
Stock, J., & Watson, M. (2002a). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics, 20*, 147–162.
Stock, J., & Watson, M. (2002b). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association, 460*, 1167–1179.
Wang, Y., Zhang, Y., & Zhou, Q. (2016). A Stein-like estimator for linear panel data models. *Economics Letters, 141*, 156–161.