Title
Combined Estimation and Forecasting for Panel Data Models: Parametric and Semi-Parametric

Author
Huang, Bai

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Combined Estimation and Forecasting for Panel Data Models: Parametric and Semi-Parametric

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Bai Huang

June 2017

Dissertation Committee:

Professor Aman Ullah, Co-Chairperson
Professor Tae-Hwy Lee, Co-Chairperson
Professor Gloria Gonzalez-Rivera
Professor Jang-Ting Guo

The Dissertation of Bai Huang is approved:

Co-Chairperson

Co-Chairperson

University of California, Riverside

Acknowledgments

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my parents.

I would like to express my deepest gratitude to my commmitte chairs, Professor Aman Ullah and Professor Tae-Hwy Lee, for their excellent guidance, patience, motivation, and immense knowledge. Professor Ullah has taught me more than I could ever give him credit for here. He has shown me, by his example, what a good economist should be. Professor Lee has given his knowledge and put in effort at all times for the benefit of this dissertation. He worked closely with me during the proposal writing and during the period of my dissertation. Without their help and encouragement this dissertation would not have been written (or ever finished!).

I am also very grateful to the rest of my dissertation committee: Professor Gloria Gonzalez-Rivera and Professor Jang-Ting Guo, for their insightful advice and feedback. Their varied perspectives have helped me to strengthen my work. I would also like to thank Professor David A. Malueg for his support and encouragement. Special thanks goes to Professor Victor Ortego-Marti, who was willing to participate in my final defense at the last moment.

I would also like to express my thanks to Gary Kuzas for his unfailing support and assistance.

Thanks must also go to all my instructors and teachers, who throughout my educational career have supported and encouraged me to believe in my abilities. They have directed me through various situations, allowing me to reach this accomplishment.

Many friends and fellow graduate students have provided much needed emotional support.

Last but not the least, I would like to thank my parents, Ningmei Hao and Wei

Huang, whose love and guidance are with me in whatever I pursue. They were always there cheering me up and stood by me through the good times and bad.

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Combined Estimation and Forecasting for Panel Data Models: Parametric and Semi-Parametric

by

Bai Huang

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2017
Professor Aman Ullah, Co-Chairperson
Professor Tae-Hwy Lee, Co-Chairperson

This dissertation covers several topics in estimation and forecasting in panel data models.

Chapter one considers the panel data model with correlated individual effects and regressors. We form a combined estimator from combining the fixed effects (FE) and random effects (RE) estimators. We derive the asymptotic distribution and the asymptotic risk of our estimator using a local asymptotic framework. We show that if the regressor dimension exceeds two, the asymptotic risk of the combined estimator is strictly less than that of FE estimator. Our simulation result shows that the combined estimator can reduce finite sample MSE relative to the FE estimator for all degrees of endogeneity and heterogeneity, as well as relative to the RE estimator for moderate to large degrees of endogeneity and heterogeneity. We also apply the combined estimator to revisit the relationship between public capital infrastructure and private economic performance.

Chapter two extends chapter one into the semi-parametric (SP) framework, and proposes a combined SP-FE and SP-RE estimator.

Chapter three considers the panel data model with correlated residuals and regressors. In the presence of such correlation, both FE and RE estimators yield biased and inconsistent estimates of the parameter. We propose a combined FE and FE-2SLS estimator, and a combined RE and RE-2SLS estimator.

Chapter four considers regression models for panel data that exhibit cross-section dependence due to common shocks. Model with factor structures for errors and regressors are considered. In this case, the FE estimator is inconsistent. To solve this problem, Pesaran (2006) introduced the common correlated effects pooled (CCEP) estimator. We propose a combined FE and CCEP estimator, and show that under certain conditions, the combined estimator has strictly smaller risk than the CCEP estimator. Finally, we use Holly et al. (2010) state-level housing data to show the applicability of the combined estimator.

Chapter five proposes a combined approach to econometric forecasting. Monte-Carlo simulations are conducted to evaluate the performance of the combined forecast in finite samples. We contrast the out-of-sample forecast performance of the FE, RE and the combined approaches using the electricity and natural gas data sets.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The endogeneity problem in the panel data model has received a great deal of attention in the literature. Hausman (1978) takes the difference between an efficient estimator and a robust estimator and derive the Hausman test. We show that it is possible to make a uniform improvement on the consistent estimator in terms of risk when there exhibits weak endogeneity. Consider a panel data model, let

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}$$

where x_{it} is the ith observation on q explanatory variables, β is a $q \times 1$ unknown parameter, α_i are known as the individual effects and u_{it} is the random error.

Chapter one examines the case of potential correlation of α_i with the columns of X within the parametric panel data framework. In the presence of such correlations, the random effects (RE) estimator yields biased and inconsistent estimates of the parameters. The traditional technique to overcome this problem is to eliminate the individual effects in the sample by transforming the data into deviations from individual means, which is known as fixed effects (FE) estimator. Researchers commonly approach the panel data using FE estimator to avoid endogeneity problem. Unfortunately, in some applications, primary interest is attached to the unknown coefficients of time-invariant variables which can not be estimated by the FE estimator. On the other hand, if α_i are not correlated with the other regressors in the model, both RE and FE estimators are consistent and RE estimator is efficient. In this chapter, we propose a combined estimator, which is a weighted combination of FE and RE estimators with weights depending on Hausman test statistic. We derive the asymptotic distribution of the combined estimator using a local-to-exogeneity condition and calculate the asymptotic risk of the estimator and find that the asymptotic risk of the combined estimator is strictly less than that of FE estimator, when the number of regressors exceeds two. Our simulation result shows that the combined estimator can reduce finite sample MSE relative to the FE estimator for all degrees of endogeneity and heterogeneity, as well as relative to the RE estimator for moderate to large degrees of endogeneity and heterogeneity. Finally, we use a panel data for the 48 contiguous U.S. states in each year between 1970 and 1986 to revisit the relationship between public infrastructure and private economic performance.

Chapter two extends chapter one into the semi-parametric framework. We propose a combined estimator, which is a weighted combination of the semi-parametric fixed effects and semi-parametric random effects estimators. We show that the combined estimator uniformly dominates the semi-parametric fixed effects estimator for all degrees of endogeneity and heterogeneity. It also has smaller asymptotic risk compared to the semi-parametric random effects estimator unless the endogeneity is very weak. Based on simulations, we find that the above results also hold for small samples. The magnitude of efficiency of the combined estimator varies with respect to the degree of endogeneity and heterogeneity.

Chapter three (joint with Aman Ullah) and chapter four (joint with Tae-Hwy Lee and Aman Ullah) examine the case of potential correlation of u_{it} with the columns of

X. In the presence of such correlations, both FE and RE estimators yield biased and inconsistent estimates of the parameter. The traditional technique to overcome this problem is to find instruments for those explanatory variables which are potentially correlated with idiosyncratic errors. However, the finite sample properties of the 2SLS estimator are often problematic. In this chapter, we propose two combined estimators, which are weighted average of FE and FE-2SLS estimators, and weighted average of RE and RE-2SLS estimators with the weights depending on Hausman statistic. The asymptotic distribution and risk of the combined estimators are derived using a local asymptotic framework. In the Monte Carlo study, we show that the combined estimators uniformly dominate the individual effects estimators for all degrees of endogeneity and heterogeneity. The combined estimators are also better than the individual effects estimators except when the degree of endogeneity is very small, or when both very small sample size and very weak instruments are satisfied. Finally, we use a panel data on 90 counties in North Carolina over the period 1981–1987 to revisit the effect of police on crime using the FE, FE-2SLS and combined estimators.

The case examined by chapter four features the panel data that exhibits cross-section dependence due to common shocks. 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. One popular approach to this problem is the common correlated effects method proposed by Pesaran (2006) which eliminates the error cross-sectional dependence using cross-sectional averages of the data. This approach yields consistent and asymptotically normal parameter estimates when T is fixed as $N \to \infty$. In this chapter, we propose a combined estimator which is a weighted combination of the fixed effects estimator and the common correlated effects pooled estimator of Pesaran (2006). 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 estimator has strictly smaller risk than the CCEP estimator. Following Holly, Pesaran, and Yamagata (2010) analysis of changes in real house prices, we examine the performance of the combined estimator using a panel of 49 states over the period 1975–2011.

Chapter 5 contributes to the literature on forecast uncertainty by investigating the forecast model combination in the panel data model. First, we calculate the coefficients based combination weights depending on Hausman test statistic. Second, we show that under endogeneity, the forecast combining both fixed effects and random effects models using the weights from step one outperforms forecast with fixed effects in terms of mean squared forecast error. We illustrate this method with an application to forecasting electricity and natural-gas demands for 51 U.S. states over the period 1997–2012. Overall, these results show promise for the combined forecast.

Chapter 2

A Combined Estimator for the Panel Data Model

2.1 Introduction

In a panel data model, the individual effect terms can be modeled as either random or fixed effects. Historically, econometricians have argued whether we should use fixed effects or random effects estimator when estimating a panel data model. Some econometricians, and most statisticians, have been in favor of using random effects estimator. Balestra and Nerlove (1966) and Maddala (1971) were advocates of the random error component model. They believe that the heterogeneity parameters appearing in the panel data models should be treated as random like we treat equational errors (see, for example, Hyslop 1999, Olsen and Schafer 2001, Scheike et al. 2010). However, the random effects estimator becomes inconsistent since the individual effect terms are often correlated with regressors, which are endogenous. In view of this, large number of econometricians, especially practitioners, use fixed effects estimator. Mundlak (1961) and Wallace and Hussain (1969) were advocates of the fixed effects model. The fixed effects estimator is a consistent estimator under the endogeneity of regressors (see, for example, Rockoff 2004, Rivkin et al. 2005, Pedroni 2007). Since the fixed effects model de-means all variables in the model and thereby eliminates the correlated individual effects, it is widely used to control for individual heterogeneity in the data. In fact, this practice goes on ignoring the use of Hausman test (Hausman, 1978) of endogeneity which is based upon a contrast between fixed effects and random effects estimators. Mundlak (1978) argued that the random effects model assumes exogeneity of all the regressors with the random individual effects. In contrast, the fixed effects model allows for endogeneity of all the regressors and the individual effects. Hausman and Taylor (1981) allowed for some of the regressors to be correlated with the individual effects, as opposed to this all or nothing choice.

Under this scenario, we provide a fixed effects and random effects combined estimator which can improve the estimation efficiency. Motivated by Hansen (2014), we derive the asymptotic distribution of the combined estimator using a local-to-exogeneity condition and calculate the asymptotic risk of the estimator. We find that the combined estimator has strictly smaller asymptotic risk that the fixed effect estimator, when the number of regressors exceeds two. The properties of the combined fixed effects and random effects estimator for the parametric panel data model are discussed in Huang (2015) and Wang et al. (2016). Our simulation result shows that the combined estimator can reduce finite sample MSE relative to the fixed effects estimator for all degrees of endogeneity and heterogeneity, as well as relative to the random effects estimator for moderate to large degrees of endogeneity and heterogeneity. We also discuss four combined estimators from combining the fixed effects and random effects estimators in the panel data model, including leave-one-out estimator, inverse-variance weighted combined estimator, the proposed combined estimator and the combined estimator using optimal weights. We compare the performance of these estimators using a series of Monte Carlo experiments that vary the sample sizes, degrees of endogeneity and degrees of hererogeniety. Finally, to examine the applicablity of the combined estimator, we use a panel data for the 48 contiguous U.S. states in each year between 1970 and 1986 to revisit the relationship between public infrastructure and private economic performance.

The rest of this chapter is organized as follows. Section 2 presents the model and estimators. Section 3 presents the asymptotic distribution of the combined estimator. Section 4 derives the asymptotic risk of the combined estimator. Monte Carlo simulation is provided in Section 5. Empirical example is given in Section 6. Section 7 concludes.

2.2 The Model and Estimators

2.2.1 The Fixed Effects Estimator

Consider a panel data regression model

$$y_{it} = x_{it}\beta + \alpha_i + u_{it} \tag{2.1}$$

where $i = 1, 2, \ldots n$ and $t = 1, 2, \ldots T$. β is $q \times 1$ and x_{it} is the ith observation on q explanatory variables, β is a $q \times 1$ unknown parameter, α_i is known as the individual effect and u_{it} is the random error.

For the fixed effects (FE) case, the α_i are assumed to be fixed parameters to be estimated and remainder disturbances stochastic with u_{it} independent and identically distributed $(0, \sigma_u^2)$. The x_{it} are assumed independent of the u_{it} for all i and t. Write

(2.1) in matrix form

$$y = X\beta + D\alpha + u \tag{2.2}$$

where $y = (y_{11}, \ldots, y_{1T}, y_{21}, \ldots, y_{nT})'$ is $nT \times 1$, $X = (x_{11}, \ldots, x_{1T}, \ldots, x_{n1}, \ldots, x_{nT})$ is $nT \times q$ and $u \sim (0,\ \sigma_u^2 I_{nT})$. Let ι_T be a vector of ones, $D = I_n \otimes \iota_T$ is $nT \times n$. Note that $DD' = I_n \otimes J_T$ where $J_T = \iota_T \iota_T'$, $D'D = TI_n$. $P = I_n \otimes \bar{J}_T$ where $\bar{J}_T = J_T/T$. P is a matrix which averages the observation across time for each individual, and $Q = I_{nT} - P$ is a matrix which obtains the deviations from individual means.

One can premultiply (2.2) by Q and perform OLS on the resulting transformed model:

$$Qy = QX\beta + QW$$

The $\hat{\beta}_{FE}$ is

$$\hat{\beta}_{FE} = \left(X'QX\right)^{-1} X'Qy \tag{2.3}$$

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

$$\sqrt{n}\left(\hat{\beta}_{FE} - \beta\right) \xrightarrow{d} N\left(0, V_2\right)$$

where $V_2 = \sigma_u^2 \left(\text{plim}\frac{X'QX}{n}\right)^{-1}$.

2.2.2 The Random Effects Estimator

The random effects (RE) model assumes $\alpha_i \sim$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 x_{it} are independent of the α_i and u_{it} for all i and t. Under this assumption, we can write

$$y_{it} = x_{it}\beta + v_{it}, \quad \mathbb{E}\left(v_{it}|x_i\right) = 0 \tag{2.4}$$

where $v_{it} = \alpha_i + u_{it}$. Write (2.4) in matrix form

$$y = X\beta + v, \quad v = D\alpha + u \tag{2.5}$$

The variance-covariance matrix of Ω is given by

$$\Omega = \sigma_\alpha^2 (I_n \otimes J_T) + \sigma_u^2 (I_n \otimes I_T) = \sigma_1^2 P + \sigma_u^2 Q$$

where $\sigma_1^2 = T\sigma_\alpha^2 + \sigma_u^2$. The feasible estimator of $\hat{\Omega}$ of Ω can be obtained by running the OLS regression y on X, define

$$\hat{\sigma}_1^2 = \frac{T}{n}\sum_{i=1}^{n} \bar{\hat{v}}_i^2, \quad \hat{\sigma}_u^2 = \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\hat{v}_{it} - \bar{\hat{v}}_i\right)^2$$

where $\hat{v}_{it} = y_{it} - x_{it}\hat{\beta}_{OLS}$ is the OLS residual and $\hat{\beta}_{OLS} = (X'X)^{-1} X'y$. Noting that $\hat{\sigma}_1^2$ and $\hat{\sigma}_u^2$ estimate $\hat{\sigma}_\alpha^2 = \frac{1}{T}(\hat{\sigma}_1^2 - \hat{\sigma}_u^2)$. With these estimates, one can obtain the generalized least squares (GLS) of β based on (2.5) is

$$\hat{\beta}_{RE} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y \tag{2.6}$$

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

$$\sqrt{n}\left(\hat{\beta}_{RE} - \beta\right) \xrightarrow{d} N(0, V_1)$$

where $V_1 = \left(\text{plim}\frac{X'\Omega^{-1}X}{n}\right)^{-1}$.

2.2.3 The Combined Estimator

See Hausman (1978), 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 $\mathbb{E}(\alpha_i x_{it}) \neq 0$, $\hat{\beta}_{RE}$ is biased and inconsistent while $\hat{\beta}_{FE}$ is not affected. Motivated by this observation, we would like to see if combination of $\hat{\beta}_{RE}$ and $\hat{\beta}_{FE}$ can result in an improved estimation. We propose the following combined estimator of β:

$$\hat{\beta}_c = w\hat{\beta}_{RE} + (1-w)\hat{\beta}_{FE} \tag{2.7}$$

where

$$
\begin{aligned}
w &= \begin{cases} \frac{\tau}{H_n} & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases} \\
H_n &= (\hat{\beta}_{FE} - \hat{\beta}_{RE})'(\hat{V}_{FE} - \hat{V}_{RE})^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE})
\end{aligned}
$$

and τ is a shrinkage parameter. We suggest set $\tau = q - 2$ when $q > 2$. The degree of shrinkage depends on the ratio τ/H_n. When $H_n < \tau$ then $\hat{w} = \hat{\beta}_{RE}$, When $H_n \geq \tau$ then $\hat{\beta}_c$ is a weighted average of $\hat{\beta}_{RE}$ and $\hat{\beta}_{FE}$, with more weight on $\hat{\beta}_{RE}$ when τ/H_n is large. Alternatively, it can be written as a positive-part James-Stein Estimator

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

where $(b)^+ = b$ if $b > 0$, and 0 if $b \leq 0$.

2.2.4 Asymptotic Distribution

We use the local asymptotic approach. Write α_i as a linear function of $\bar{x}_i = \sum_t x_{it}/T$

$$
\begin{aligned}
\alpha_i &= \bar{x}_i \rho + \epsilon_i, & (2.8) \\
E(x_{it}\epsilon_i) &= 0
\end{aligned}
$$

The variable x_{it} are exogenous if α_i and x_{it} are uncorrelated, or equivalently that the coefficient ρ is zero. For fixed T, ρ is local to zero

$$\rho = \frac{1}{\sqrt{n}}\delta \tag{2.9}$$

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 endogeneity and FE estimator is chosen. If $\delta = 0$, x_{it} are exogenous and RE estimator is preferred.

Now, we make the following assumptions:

Assumption 2.1. $(x_i,\ \alpha_i,\ u_i)$ are i.i.d over i, u_{it} is i.i.d over t, $\mathbb{E}\left(u_{it} \mid x_{it},\ \alpha_i\right) = 0$, $\mathbb{E}\left(u_{it}^4 \mid x_{it},\ \alpha_i\right) < \infty$.

Assumption 2.2. $\mathbb{E} \parallel x_{it} \parallel^{2+k} < \infty$ and $\mathbb{E}|u_{it}|^{2+k} < \infty$ for some $k > 0$.

Assumption 2.3: $\hat{\sigma}_u^2 = \sigma_u^2 + o_p\left(1\right)$, $\hat{\sigma}_\alpha^2 = \sigma_\alpha^2 + o_p\left(1\right)$.

Assumption 2.1 and 2.2 specify that the variables have finite fourth moments (so that conventional central limit theory applies) and that the error is conditionally homoskedastic given the regressors, which is used to simplify the asymptotic covariance expressions.

Theorem 1 *Under Assumption 2.1-2.3,*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{RE} - \beta \\ \hat{\beta}_{FE} - \beta \end{pmatrix} \xrightarrow{d} h + \xi \tag{2.10}$$

where

$$h = \begin{pmatrix} \sigma_1^2 V_1 \bar{X}'\bar{X}\delta \\ 0 \end{pmatrix}$$

and $\xi \sim N(0, V)$ *where*

$$V = \begin{pmatrix} V_1 & V_1 \\ V_1 & V_2 \end{pmatrix}$$

with

$$\begin{aligned} V_1 &= \left(\frac{X'\Omega^{-1}X}{n}\right)^{-1} \\ V_2 &= \sigma_u^2\left(\frac{X'QX}{n}\right)^{-1} \end{aligned}$$

Furthermore,

$$H_n \rightarrow \left(h + \xi\right)' B(h + \xi) \tag{2.11}$$

and

$$\sqrt{n}\left(\hat{\beta}_c-\beta\right)\xrightarrow{d}\Psi=G_2'\xi-\left(\frac{\tau}{(h+\xi)'\,B(h+\xi)}\right)_1 G'\left(h+\xi\right) \tag{2.12}$$

where $\bar{X}=(\bar{x}_1,\ldots,\bar{x}_q)$ *is* $n\times q$, $B=G\left(V_2-V_1\right)^{-1}G'$, $G=\begin{pmatrix}-I & I\end{pmatrix}'$, $G_2=\begin{pmatrix}0 & I\end{pmatrix}'$, *and* $(a)_1$ $=min[1,a]$.

Theorem 1 presents the joint asymptotic distribution of $\hat{\beta}_{RE}$ and $\hat{\beta}_{FE}$, the Hausman statistic, and $\hat{\beta}_c$ under the local exogeneity assumption. The joint asymptotic distribution of $\hat{\beta}_{RE}$ and $\hat{\beta}_{FE}$ is normal with a classic covariance matrix. $\hat{\beta}_{RE}$ has an asymptotic bias when $\delta\neq 0$ but not $\hat{\beta}_{FE}$. 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 $\hat{\beta}_c$ is a nonlinear function of the normal random vector and a function of the noncentrality parameter h.

2.3 Asymptotic Risk

The asymptotic risk of any sequence of estimators β_n of β is defined as

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

so long as the estimator has an asymptotic distribution

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

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

$$R\left(\beta\right)=\mathbb{E}\left(\psi'W\psi\right)=\mathrm{tr}\left(W\mathbb{E}\left(\psi\psi'\right)\right) \tag{2.13}$$

Define the largest eigenvalue of the matrix $W(V_2 - V_1)$

$$\lambda_1 = \lambda_{\max}\left(W\left(V_2 - V_1\right)\right),$$

and the ratio

$$d = \frac{\operatorname{tr}\left(W\left(V_2 - V_1\right)\right)}{\lambda_1}. \tag{2.14}$$

Notice that (2.14) satisfies $1 \leq d \leq q$. In the case $W = \left(V_2 - V_1\right)^{-1}$, $\lambda_1 = 1$ and we have the simplification $d = q$.

Theorem 2 *Under Assumption 2.1-2.3, if*

$$d > 2$$

and

$$0 < \tau \leq 2\left(d - 2\right) \tag{2.15}$$

then

$$\begin{aligned} R\left(\hat{\beta}_{FE}\right) &= tr\left(WV_2\right) \\ R\left(\hat{\beta}_c\right) &< R\left(\hat{\beta}_{FE}\right) - \frac{\tau\lambda_1\left[2\left(d-2\right)-\tau\right]}{\sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1\left(V_2 - V_1\right)^{-1}V_1\bar{X}'\bar{X}\delta + q}. \end{aligned} \tag{2.16}$$

Equation (2.16) shows that the asymptotic risk of $\hat{\beta}_c$ is strictly less than that of $\hat{\beta}_{FE}$, so long as τ satisfies the condition (2.15).

The assumption $d > 2$ is the critical condition needed to ensure that $\hat{\beta}_c$ can have smaller asymptotic risk than that of $\hat{\beta}_{FE}$. It is necessary in order for the right-hand-side of (2.15) to be positive, which is necessary for the existence of τ.

τ appears in the risk bound (2.16) as a quadratic expression, so there is a unique choice $\tau^* = d - 2$ which minimizes this bound. For practical implementation we recommend replacing the maximum eigenvalue λ_1 with the average $\frac{\operatorname{tr}(W(V_2 - V_1))}{q}$. Substituting into the expression for τ^* we obtain $\tau = q - 2$.

In the general weight matrix case, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q$ denote the ordered eigenvalues of $W(V_2 - V_1)$. $d > 2$ is equivalent to $\lambda_2 + \cdots + \lambda_q > \lambda_1$. This is violated only if λ_1 is much larger than the other eigenvalues. (2.15) is equivalent to $0 < \tau \leq 2\left(\sum_{i=1}^{q} \frac{\lambda_i}{\lambda_1} - 2\right)$.

Corollary 3 *$R\left(\hat{\beta}_c\right) - R\left(\hat{\beta}_{FE}\right) < 0$, for $d > 2$ and $0 < \tau \leq 2(d-2)$. When $W = (V_2 - V_1)^{-1}$, the condition simplify to $0 < \tau \leq 2(q-2)$ and $q > 2$, which is Stein's (1956) classic condition for shrinkage.*

As shown in Stein's (1956), $q > 2$ 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 4 *$R\left(\hat{\beta}_{RE}\right) = tr(WV_1) + \sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1WV_1\bar{X}'\bar{X}\delta$; $R\left(\hat{\beta}_{RE}\right) \leq R\left(\hat{\beta}_{FE}\right)$ when $\sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1WV_1\bar{X}'\bar{X}\delta \leq q$ and $R\left(\hat{\beta}_{RE}\right) > R\left(\hat{\beta}_{FE}\right)$ when otherwise.*

The result in Corollary 4 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 5 *$R\left(\hat{\beta}_c\right) - R\left(\hat{\beta}_{RE}\right) < 0$, for $q < \sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1WV_1\bar{X}'\bar{X}\delta$, $d > 2$, and $0 < \tau \leq 2(d-2)$.*

The result in Corollary 5 indicates that when endogeneity is strong, $d > 2$, $0 < \tau \leq 2(d-2)$, the combined estimator performs best among these three estimators.

Remark 6 *As in Wang et al. (2016), if the weight w is non-stochastic, we can obtain the asymptotic optimal w by minimizing*

$$w^2 R\left(\hat{\beta}_{RE}\right) + (1-w)^2 R\left(\hat{\beta}_{FE}\right) + 2w(1-w)\mathbb{E}\left(\left(\hat{\beta}_{RE} - \beta\right)' W \left(\hat{\beta}_{FE} - \beta\right)\right)$$

w^* is given by

$$w^* = tr\left(V_2 - V_1\right) / \left(tr\left(V_2 - V_1\right) + \sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1V_1\bar{X}'\bar{X}\delta\right)$$

which depends on the localizing parameter δ.

To understand the magnitude of the risk improvement, define

$$a_q = \frac{\mathrm{tr}\left(WV_2\right)}{\mathrm{tr}\left(W\left(V_2 - V_1\right)\right)}$$

and

$$c_q = \frac{\sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1\left(V_2 - V_1\right)^{-1}V_1\bar{X}'\bar{X}\delta}{\mathrm{tr}\left(W\left(V_2 - V_1\right)\right)}$$

In the special case when $W = \left(V_2 - V_1\right)^{-1}$

$$\begin{aligned} a_q &= \frac{\mathrm{tr}\left(\left(V_2 - V_1\right)^{-1}V_2\right)}{q}, \\ c_q &= \frac{\sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1\left(V_2 - V_1\right)^{-1}V_1\bar{X}'\bar{X}\delta}{q}. \end{aligned}$$

The a_q is a nonlinear function of the ratio $\frac{\sigma_\alpha}{\sigma_u} = \theta$ or $\rho^* = \theta/\left(1+\theta\right)$ $\left(0 \leq \rho^* \leq 1\right)$ which controls the strength of heterogeneity. c_q is a scalar measure of the strength of endogeneity δ (or ρ). c_q is increasing as the degree of endogeneity increases. From (2.16) we can calculate

$$\begin{aligned} r &= \frac{R\left(\hat{\beta}_c\right)}{R\left(\hat{\beta}_{FE}\right)} \\ &\leq 1 - \frac{\mathrm{tr}\left(W\left(V_2 - V_1\right)\right) - 2\lambda_1}{\mathrm{tr}\left(WV_2\right)} \times \frac{\mathrm{tr}\left(W\left(V_2 - V_1\right)\right) - 2\lambda_1}{\sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1\left(V_2 - V_1\right)^{-1}V_1\bar{X}'\bar{X}\delta + q} \\ &\simeq 1 - \frac{1}{a_q\left(c_q + 1\right)} \end{aligned}$$

The percentage reduction in asymptotic risk achieved by the combined estimator relative to the fixed effects estimator is approximately $0 \leq 1/a_qc_q \leq 1$. Note that $dr/d\delta > 0$, thus we expect the combined estimator to achieve large risk reductions when the degree of endogeneity is small.

2.4 Monte Carlo Study

The observations are generated by the progress

$$
\begin{aligned}
y_{it} &= x_{it}\beta + \alpha_i + u_{it} \\
\alpha_i &= \rho\sqrt{T}\bar{x}_i\frac{\iota}{q} + \sqrt{1-\rho^2}\ \epsilon_i
\end{aligned}
$$

$\{x_{it},\ u_{it}\}$ are i.i.d $N\left(0,\ I_{q+1}\right)$ across i, t. ϵ_i are i.i.d $N\left(0,\ 1\right)$ independent of $\{x_{it},\ u_{it}\}$. $\mathrm{Var}(\alpha_i) = 1$. The distribution are invariant to β so we set it to zero, $\beta = 0$.

We vary $n = \{20, 100\}$, $T = 3$, $q = \{4, 8\}$, and ρ on a 40-point grid on $[0, 0.975]$. ρ controls the degree of endogeneity, ranging in $(0, 1)$ ($\rho = 0$ is the case of exogenous regressors; large ρ is the case of strong endogeneity). We also set $\sqrt{\theta} = \frac{\sigma_\alpha}{\sigma_u} \in \left\{\frac{1}{4}, 1, 4\right\}$ so $\rho^* = \frac{\theta}{1+\theta} = \{.06,\ .05,\ .94\}$. ρ^* controls the degree of heterogeneity which is the temporal correlation between $\alpha_i + u_{it}$ and $\alpha_i + u_{it'}$.

Generated 100,000 samples on each calculated $\hat{\beta}_{RE}$, $\hat{\beta}_{FE}$, $\hat{\beta}_c$. To compare the estimators, calculate the median squared error (MSE) of each estimator and plot the relative median square error, that is

$$
\frac{median\left[\left(\hat{\beta}-\beta\right)'\left(\hat{\beta}-\beta\right)\right]}{median\left[\left(\hat{\beta}_{FE}-\beta\right)'\left(\hat{\beta}_{FE}-\beta\right)\right]}
$$

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

Figure 2.1 is the case $q = 8$, and Figure 2.2 is the case $q = 4$. The 6 plots in figure 2.2 look similar to the plots in Figure 2.1. By contrasting Figure 2.1 and 2.2, we can see that the improvement in the combined estimator over FE estimator with different values of ρ^* are greater in the cases of larger number of regressors.

Figure 2.1(a), 2.1(c), 2.1(e) are the cases $n = 20$, and Figure 2.1(b), 2.1(d), 2.1(f) are the cases $n = 100$. The region of dominance for the combined estimator over FE estimator is greater for small n.

Consider the case of eight endogenous regressors. Figure 2.1(a) and Figure 2.1(b) are the cases $\rho^* = 0.5$. For small ρ the RE estimator has lower MSE than the combined estimator, but the ranking is reversed for larger values of ρ. Figure 2.1(c) and Figure 2.1(d) are the cases $\rho^* = 0.06$. The MSE of the RE estimator is smaller than that of the combined estimator for all the values of ρ when $n = 20$ in Figure 2.1(c). The combined estimator has smaller MSE than that of the RE estimator when $n = 100$ in Figure 2.1(d). The MSE of the RE and combined estimators are uniformly smaller than that of the FE estimator. Figure 2.1(e) and Figure 2.1(f) are the cases $\rho^* = 0.94$. The FE and the combined estimators are near equivalents. RE has similar MSE to FE and combined estimators for small ρ, but the MSE of RE estimator increases dramatically after intermediate values of ρ.

Figure 2.3 is a 3D graph of the case $n = 20, T = 3$, $q = 8$. The improvements in combined estimator over FE estimator are greater for smaller heterogeneity ρ^*. For very small ρ^*, RE estimator tends to be better than both FE and combined estimators. 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.

In summary, the simulation results provide strong finite sample confirmation of Theorem 2.

Next, we would like to compare the performance of following combined estimators:

1. The proposed combined estimator (Stein Estimator):

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

2. The proposed combined estimator with optimal weights:

$$\hat{\beta}_{Opt} = w_{Opt}\hat{\beta}_{RE} + (1 - w_{Opt})\,\hat{\beta}_{FE}$$

where

$$w_{Opt} = \operatorname{tr}(V_{FE} - V_{RE}) \,/\, \left(\operatorname{tr}(V_{FE} - V_{RE}) + \sigma_1^{-4}\delta'\bar{X}'\bar{X}V_{RE}V_{RE}\bar{X}'\bar{X}\delta\right)$$

3. The combined estimator by inverse-variance weighting method:

$$\hat{\beta}_{Var} = w_{RE}\hat{\beta}_{RE} + w_{FE}\hat{\beta}_{FE}$$

where

$$\begin{aligned} w_{RE} &= \frac{\hat{V}_{RE}^{-1}}{\hat{V}_{FE}^{-1} + \hat{V}_{RE}^{-1}}, \\ w_{FE} &= \frac{\hat{V}_{FE}^{-1}}{\hat{V}_{FE}^{-1} + \hat{V}_{RE}^{-1}}. \end{aligned}$$

Observations with larger disturbance variance contain less information than observations with smaller disturbance variance. The inverse-variance weighted combination method obtains weights inversely proportional to the respective variance. Therefore, the smaller variance estimates gets the larger weight.

4. The combined estimator by Leave one out (LOO) method:

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

We use the above DGP, and generate 100,000 samples for each configuration, and on each calculated $\hat{\beta}_{RE}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{LOO}$, $\hat{\beta}_{Var}$, $\hat{\beta}_{Stein}$ and $\hat{\beta}_{Opt}$. To compare the estimators, we calculated the MSE of each estimator and plot the relative MSE.

Figure 2.4 is the case $q = 8$, and Figure 2.5 is the case $q = 4$. The 6 plots in figure 2.4 look similar to the plots in Figure 2.5. Figure 2.4(a), 2.4(c), 2.4(e), 2.5(a), 2.5(c), 2.5(e) are the cases $n = 20$, and Figure 2.4(b), 2.4(d), 2.4(f), 2.5(b), 2.5(d), 2.5(f) are the cases $n = 100$. Consider the case of eight endogenous regressors. The LOO estimator and the FE estimator are near equivalent. The combined estimator dominates the FE estimator for all the values of ρ and ρ^*. For small n, the optimal weights combined (Opt) estimator has quite close MSE to that of the RE estimator regardless of the degree of heterogeneity. Figure 2.4(a) and Figure 2.4(b) are the cases $\rho^* = 0.5$. For small ρ the Opt estimators has the smallest MSE, but for larger values of ρ, the combined estimator performs the best. Figure 2.4(c) and Figure 2.4(d) are the cases $\rho^* = 0.06$. The MSE of the RE and Opt estimators are uniformly smaller than the other estimators, except when both n and ρ^* are large (the inverse-variance weighted combined (IVW) estimator has the smallest MSE for this situation). Figure 2.4(e) and Figure 2.4(f) are the cases $\rho^* = 0.94$. All the estimators have similar MSE for small values of ρ. But the MSE of RE, Opt and IVW estimators increase dramatically after intermediate values of ρ.

In summary, for small ρ, the optimal weights combined estimator performs the best. It is also recommended to use for moderate ρ, small and moderate ρ^*, or large ρ, small ρ^*, small n. For large ρ, small ρ^* and large n, the inverse-variance weighted combined estimator performs the best. The combined estimator is recommended for moderate and large values of ρ and ρ^*.

2.5 Empirical Results

In this section, we use a panel data for the 48 contiguous U.S. states in each year between 1970 and 1986. To these data, we fit Cobb-Douglas and translog production function to revisit the relationship between public infrastructure and private economic performance. Details on this data set can be found in Munnell (1990). A large body of research has explored the public-sector capital and private economic performance relationship. Some theories support a positive and significant impact of public capital stock on private sector output [see, e.g., Munnell (1990)]. However, many studies believe that the public capital had negative and significant effects on private productivity [see, e.g., Evans and Karras (1994)]. In addition, another type of findings is that the contribution of the public infrastructure does not have quantitatively significant spillover effects on private sector across states [see, e.g., Holtz-Eakin (1994), Baltagi and Pinnoi (1995)].

The following panel data model is estimated:

$$\begin{aligned} \log(Y_{it}) &= \beta_0 + \beta_1 \log(KG_{i,t-2}) + \beta_2 \log(KPR_{i,t-2}) \\ &\quad + \beta_3 \log(L_{i,t-2}) + \beta_4 UNEM_{i,t-2} + \alpha_i + u_{it} \end{aligned}$$

where $i = 1, \ldots, 48, t = 1, \ldots, 17$, Y_{it} denotes the gross private nonagricultural product of state i in period t, KG denotes public capital which includes highways and streets, water and sewer facilities and other public buildings and structures, KPR is the private capital stock estimated from the Bureau of Economics Analysis, L is labor input measured as employment in nonagricultural payrolls, and $UNEM$ stands for the states unemployment rate, included to control for business cycle effects as in the previous literature. $X_{i,t-2}$ is used in the regression to take into account the time delay effects, since it takes time for the investments to be fully utilized. Fixed effects for each state will pick up state specific factors such as natural resources, the quality of public infrastructure, physical characteristics of a state. Furthermore, the spillover effects of infrastructure improvement from other states could also be included in the state-specific effects.

In order to obtain the mean square errors (MSE) and the standard errors for these estimates, we bootstrap the data 10000 times by resampling across individuals and keep the time series structure for each individual unchanged. We obtain estimates of the average elasticities and coefficients for each bootstrap data, based on which we can calculate the bootstrap MSE and the standard errors for the above estimates. The MSE for FE, RE and combined estimators are 4.9935e-04, 0.0141 and 3.6275e-04, respectively.

Table 2.1 suggests that the estimated coefficients of gross private nonagricultural product for FE estimator with respect to KG, KPR, L and $UNEMP$ are -0.0261, 0.2920, 0.7682 and -0.0053, respectively; The estimated coefficients for RE estimator with respect to KG, KPR, L and $UNEMP$ are 0.0044, 0.3105, 0.7297 and -0.0062, respectively; The estimated coefficients for the combined estimator with respect to KG, KPR, L and $UNEMP$ are -0.0167, 0.2977 and 0.7563, and -0.0056, respectively. Both FE and combined estimators report that the public capital is counter productive and insignificantly in the state private production. In contrast, the RE estimator finds that the public capital is productive and insignificantly. Note that the Hausman statistic is 11.7181. Thus, the null hypothesis is easily rejected at the one percent level of significance. This indicates that there exists a huge problem of endogeneity. In this circumstance, the FE estimator solves the problem. Thus, it would be more appropriate to treat α as fixed. And the RE estimator seems overwhelming that public capital has a positive impact on private sector output. The combined estimator

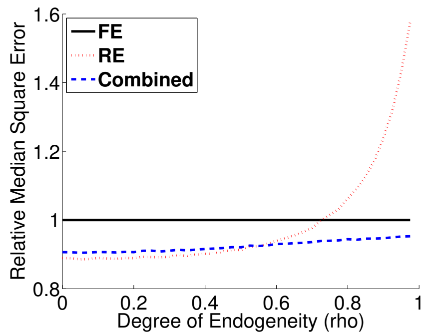
	$\hat{\beta}_{KG}$	$\hat{\beta}_{KPR}$	$\hat{\beta}_{L}$	$\hat{\beta}_{UNEMP}$
FE	-0.0261	0.2920	0.7682	-0.0053
	(0.0210)	(0.0227)	(0.0239)	(0.0008)
RE	0.0044	0.3105	0.7297	-0.0062
	(0.0392)	(0.0314)	(0.0407)	(0.0016)
Combined	-0.0167	0.2977	0.7563	-0.0056
	(0.0219)	(0.0185)	(0.0196)	(0.0007)

Table 2.1: Economics of Private Sector Output Estimates for 48 U.S. states, 1970-1986 (standard errors in parentheses)

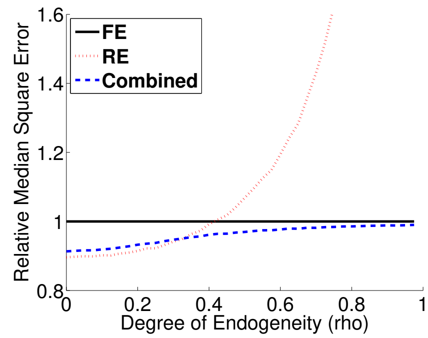
result is consistent with FE estimator. Our empirical analysis is in agreement with the findings of Holtz-Eakin (1994) and Baltagi and Pinnoi (1995) that there is no quantitatively important spillover effects across states. The estimated $\rho^* = 0.8193$, which may also explain why the combined estimator result is closer to FE estimator result. A careful weighting of the evidence available from state-level data indicates that the best estimate of the elasticity of private output or productivity with respect to state government capital is essentially zero. As a result, the combined estimator is more reliable under this scenario. However, one should not disregard the importance of public infrastructure based on exclusively on this evidence. More insight can be gained by examining the results of disaggregating the model.

2.6 Conclusion

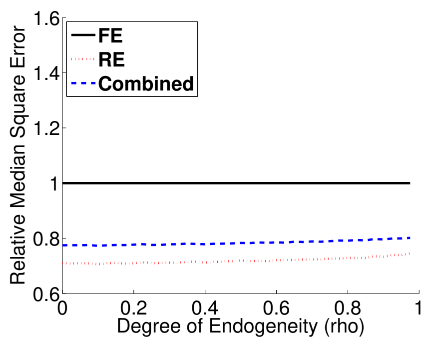
This chapter provides a combined fixed effects and random effects estimator, with the weights depending on the Hausman statistic. We show that the combined estimator has strictly smaller asymptotic risk than the fixed effects estimator. The combined estimator also has smaller asymptotic risk compared to the random effetcs estimator unless the endogeneity is very weak. Our simulation experiment finds that the results also hold for small samples. The magnitude of efficiency of the combined estimator over random effects and fixed effects estimators varies with respect to the degree of endogeneity and heterogeneity. In the simulation, we also discuss four combined estimators from combining the fixed effects and random effects estimators in the panel data model. The results offer a typology of data set characteristics to help researchers choose a preferred combined estimator. We use the combined estimator over fixed effetcs and random effects to revisit the relationship between the public capital stock and the private sector output. We confirm findings of Holtz-Eakin (1994) and Baltagi and Pinnoi (1995) that the contribution of the public infrastructure does not have quantitatively significant spillover effects on private sector across states. In this case, the combined estimator gives smaller MSE and is a more reliable estimator.



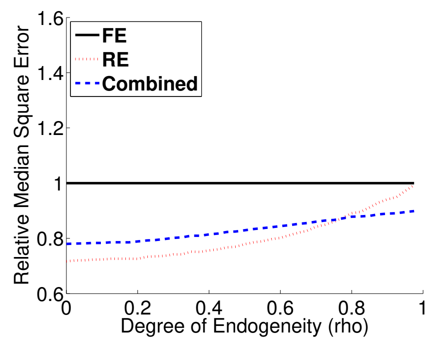
(a) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .5$



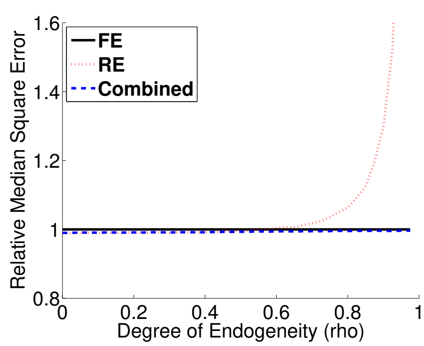
(b) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .5$



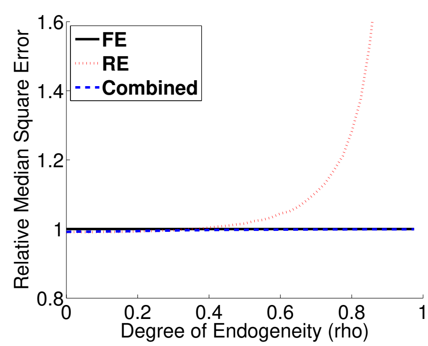
(c) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .06$



(d) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .06$

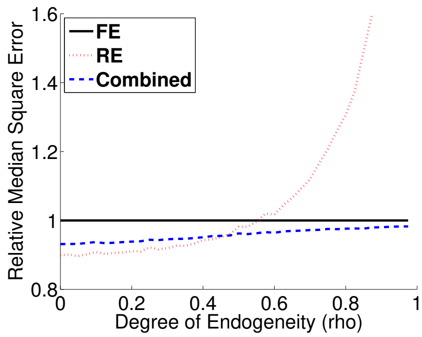


(e) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .94$

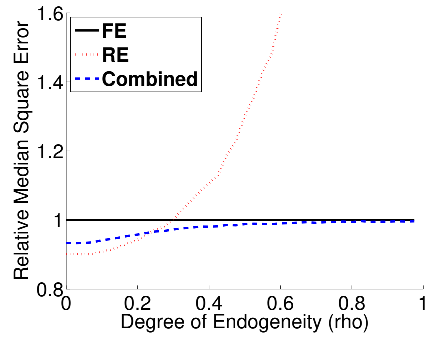


(f) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .94$

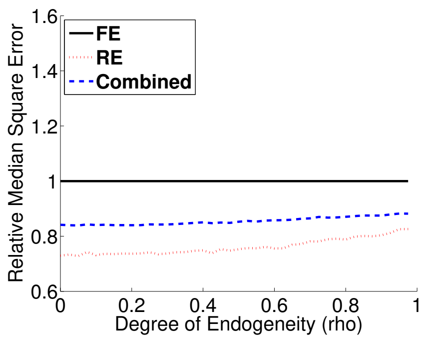
Figure 2.1: Relative MSE of FE, RE and Combined Estimators, $n = \{20, 100\}$, $T = 3$, $q = 8$, $\rho^* = \{.5, .06, .94\}$.



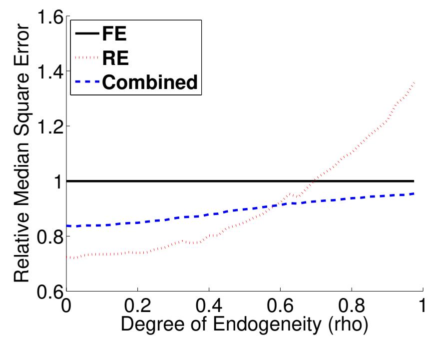
(a) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .5$



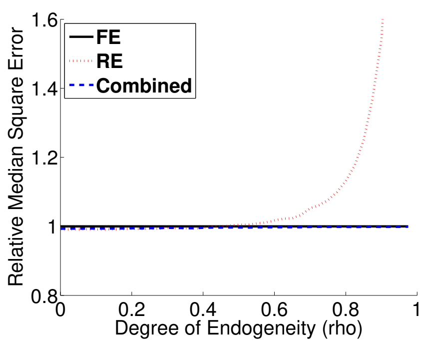
(b) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .5$



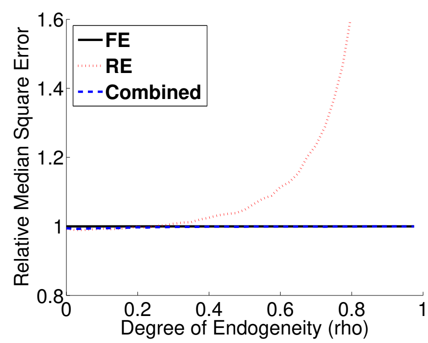
(c) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .06$



(d) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .06$



(e) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .94$



(f) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .94$

Figure 2.2: Relative MSE of FE, RE and Combined Estimators, $n = \{20, 100\}$, $T = 3$, $q = 4$, $\rho^* = \{.5, .06, .94\}$.

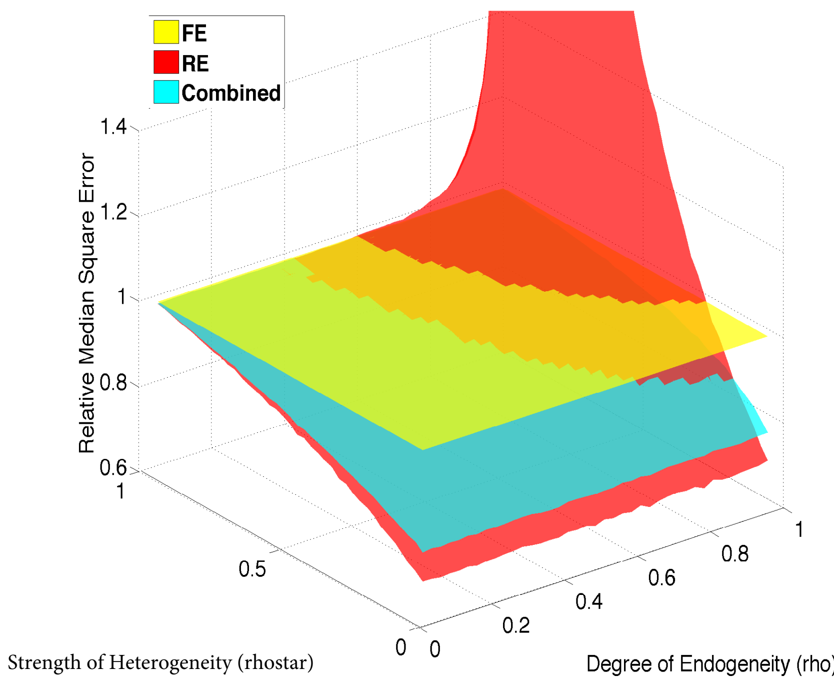
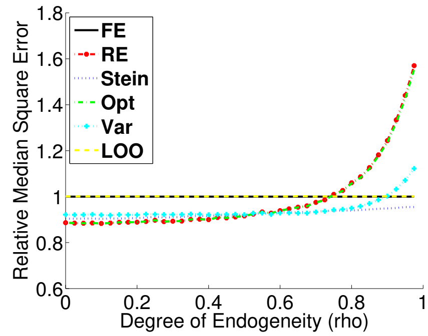
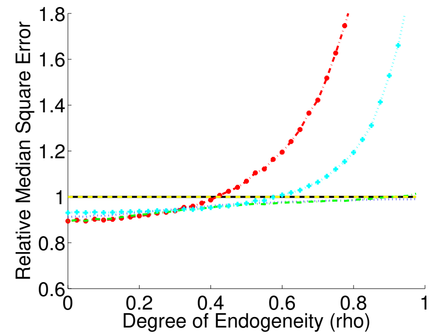


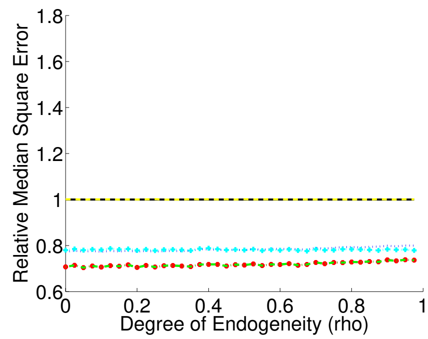
Figure 2.3: Relative MSE of FE, RE and Combined Estimators 3D Graph, $n = 20$, $T = 3$, $q = 8$



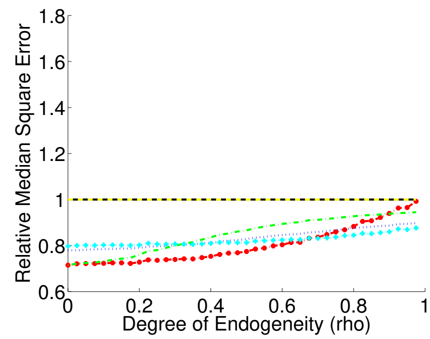
(a) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .5$



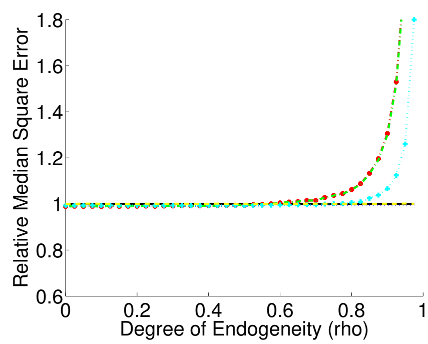
(b) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .5$



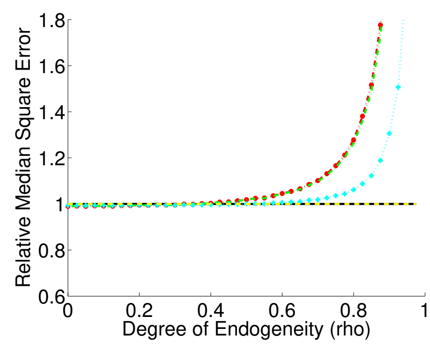
(c) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .06$



(d) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .06$

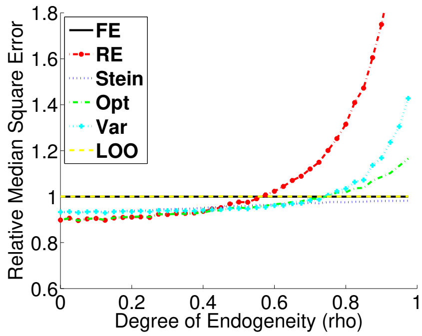


(e) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .94$

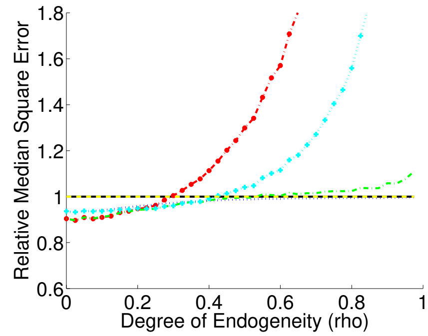


(f) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .94$

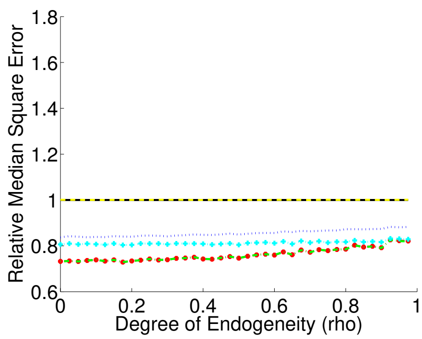
Figure 2.4: Relative MSE of Four Combined Estimators, $n = \{20, 100\}$, $T = 3$, $q = 8$, $\rho^* = \{.5, .06, .94\}$



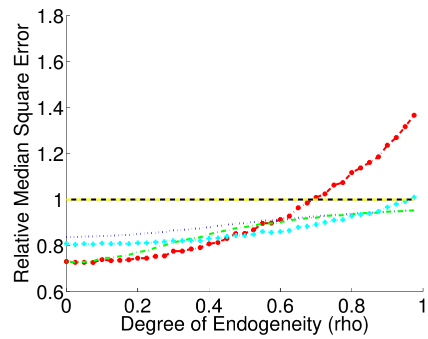
(a) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .5$



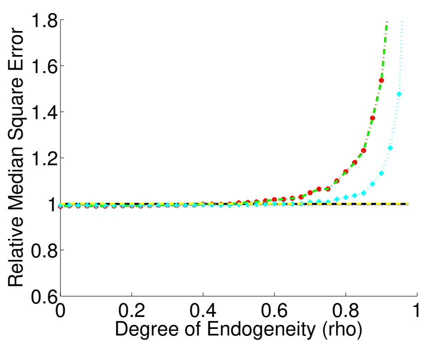
(b) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .5$



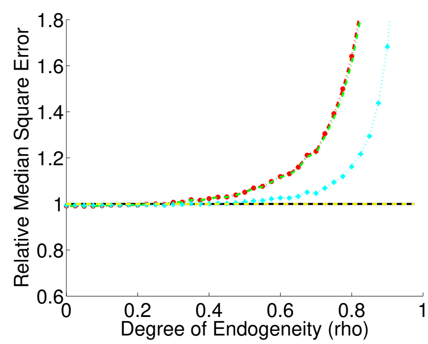
(c) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .06$



(d) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .06$



(e) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .94$



(f) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .94$

Figure 2.5: Relative MSE of Four Combined Estimators, $n = \{20, 100\}$, $T = 3$, $q = 4$, $\rho^* = \{.5, .06, .94\}$.

Chapter 3

A Combined Semi-parametric Estimator for Panel Data Model

3.1 Introduction

Semi-parametric modelling (SP) is, as its name suggests, a hybrid of the parametric and nonparametric approaches. The flexibility of semi-parametric modelling has made it a widely accepted statistical technique. In this chapter, we adopt a semi-parametric approach to modelling a general partially linear panel data model. There is a rich literature on semi-parametric estimation of panel data, see Horowitz and Markatou (1996), Ullah and Roy (1998) and Li and Hisao (1998), to mention only a few. Li and Ullah(1998) discusses the partially linear panel data model with random effects. Balagi and Li (2002), Su and Ullah (2006) consider the estimation of the partially linear panel data models with fixed effects. If the individual effects are correlated with the other regressors in the model, the fixed effect model is consistent and the random effects model is inconsistent. A random effects approach involves an assumption that is rarely palatable, that the individual effects are uncorrelated with the regressors. When this assumption fails, the random effects estimator is biased. On the other hand, if the individual effects are not correlated with the other regressors in the model, both random and fixed effects estimators are consistent and random effects estimator is efficient. Therefore, there is a trade-off between inefficient fixed effects estimation and biased random effects estimation.

Under this scenario, we propose a combined estimator, which is a weighted combination of the SP fixed effects, and SP random effects estimators with weights depending on Hausman test (1978) statistic. The parametric combined estimator can be viewed as a special case of the semi-parametric combined estimator. The asymptotic distribution and risk of the combined estimator are derived using a local asymptotic framework. We show that under certain conditions, the combined estimator has strictly smaller risk than SP fixed effects estimator. Further, in the Monte Carlo study we show that the combined estimator performs better than the fixed effects estimator, as well as compared to the random effects estimator except when the degree of endogeneity or heterogeneity is very small. Finally, We examine the role for public sector in affecting private sector economic performance using a panel data for the 48 contiguous U.S. states over the period 1970–1986.

The rest of this chapter is organized as follows. Section 2 presents the model and estimators. Section 3 presents the asymptotic distribution of the combined estimator. Section 4 derives the asymptotic risk of the combined estimator. Monte Carlo simulation is provided in Section 5. Empirical example is given in Section 6. Section 7 concludes.

3.2 The Model and Estimators

3.2.1 The Fixed Effects Estimator

Consider the following semi-parametric regression model with fixed effects (FE):

$$y_{it} = x_{it}\beta + m(z_{it}) + \alpha_i + u_{it}, \quad i = 1, 2, \ldots, n, \; t = 1, 2, \ldots, T \tag{3.1}$$

where x_{it} and z_{it} are of dimensions $1 \times q$ and $1 \times p$, respectively, and β is a $q \times 1$ vector of unknown parameters, $m(\cdot)$ is an unknown smooth function. $\alpha_i's$ are fixed effects and $u_{it}'s$ are the random disturbances. We consider the usual panel data case of large n and small T. Hence all the asymptotics are for $n \to \infty$ for a fixed T. In matrix notation, (3.1) can be written as

$$y = X\beta + m(Z) + D\alpha + u$$

$D = I_n \otimes \iota_T$ is $nT \times n$, α is $n \times 1$, and $u \sim$i.i.d.$(0, \; \sigma_u^2 I_{nT})$.

A local linear approximation of the model (3.1) can be written as

$$\begin{aligned} y_{it} &\approx x_{it}\beta + m(z) + (z_{it} - z)\dot{m}(z) + \alpha_i + u_{it} \\ &= x_{it}\beta + Z_{it}(z)\delta(z) + \alpha_i + u_{it} \end{aligned} \tag{3.2}$$

where $Z_{it}(z) = (1, \; (z_{it} - z))$, $\delta(z) = \left(m(z), (\dot{m}(z))'\right)'$, $\dot{m}(\cdot)$ is the first derivative of $m(\cdot)$. In a vector form, we can write

$$y = X\beta + Z(z)\delta(z) + D\alpha + u$$

where $y = (y_{11}, \ldots, y_{1T}, y_{21}, \ldots, y_{nT})'$, $X = (x_{11}, \ldots, x_{1T}, \ldots, x_{n1}, \ldots, x_{nT})$ and $Z(z) = (Z_{11}'(z), \ldots, Z_{1T}'(z), Z_{21}'(z), \ldots, Z_{nT}(z))'$.

Let K denote a kernel function on $\mathbb{R}^p$ and $H =$diag$(h_1, \ldots, h_p)$, a matrix of bandwidth sequences. Set $K_H(z) = |H|^{-1}K(H^{-1}z)$, where $|H|$ is the determinant of H. Further denote that

$\mathbf{K}(H^{-1}z)=$diag$(K(H^{-1}(z_{11}-z)),\ldots,K(H^{-1}(z_{1T}-z)),\ldots,K(H^{-1}(z_{nT}-z)))$.

Su and Ullah (2006) estimate δ by minimizing the following criterion function

$$(y-X\beta-D\alpha-Z(z)\delta)'\mathbf{K}_H(z)(y-X\beta-D\alpha-Z(z)\delta)$$

Define the smoothing operator by $S(z)=\left[Z(z)'\mathbf{K}_H(z)Z(z)\right]^{-1}Z(z)'\mathbf{K}_H(z)$, then

$$\begin{aligned}\delta(z) &= S(z)(y-D\alpha-X\beta)\\ m(Z) &= S(y-X\beta-D\alpha)\end{aligned}\tag{3.3}$$

where $S=(s_{11},\ldots,s_{1T},s_{21},\ldots,s_{nT})$, and $s_{it}=s(z_{it})$. In particular, $m(z)$ is given by

$$m(z)=s(z)'(y-X\beta-D\alpha)$$

where $s(z)'=eS(z)$, and $e=(1,0,\ldots,0)$ is a $1\times(p+1)$ vector. The parameter β is then estimated by the profile likelihood method by minimizing

$$(y-X\beta-D\alpha-m(Z))'(y-X\beta-D\alpha-m(Z))\tag{3.4}$$

where $m(Z)=(m(z_{11}),\ldots,m(z_{1T}),\ldots,m(z_{nT}))'$. Plugging (3.3) into (3.4) and using the formula for partitioned regression, we can obtain

$$\begin{aligned}\hat{\beta}_{SP,FE} &= (X^{*\prime}Q^*X^*)^{-1}X^{*\prime}Q^*y^*, \qquad (3.5)\\ \hat{\alpha}_{SP,FE} &= (D^{*\prime}D^*)^{-1}D^{*\prime}y^*\end{aligned}$$

where $D^*=(I_{nT}-S)D$, $y^*=(I_{nT}-S)y$, $X^*=(I_{nT}-S)X$ and $Q^*=I_{nT}-D^*(D^{*\prime}D^*)^{-1}D^{*\prime}$. The profile likelihood estimator for $m(Z)$ is given by

$$\hat{m}(Z)=S\left(y-X\hat{\beta}_{SP,FE}-D\hat{\alpha}_{SP,FE}\right)$$

In particular, the profile likelihood estimator for $m(z)$ is

$$\hat{m}(z)=s(z)'\left(y-X\hat{\beta}_{SP,FE}-D\hat{\alpha}_{SP,FE}\right)$$

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

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

where $V_{SP,FE}=\sigma_u^2\left(\text{plim}\frac{X^{*\prime}Q^{*}X^{*}}{n}\right)^{-1}$.

3.2.2 The Random Effects Estimator

Now, we present the semi-parametric regression model with random effects (RE):

$$y_{it}=x_{it}\beta+m(z_{it})+\alpha_i+u_{it},\quad i=1,2,\ldots,n,\ t=1,2,\ldots,T \tag{3.6}$$

in matrix notation, (3.6) can be written as

$$y=X\beta+m(Z)+v \tag{3.7}$$

the error v in (3.7) follows an one-way error components structure:

$$v=D\alpha+u$$

where $\alpha\sim\left(0,\ \sigma_\alpha^2 I_n\right)$, $u\sim\left(0,\ \sigma_u^2 I_{nT}\right)$, $v\sim\left(0,\ \Omega\right)$. The variance-covariance matrix of v is given by $\Omega=\sigma_u^2 I_{nT}+\sigma_\alpha^2 DD'=\sigma_1^2 P+\sigma_u^2 Q$, where $Q=I_{nT}-P$, $P=D\left(D'D\right)^{-1}D'$, $\sigma_1^2=T\sigma_\alpha^2+\sigma_u^2$. The inverse matrix of Ω is given by $\Omega^{-1}=\frac{1}{\sigma_1^2}P+\frac{1}{\sigma_u^2}Q$. By taking expectation of (3.6) conditional on z_{it}, obtain

$$E\left(y_{it}\mid z_{it}\right)=E\left(x_{it}\mid z_{it}\right)\beta+m(z_{it}) \tag{3.8}$$

Then one can eliminate the unknown function $m\left(\cdot\right)$ by subtracting (3.8) from (3.6) to get

$$y_{it}-E\left(y_{it}\mid z_{it}\right)=\left(x_{it}-E\left(x_{it}\mid z_{it}\right)\right)\beta+v_{it}$$

In vector notation,

$$\begin{aligned} y-E\left(y\mid Z\right) &= \left(X-E\left(X\mid Z\right)\right)\beta+v \\ y^{*} &= X^{*}\beta+v \end{aligned} \tag{3.9}$$

where the conditional expectations $E(y \mid Z)$ and $E(X \mid Z)$ can be estimated by local linear least squares estimators (LLLS). Therefore, y^* and X^* in (3.9) are

$$\begin{aligned} y^* &= (I_{nT} - S)\, y, \\ X^* &= (I_{nT} - S)\, X. \end{aligned}$$

The feasible estimator of $\hat{\Omega}$ of Ω can be obtained by running the OLS regression y^* on X^*. Define

$$\hat{\sigma}_1^2 = \frac{T}{n} \sum_{i=1}^{n} \bar{\hat{v}}_i^2, \quad \hat{\sigma}_u^2 = \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\hat{v}_{it} - \bar{\hat{v}}_i \right)^2$$

where $\hat{v} = y^* - X^* \hat{\beta}_{OLS}$ is the OLS residual and $\hat{\beta}_{OLS} = (X^{*\prime} X^*)^{-1} X^{*\prime} y^*$. Noting that $\hat{\sigma}_1^2$ and $\hat{\sigma}_u^2$ estimate $\hat{\sigma}_\alpha^2 = \frac{1}{T}\left(\hat{\sigma}_1^2 - \hat{\sigma}_u^2\right)$. With these estimates, one can obtain the generalized least squares (GLS) of β based on (3.9) is

$$\hat{\beta}_{SP,RE} = \left(X^{*\prime} \hat{\Omega}^{-1} X^* \right)^{-1} X^{*\prime} \hat{\Omega}^{-1} y^* \tag{3.10}$$

and $\hat{\beta}_{SP,RE}$ has an asymptotic distribution as

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

where $V_{SP,RE} = \left(\operatorname{plim} \frac{X^{*\prime} \Omega^{-1} X^*}{n} \right)^{-1}$.

3.2.3 The Combined Estimator

See Hausman (1978), under the RE specification, the RE estimator is the asymptotically efficient estimator while the FE estimator is unbiased and consistent but not efficient. If $E(\alpha_i x_{it}) \neq 0$, the RE estimator is biased and inconsistent while the FE estimator is not affected. Motivated by this observation, we would like to see if combination of $\hat{\beta}_{SP,RE}$ and $\hat{\beta}_{SP,FE}$ can result in an improved estimation. We propose the following combined estimator of β, which is a weighted combination of $\hat{\beta}_{SP,FE}$ and $\hat{\beta}_{SP,RE}$ with weights depending on Hausman statistic:

$$\hat{\beta}_{SP,c} = w\hat{\beta}_{SP,RE} + (1-w)\hat{\beta}_{SP,FE} \tag{3.11}$$

where

$$\begin{aligned} w &= \begin{cases} \frac{\tau}{H_n} & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases} \\ H_n &= (\hat{\beta}_{SP,FE} - \hat{\beta}_{SP,RE})'\hat{V}(\hat{\beta}_{SP,FE} - \hat{\beta}_{SP,RE})^{-1}(\hat{\beta}_{SP,FE} - \hat{\beta}_{SP,RE}) \end{aligned}$$

where τ is a shrinkage parameter. The degree of shrinkage depends on the ratio τ/H_n. When $H_n < \tau$ then $\hat{\beta}_{SP,c} = \hat{\beta}_{SP,RE}$, When $H_n \geq \tau$ then $\hat{\beta}_{SP,c}$ is a weighted average of $\hat{\beta}_{SP,RE}$ and $\hat{\beta}_{SP,FE}$, with more weight on $\hat{\beta}_{SP,RE}$ when τ/H_n is large. Alternatively, it can be written as a positive-part James-Stein Estimator

$$\hat{\beta}_{SP,c} = \hat{\beta}_{SP,RE} + \left(1 - \frac{\tau}{H_n}\right)^{+} (\hat{\beta}_{SP,FE} - \hat{\beta}_{SP,RE})$$

where $(b)^{+} = b$ if $b > 0$, and 0 if $b \leq 0$.

3.3 Asymptotic Distribution

Write α_i as a linear function of $\bar{x}_i = \sum_t x_{it}/T$

$$\begin{aligned} \alpha_i &= \bar{x}_i\rho + \epsilon_i, \\ \mathbb{E}\left(\bar{x}_i\epsilon_i\right) &= 0 \end{aligned} \tag{3.12}$$

The variable x_{it} is exogenous if α_i and x_{it} are uncorrelated, or equivalently that the coefficient ρ is zero. We use the local asymptotic approach. For fixed T, ρ is local to zero

$$\rho = \frac{1}{\sqrt{n}}\delta \tag{3.13}$$

δ is a $q \times 1$ localizing parameter, which is the degree of correlation between $\bar{x}_i$ and α_i. When $\delta = 0$, x_{it} are exogenous. When $\delta \neq 0$, x_{it} are endogenous. ρ (or δ) controls the degree of endogeneity.

Now, we make the following assumptions:

Assumption 3.1. $(\alpha_i, u_i, x_i, z_i), i = 1, \ldots, n$, are i.i.d. over i, u_{it} is i.i.d over t, where $u_i = (u_{i1}, \ldots, u_{iT})'$ and x_i and z_i are similarly defined. $\mathbb{E}(u_{it} \mid x_{it},\ \alpha_i) = 0$ and $\mathbb{E}(u_{it}^4 \mid x_{it},\ \alpha_i) < \infty$.

Assumption 3.2. $\mathbb{E} \parallel x_{it} \parallel^{2+k} < \infty$ and $\mathbb{E}|u_{it}|^{2+k} < \infty$ for some $k > 0$.

Assumption 3.3. Let $x_{it}^* \equiv x_{it} - \mathbb{E}(x_{it} \mid z_{it})$, $\sum_t \mathbb{E}\left\{x_{it}^* \left[x_{it}^* - \sum_s x_{is}^*/T\right]'\right\}$ is positive definite.

Assumption 3.4. The kernel function $K(\cdot)$ is a continuous density with compact support on $\mathbb{R}^q$. All odd order moments of K vanish.

Assumption 3.5. As $n \to \infty$, $\parallel H \parallel \to 0$, $n|H|^2 \to \infty$, $\parallel H \parallel^4 |H|^{-1} \to 0$ and $n|H| \parallel H \parallel^4 \to c \in [0, \infty)$, where $|H|$ is the determinant of H.

Assumption 3.6. $\mathbb{E} \parallel x_{it} \parallel^4 < \infty$; $\sigma_u^2 \left(\operatorname{plim} \frac{X^{*\prime} Q^* X^*}{n}\right)^{-1} = V_2$, $\operatorname{plim}\left(\frac{X^{*\prime}\Omega^{-1}X^*}{n}\right)^{-1} = V_1$ and $\sigma_u^2 \left(\operatorname{plim} \frac{(X^{*\prime}Q^*X^*)^{-1} X^{*\prime} Q^* \Omega^{-1} X^* \left(X^{*\prime}\Omega^{-1}X^*\right)^{-1}}{n}\right) = V_{21}$ as $n \to \infty$.

Assumption 3.7. $\hat{\sigma}_u^2 = \sigma_u^2 + o_p(1)$, $\hat{\sigma}_\alpha^2 = \sigma_\alpha^2 + o_p(1)$.

Assumption 3.1 and 3.2 are standard in the literature. Assumption 3.3 rules out time-invariant terms x_{it}. Assumption 3.4 are requirements that K is compactly supported. Assumption 3.5 is easily satisfied by considering $H = \operatorname{diag}(h_1, \ldots, h_p)$ with $h_i \propto n^{-1/(4+p)}$, for $p < 4$. When $p \geq 4$, higher order local polynomial can be used to achieve bias reduction. Nevertheless, due to the "the curse of dimensionality", we do not expect large p in practice.

Theorem 7 *Under Assumptions 3.1-3.7,*

$$\sqrt{n}\begin{pmatrix}\hat{\beta}_{SP,RE}-\beta\\ \hat{\beta}_{SP,FE}-\beta\end{pmatrix}\to_d h+\xi \tag{3.14}$$

where

$$h=\begin{pmatrix}\sigma_1^{-2}V_1\bar{X}^{*\prime}\bar{X}\delta\\ 0\end{pmatrix}$$

and $\xi\sim N(0,V)$ *with*

$$V=\begin{pmatrix}V_1 & V_{21}'\\ V_{21} & V_2\end{pmatrix}$$

Furthermore,

$$H_n\to(h+\xi)'B(h+\xi) \tag{3.15}$$

and

$$\sqrt{n}\left(\hat{\beta}_{SP,c}-\beta\right)\xrightarrow{d}\Psi=G_2'\xi-\left(\frac{\tau}{(h+\xi)'B(h+\xi)}\right)_1 G'(h+\xi) \tag{3.16}$$

where $\bar{X}=(\bar{x}_1,\ldots,\bar{x}_q)$, $\bar{X}^*=\left(\bar{x}_1^*,\ldots,\bar{x}_q^*\right)$. $B=G\left(V_1+V_2-(V_{21}+V_{21}')\right)^{-1}G'$, $G=\begin{pmatrix}-I & I\end{pmatrix}'$, $G_2=\begin{pmatrix}0 & I\end{pmatrix}'$, *and* $(a)_1=min[1,a]$.

Theorem 7 presents the joint asymptotic distribution of $\hat{\beta}_{SP,RE}$ and $\hat{\beta}_{SP,FE}$, the Hausman statistic, and $\hat{\beta}_{SP,c}$ under the local exogeneity assumption. The joint asymptotic distribution of $\hat{\beta}_{SP,RE}$ and $\hat{\beta}_{SP,FE}$ is normal. $\hat{\beta}_{SP,RE}$ has an asymptotic bias when $\delta\neq 0$ but not $\hat{\beta}_{SP,FE}$. 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 $\hat{\beta}_{SP,c}$ is a nonlinear function of the normal random vector and a function of the noncentrality parameter h.

3.4 Asymptotic Risk

The asymptotic risk of any sequence of estimators β_n of β is defined as

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

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

$$R(\beta_n) = \mathbb{E}(\psi' W \psi) = \operatorname{tr}(W\mathbb{E}(\psi\psi')) \tag{3.17}$$

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

$$\begin{aligned} \lambda_1 &= \lambda_{\max}\left(\frac{A+A'}{2}\right) \\ \lambda_1^* &= \lambda_{\max}\left(\frac{A^*+A^{*\prime}}{2}\right) \end{aligned}$$

where

$$\begin{aligned} A &= (V_1 + V_2 - (V_{21} + V_{21}'))^{\frac{1}{2}} W (V_2 - V_{21}) (V_1 + V_2 - (V_{21} + V_{21}'))^{-\frac{1}{2}}, \\ A^* &= (V_1 + V_2 - (V_{21} + V_{21}'))^{\frac{1}{2}} W (V_1 + V_2 - (V_{21} + V_{21}'))^{\frac{1}{2}}. \end{aligned}$$

Let

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

Theorem 8 *Under Assumptions 3.1-3.7, if*

$$d > 2 \tag{3.18}$$

and

$$0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}, \tag{3.19}$$

then $R\left(\hat{\beta}_{SP,c}\right) = tr[WE\left(\Psi\Psi'\right)]$,

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

and

$$R\left(\hat{\beta}_{SP,c}\right) < R\left(\hat{\beta}_{SP,FE}\right) - \frac{\tau\left(2\lambda_1\left(d-2\right)-\lambda_1^*\tau\right)}{\sigma_1^{-4}\delta'\bar{X}'\bar{X}^*V_1\left(V_1+V_2-\left(V_{21}+V_{21}'\right)\right)^{-1}V_1\bar{X}^{*\prime}\bar{X}\delta+q}. \tag{3.20}$$

Equation (3.20) shows that the asymptotic risk of $\hat{\beta}_{SP,c}$ is strictly less than that of $\hat{\beta}_{SP,FE}$, so long as τ satisfies the condition (3.19). The assumption $d > 2$ is the critical condition needed to ensure that $\hat{\beta}_{SP,c}$ can have smaller asymptotic risk than that of $\hat{\beta}_{SP,FE}$. It is necessary in order for the right-hand-side of (3.19) to be positive, which is necessary for the existence of τ satisfying (3.19). τ appears in the risk bound (3.20) as a quadratic expression, so there is an optimal choice $\tau^* = \frac{\mathrm{tr}(W(V_2-V_{21}))-2\lambda_1}{\lambda_1^*}$ which minimizes this bound.

Corollary 9 $R\left(\hat{\beta}_{SP,c}\right) - R\left(\hat{\beta}_{SP,FE}\right) < 0$, *for* $d > 2$ *and* $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$. *When* $W = \left(V_2 - V_{21}\right)^{-1}$, $0 < \tau \leq 2\left(\frac{q-2}{\lambda_1^*}\right)$ *and* $q > 2$, *which is Stein's (1956) classic condition for shrinkage.*

Corollary 10 $R\left(\hat{\beta}_{SP,RE}\right) = tr(WV_1) + \sigma_1^{-4}\delta'\bar{X}'\bar{X}^*V_1WV_1\bar{X}^{*\prime}\bar{X}\delta$;

$R\left(\hat{\beta}_{SP,RE}\right) \leq R\left(\hat{\beta}_{SP,FE}\right)$ *when* $\sigma_1^{-4}\delta'\bar{X}'\bar{X}^*V_1WV_1\bar{X}^{*\prime}\bar{X}\delta \leq q$, *and* $R\left(\hat{\beta}_{SP,RE}\right) > R\left(\hat{\beta}_{SP,FE}\right)$ *otherwise.*

The result in Corollary 10 indicates that when endogeneity is weak (ρ and hence δ is close to zero), $\hat{\beta}_{SP,RE}$ may perform better than $\hat{\beta}_{SP,FE}$.

Corollary 11 $R\left(\hat{\beta}_{SP,c}\right) - R\left(\hat{\beta}_{SP,RE}\right) < 0$, *for* $q < \sigma_1^{-4}\delta'\bar{X}'\bar{X}^*V_1WV_1\bar{X}^{*\prime}\bar{X}\delta$, $d > 2$, *and* $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$.

The result in Corollary 11 indicates that when endogeneity is strong, $d > 2$, $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$, $\hat{\beta}_{SP,c}$ performs best among these three estimators.

Remark 12 *If the weight w is non-stochastic, we can obtain the asymptotic optimal w by minimizing*

$$w^2 R\left(\hat{\beta}_{SP,RE}\right) + (1-w)^2 R\left(\hat{\beta}_{SP,FE}\right) + 2w(1-w)\mathbb{E}\left(\left(\hat{\beta}_{SP,RE} - \beta\right)' W \left(\hat{\beta}_{FE} - \beta\right)\right)$$

w^* is given by

$$w^* = tr(V_2 - V_{21}) \,/\, \left(tr(V_1 + V_2 - (V_{21} + V_{21}')) + \sigma_1^{-4}\delta'\bar{X}'\bar{X}^* V_1 W V_1 \bar{X}^{*\prime}\bar{X}\delta\right)$$

which depends on the localizing parameter δ.

Remark 13 *A parametric combined estimator can be viewed as a special case of the semi-parametric combined estimator.*

Write (3.2) as

$$\begin{aligned} y_{it} &\approx x_{it}\beta + \alpha(z) + z_{it}\dot{m}(z) + \alpha_i + u_{it} \\ &= x_{it}\beta + Z\delta(z) + \alpha_i + u_{it} \end{aligned}$$

where $\alpha(z) = m(z) - z\dot{m}(z)$, $Z = (1,\ z_{it})$, $\delta(z) = \left(\alpha(z), (\dot{m}(z))'\right)'$. As $h \to \infty$, the weighted function $\mathbf{K}_H(z) \to \mathbf{K}(0)$ and local minimization becomes global minimization:

$$(y - X\beta - D\alpha - Z\delta)'(y - X\beta - D\alpha - Z\delta)$$

In this case, dropping Z or assuming it is in X, one can obtain the combined estimator for the parametric model as

$$\hat{\beta}_{SP,c} = w\hat{\beta}_{SP,RE} + (1-w)\hat{\beta}_{SP,FE}$$

where $\hat{\beta}_{FE} = (X'QX)^{-1} X'Qy,\ \hat{\beta}_{RE} = \left(X'\hat{\Omega}^{-1}X\right)^{-1} X'\hat{\Omega}^{-1}y$. Noting that if Z is included in regression, the only difference is that X will be replaced by MX and y will be replaced by My where M is the same as Q but based on Z.

The properties of the combined FE and RE estimator for the parametric panel data model were discussed in Huang (2015) and Wang et al. (2016).

3.5 Monte Carlo Simulation

The observations are generated by the progress

$$\begin{aligned} y_{it} &= x_{it}\beta + m(z_{it}) + \alpha_i + u_{it} \\ \alpha_i &= \rho\sqrt{T}\bar{x}_i \frac{\iota}{\sqrt{q}} + \sqrt{1-\rho^2}\ \varepsilon_i \end{aligned}$$

$\{x_{it},\ u_{it}\}$ are i.i.d $N(0,\ I_{q+1})$ across i, t. ϵ_i are i.i.d $N(0,\ 1)$ independent of $\{x_{it},\ u_{it}\}$. $\mathrm{Var}(\alpha_i) = 1$. The distribution are invariant to β so we set it to zero, $\beta = 0$. And set $m(z) = 2z + e^{-4(z-0.5)^2} - 1$.

Vary $n = \{20, 100\}$, $T = 3$, $q = \{4, 8\}$, and ρ on a 40-point grid on $[0, 0.975]$. ρ controls the degree of endogeneity, ranging in $(0,1)$ ($\rho = 0$ is the case of exogenous regressors; large ρ is the case of strong endogeneity). We also set $\sqrt{\theta} = \frac{\sigma_\alpha}{\sigma_u} \in \left\{\frac{1}{4}, 1, 4\right\}$ so $\rho^* = \frac{\theta}{1+\theta} = \{.06,\ .05,\ .94\}$. ρ^* $(0 \leq \rho^* \leq 1)$ controls the degree of heterogeneity which is the temporal correlation between $\alpha_i + u_{it}$ and $\alpha_i + u_{it'}$.

Generated 100,000 samples on each calculated $\hat{\beta}_{SP,RE}$, $\hat{\beta}_{SP,FE}$, $\hat{\beta}_{SP,c}$. To compare the estimators, calculate the median squared error (MSE) of each estimator and plot the relative median square error, that is

$$\frac{median\left[\left(\hat{\beta}-\beta\right)'\left(\hat{\beta}-\beta\right)\right]}{median\left[\left(\hat{\beta}_{SP,FE}-\beta\right)'\left(\hat{\beta}_{SP,FE}-\beta\right)\right]}$$

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

Figure 3.1 is the case $q = 8$, and Figure 3.2 is the case $q = 4$. The 6 plots in figure 3.2 look similar to the plots in Figure 1. By contrasting Figure 3.1 and 3.2, we can see that the improvement in the combined estimator over FE estimator with different values of ρ^* are greater in the cases of larger number of regressors.

Figure 3.1(a), 3.1(c), 3.1(e) are the cases $n = 20$, and Figure 3.1(b), 3.1(d), 3.1(f) are the cases $n = 100$. The region of dominance for the combined estimator over FE estimator is greater for small n.

Consider the case of eight endogenous regressors. Figure 3.1(a) and Figure 3.1(b) are the cases $\rho^* = 0.5$. For small ρ the RE estimator has lower MSE than the combined estimator, but the ranking is reversed for larger values of ρ. Figure 3.1(c) and Figure 3.1(d) are the cases $\rho^* = 0.06$. The MSE of the RE estimator is smaller than that of the combined estimator for all the values of ρ when $n = 20$ in Figure 3.1(c). The combined estimator has smaller MSE than that of the RE estimator when $n = 100$ in Figure 3.1(d). The MSE of the RE and combined estimators are uniformly smaller than that of the FE estimator. Figure 3.1(e) and Figure 3.1(f) are the cases $\rho^* = 0.94$. The FE and the combined estimators are near equivalents. RE has similar MSE to FE and combined estimators for small ρ, but the MSE of RE estimator increases dramatically after intermediate values of ρ.

Generally, the dominance for combined estimator over FE estimator is greater for small sample size. For very small ρ^*, RE estimator performs better than both FE and combined estimators except when n and ρ are large. For moderate ρ^* and larger ρ, the combined estimator performs better than RE estimator, but for very small ρ it can be beaten by RE estimator. For very large ρ^* and very small ρ, the combined estimator is close to both RE and FE estimators, while both combined and FE estimators have smaller risk than RE estimator for larger values of ρ. In summary, the simulation results provide strong finite sample confirmation of Theorem 8.

3.6 Empirical Results

In this section, we employ the same data set as chapter one to reexamine the relationship between public-sector capital accumulation and private sector productivity.

We consider the following semi-parametric panel data model:

$$\begin{aligned} \log(Y_{it}) &= \beta_0 + \beta_1 \log(KG_{i,t-2}) + \beta_2 \log(KPR_{i,t-2}) \\ &\quad + \beta_3 \log(L_{i,t-2}) + m\,(UNEM_{i,t-2}) + \alpha_i + u_{it} \end{aligned}$$

where $i = 1, \ldots, 48, t = 1, \ldots, 17$.

As before, in order to obtain the mean square errors (MSE) and the standard errors for these estimates, we bootstrap the data 10000 times by resampling across individuals and keep the time series structure for each individual unchanged. We obtain estimates of the average elasticities and coefficients for each bootstrap data, based on which we can calculate the bootstrap MSE and the standard errors for the above estimates. The MSE for FE, RE and combined estimators are 0.0148, 0.0133 and 0.0122, respectively. In order to obtain these results, we use the Gaussian kernel and choose the bandwidth of $m\,(UNEM)$ according to leave-one-out Cross-validation.

Table 3.1 suggests that the elasticities of gross private nonagricultural product for FE estimator with respect to KG, KPR, and L are -0.0207, 0.3501 and 0.5237,

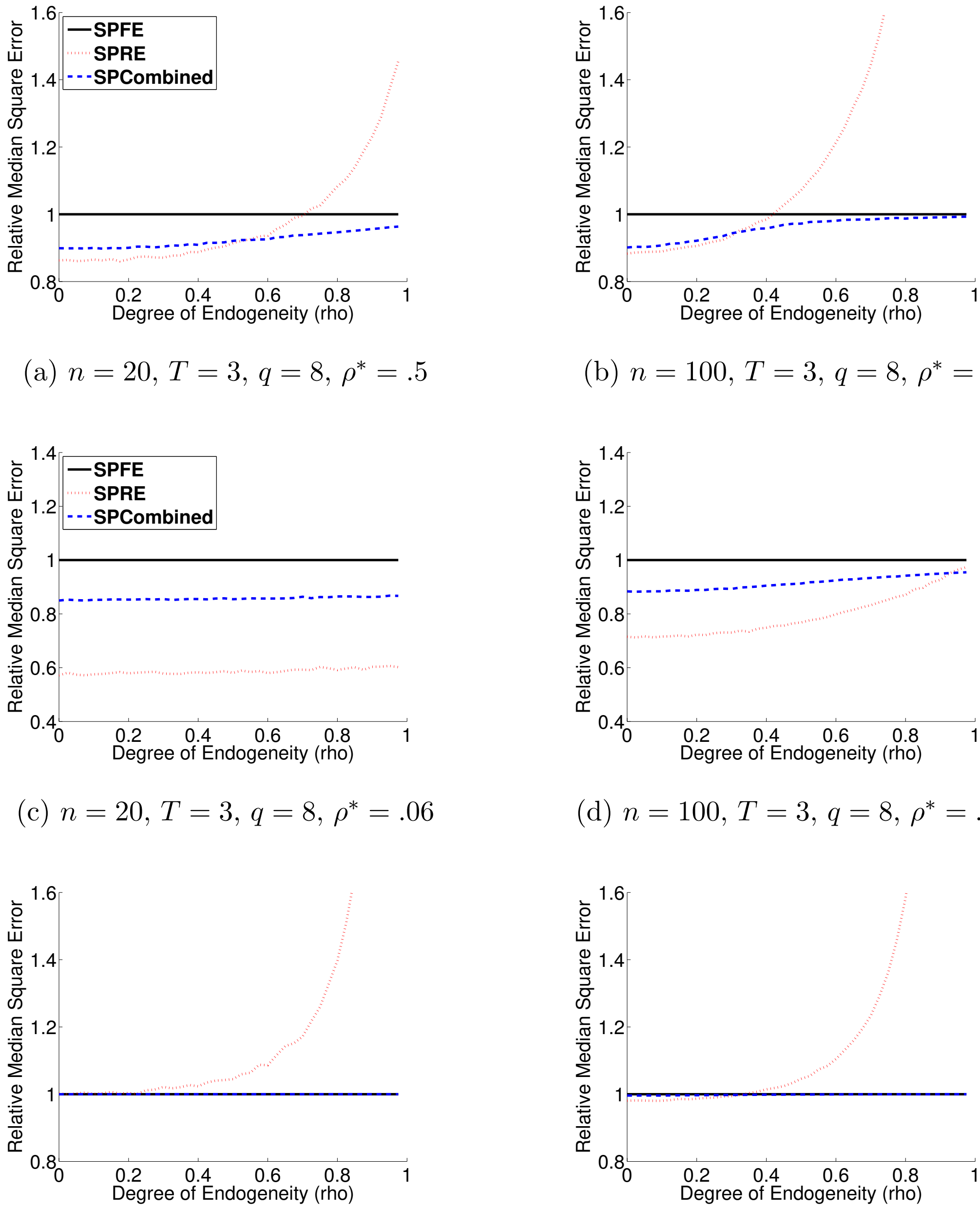
	$\hat{\beta}_{KG}$	$\hat{\beta}_{KPR}$	$\hat{\beta}_L$
FE	-0.0207	0.3501	0.5237
	(0.0379)	(0.0308)	(0.0317)
RE	0.0773	0.2440	0.5798
	(0.0435)	(0.0290)	(0.0383)
Combined	-0.0166	0.3480	0.5249
	(0.0378)	(0.0305)	(0.0313)

Table 3.1: Economics of Private Sector Output SP Estimates for 48 U.S. states, 1970-1986 (standard errors in parentheses)

respectively; The elasticities for RE estimator with respect to KG, KPR, and L are 0.0773, 0.2440 and 0.5798, respectively; The elasticities for the combined estimator with respect to KG, KPR, and L are -0.0166, 0.3480 and 0.5249, respectively. Both FE and combined estimators report that the public capital is counter productive and insignificantly in the state private production. In contrast, the RE estimator finds that the public capital is productive and insignificantly. Note that the Hausman statistic is 48.9985. Thus, the null hypothesis is easily rejected at the one percent level of significance. This indicates that there exists a huge problem of endogeneity. In this circumstance, the FE estimator solves the problem. Thus, it would be more appropriate to treat α as fixed. And the RE estimator seems overwhelming that public capital has a positive impact on private sector output. The combined estimator result is consistent with FE estimator. Our empirical analysis is in agreement with the findings of Holtz-Eakin (1994) and Baltagi and Pinnoi (1995) that there is no quantitatively important spillover effects across states. The estimated $\rho^* = 0.6486$, which may also explain why the combined estimator result is closer to FE estimator. As a result, the combined estimator is more reliable under this scenario. Thus we conclude that the public sector output has an insignificant effect on the private sector productivity.

3.7 Conclusion

This chapter provides a combined fixed effects and random effects estimator, with the weights depending on the Hausman statistic for the semi-parametric panel data model. We show that the combined estimator has strictly smaller asymptotic risk than the fixed effects estimator. The combined estimator also has smaller asymptotic risk compared to the random effetcs estimator unless the endogeneity is very weak. Our simulation result shows that the combined estimator can reduce finite sample MSE relative to the fixed effects estimator for all degrees of endogeneity and heterogeneity, as well as relative to the random effects estimator for moderate to large degrees of endogeneity and heterogeneity. We use the combined estimator to reexamine the role of public capital stock in the private sector output. We confirm the findings of Holtz-Eakin (1994) and Baltagi and Pinnoi (1995) that the public infrastructure has insignificant effects on private sector across states. Based on the empirical results, the combined estimator gives smaller MSE than both fixed effects and random effetcs estimators.



(a) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .5$

(b) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .5$

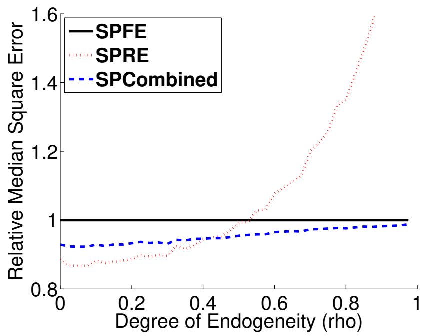
(c) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .06$

(d) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .06$

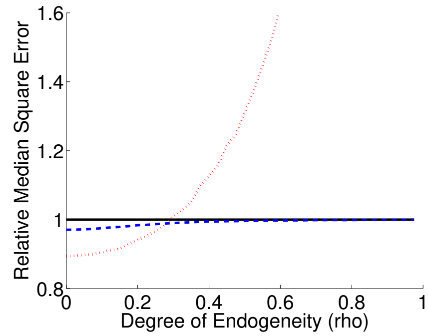
(e) $n = 20$, $T = 3$, $q = 8$, $\rho^* = .94$

(f) $n = 100$, $T = 3$, $q = 8$, $\rho^* = .94$

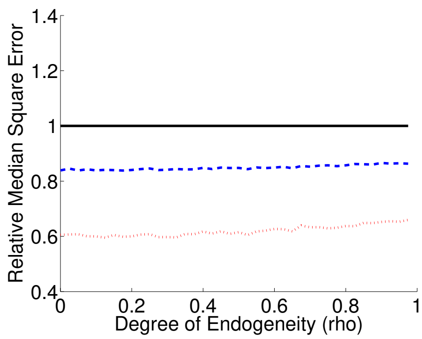
Figure 3.1: Relative MSE of SPFE, SPRE and SP Combined Estimators, $n = \{20, 100\}$, $T = 3$, $q = 8$, $\rho^* = \{.5, .06, .94\}$



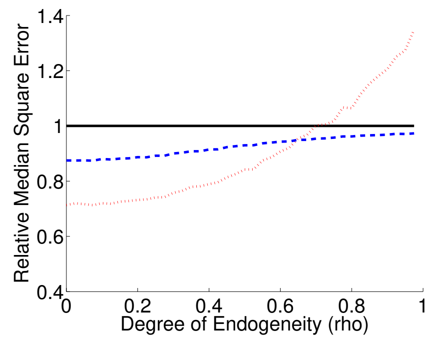
(a) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .5$



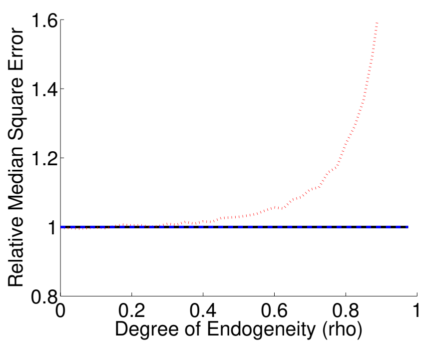
(b) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .5$



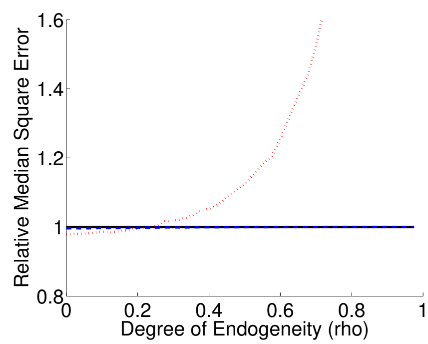
(c) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .06$



(d) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .06$



(e) $n = 20$, $T = 3$, $q = 4$, $\rho^* = .94$



(f) $n = 100$, $T = 3$, $q = 4$, $\rho^* = .94$

Figure 3.2: Relative MSE of SPFE, SPRE and SP Combined Estimators, $n = \{20, 100\}$, $T = 3$, $q = 4$, $\rho^* = \{.5, .06, .94\}$

Chapter 4

Combined Estimators for Structral Panel Data Model

4.1 Introduction

In a panel data model, the fixed effects estimator helps us to resolve the endogeneity issues that arise because of the correlated unobserved effects. However, endogeneity issues may also arise due to a nonzero correlation between explanatory variables and idiosyncratic errors. In the presence of such correlations, both fixed effects (FE) and random effects (RE) estimators yield biased and inconsistent estimates of the parameter. The resulting biases can not be removed via differencing estimation. The traditional technique to overcome this problem is to find instruments for those explanatory variables which are potentially correlated with idiosyncratic errors. For example see Hausman and Taylor (1981), Amemiya and MaCurdy (1986) and Breusch et al. (1989). These papers consider the application of instrumental-variable procedures to estimate the parameters of the model with endogenous regressors, with the error structure implied by random effects. And see Baltagi (2008) for the commonly used fixed effects 2SLS estimator.

It is well known that the finite sample properties of the 2SLS estimator are often problematic. Thus, most of the justification for the use of 2SLS estimator is asymptotic. Its performance in small samples may be poor. The precision of 2SLS estimates is lower than that of individual effects estimates. In the presence of weak instruments, the loss of precision will be severe, and 2SLS estimates may be no improvement over the individual effects estimators.

In this chapter we propose two combined estimators, which are weighted average of FE and FE-2SLS estimators, and weighted average of RE and RE-2SLS estimators with the weights depending on Hausman (1978) statistic. The asymptotic distribution and risk of the combined estimators are derived using a local asymptotic framework. We show that under certain conditions, the combined FE and FE-2SLS estimator has strictly smaller risk than the FE-2SLS estimator, and the combined RE and RE-2SLS estimator has strictly smaller risk than the RE-2SLS estimator. Further, in the Monte Carlo study we show that the combined estimators uniformly dominate the individual effects estimators for all degrees of endogeneity. The combined estimators are also better than the individual effects estimators except when the degree of endogeneity is very small, or when both very small sample size and very weak instruments are satisfied. Finally, to show the applicability of the combined estimator, we use a panel data on 90 counties in North Carolina over the period 1981–1987 to revisit the effect of police on crime.

The remainder of this chapter is organized as follows. Section 2 presents the model and estimators. Section 3 presents the asymptotic distribution of the combined estimators. Section 4 derives the asymptotic risk of the combined estimator. Monte

Carlo simulation is provided in Section 5. Section 6 is an empirical study. Section 7 concludes.

4.2 The Model and Estimators

4.2.1 The Fixed Effects and Random Effects Estimators

Consider the following panel regression model with fixed effects:

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}, \quad i = 1, 2, \ldots, n, \ t = 1, 2, \ldots, T$$

where x_{it} is $1 \times q$, and β is a $q \times 1$ vector of unknown parameters. $\alpha_i' s$ are fixed effects and $u_{it}' s$ are the random disturbances. In matrix notation, eq(1) can be written as

$$y = X\beta + D\alpha + u \tag{4.1}$$

$D = I_n \otimes \iota_T$ is $nT \times n$ where ι_T is a vector of ones, α is $n \times 1$, and $u \sim (0, \ \sigma_u^2 I_{nT})$. Pre-multiplying the model (4.1) by Q and performing OLS on the resulting transformed model:

$$Qy = QX\beta + QD\alpha + Qu \tag{4.2}$$

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

$$\hat{\beta}_{FE} = \left(X'QX\right)^{-1} X'Qy$$

where $Q = I_{nT} - D\left(D'D\right)^{-1} D'$, and $\hat{\beta}_{FE}$ has an asymptotic distribution as

$$\sqrt{n}\left(\hat{\beta}_{FE} - \beta\right) \xrightarrow{d} N\left(0, V_1\right)$$

where $V_1 = \sigma_u^2 \left(\text{plim}\frac{X'QX}{n}\right)^{-1}$.

The alternative specification for the panel data model is known as the random effects model, which assumes that α_i is drawn from an i.i.d. distribution $(0,\ \sigma_\alpha^2)$, and is uncorrelated both with u_{it} and with the x_{it}. Then

$$y_{it} = x_{it}\beta + v_{it}, \quad v_{it} = \alpha_i + u_{it} \tag{4.3}$$

In matrix notation, (4.3) can be written as

$$y = X\beta + v, \quad v = D\alpha + u \tag{4.4}$$

and the variance-covariance matrix is given by $\Omega = \sigma_u^2 I_{nT} + \sigma_\alpha^2 DD' = \sigma_1^2 P + \sigma_u^2 Q$, where $\sigma_1^2 = T\sigma_\alpha^2 + \sigma_u^2,\ P = D\left(D'D\right)^{-1} D'$. The inverse matrix of Ω is given by $\Omega^{-1} = \frac{1}{\sigma_1^2} P + \frac{1}{\sigma_u^2} Q$. The $\hat{\beta}_{RE}$ based on (4.4) is

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

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

$$\sqrt{n}\left(\hat{\beta}_{RE} - \beta\right) \xrightarrow{d} N\left(0, V_3\right)$$

where $V_3 = \left(\text{plim}\frac{X'\Omega^{-1}X}{n}\right)^{-1}$.

4.2.2 The Fixed Effects 2SLS and Random Effects 2SLS Estimators

Now, allow for the possible correlation between u_{it} and x_{it}. The vector x_{it} is treated as endogenous. Performing 2SLS on (4.2) with QZ as the set of instruments

$$Z'QQy = Z'QQX\beta + Z'QQu,$$

one gets the FE-2SLS estimator

$$\hat{\beta}_{FE,2SLS} = \left(X'H_Z X\right)^{-1} X'H_Z y \tag{4.5}$$

where $H_Z = QZ\left(Z'QZ\right)^{-1}Z'Q$. The asymptotic distribution of $\hat{\beta}_{FE,2SLS}$ follows

$$\sqrt{n}\left(\hat{\beta}_{FE,2SLS} - \beta\right) \xrightarrow{d} N\left(0, V_2\right)$$

where $V_2 = \sigma_u^2\left(\lim \frac{X'H_ZX}{n}\right)^{-1}$.

One can also perform 2SLS on (4.4) with $\Omega^{-\frac{1}{2}}Z$ as the set of instruments for $\Omega^{-\frac{1}{2}}X$,

$$Z'\Omega^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}y = Z'\Omega^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}X\beta + Z'\Omega^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}u,$$

and obtain the RE-2SLS estimator

$$\hat{\beta}_{RE,2SLS} = \left(X'R_ZX\right)^{-1}X'R_Zy \tag{4.6}$$

where $R_Z = \Omega^{-1}Z\left(Z'\Omega^{-1}Z\right)^{-1}Z'\Omega^{-1}$. The asymptotic distribution of $\hat{\beta}_{RE,2SLS}$ follows

$$\sqrt{n}\left(\hat{\beta}_{RE,2SLS} - \beta\right) \xrightarrow{d} N\left(0, V_4\right)$$

where $V_4 = \left(\lim \frac{X'R_ZX}{n}\right)^{-1}$.

The FE-2SLS estimator is preferred to the FE estimator as it is consistent under endogeneity, while the FE estimator is inconsistent. However in small samples, the FE-2SLS estimator can have much larger variance so the FE estimator have better precision. Motivated by this observation, we propose the following combined estimators of β, which is weighted average of FE and FE-2SLS estimators with the weights depending on Hausman statistic

$$\hat{\beta}_{c,FE} = w\hat{\beta}_{FE} + (1-w)\hat{\beta}_{FE,2SLS} \tag{4.7}$$

where

$$\begin{aligned} w &= \begin{cases} \frac{\tau}{H_n} & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases} \\ H_n &= \left(\hat{\beta}_{FE,2SLS} - \hat{\beta}_{FE}\right)'\left(\hat{V}_{FE,2SLS} - \hat{V}_{FE}\right)^{-1}\left(\hat{\beta}_{FE,2SLS} - \hat{\beta}_{FE}\right) \end{aligned}$$

Similarly, the RE-2SLS estimator is consistent under endogeneity, while the RE estimator is inconsistent. However in the presence of weak instruments, the loss of precision will be severe, and the RE-2SLS estimates may be no improvement over the RE estimator. In view of this, we also propose the following weighted average of RE and RE-2SLS estimators:

$$\hat{\beta}_{c,RE} = w^* \hat{\beta}_{RE} + (1 - w^*)\hat{\beta}_{RE,2SLS} \tag{4.8}$$

where

$$\begin{aligned} w^* &= \begin{cases} \frac{\tau^*}{H_n^*} & \text{if } H_n^* \geq \tau^* \\ 1 & \text{if } H_n^* < \tau^* \end{cases} \\ H_n^* &= (\hat{\beta}_{RE,2SLS} - \hat{\beta}_{RE})' \left(\hat{V}_{RE,2SLS} - \hat{V}_{RE}\right)^{-1} (\hat{\beta}_{RE,2SLS} - \hat{\beta}_{RE}) \end{aligned}$$

where τ is a shrinkage parameter. The degree of shrinkage depends on the ratio τ/H_n.

4.3 Asymptotic Distribution

First, write the reduced form equation for the endogenous variable x_{it} as

$$\begin{aligned} x_{it} &= \Pi' z_{it} + e_{it} \\ \mathbb{E}\left(e_{it} z_{it}\right) &= 0 \end{aligned} \tag{4.9}$$

Second, write the structural equation error as a linear function of the reduced form error and an orthogonal error

$$\begin{aligned} u_{it} &= e_{it}\rho + \varepsilon_{it} \\ \mathbb{E}\left(e_{it}\varepsilon_{it}\right) &= 0 \end{aligned} \tag{4.10}$$

The variable x_{it} is exogenous if u_{it} and e_{it} are uncorrelated, or equivalently that the coefficient ρ is zero. We use the local asymptotic approach. For fixed T, ρ is local to zero

$$\rho = \frac{1}{\sqrt{n}}\delta \tag{4.11}$$

δ is a $q \times 1$ localizing parameter, which indexes the degree of correlation between u_{it} and e_{it}. When $\delta = 0$, x_{it} are exogenous. When $\delta \neq 0$. x_{it} are endogenous. ρ (or δ) controls the degree of endogeneity.

Now, we make the following assumptions:

Assumption 4.1. $(x_i,\ \alpha_i,\ u_i)$ are i.i.d over i, u_{it} is i.i.d over t, $\mathbb{E}\left(u_{it} \mid x_{it},\ \alpha_i\right) = 0$ and $\mathbb{E}\left(u_{it}^4 \mid x_{it},\ \alpha_i\right) < \infty$, $\mathbb{E}\left(\varepsilon_{it} \mid e_{it}\right)^4 < \infty$.

Assumption 4.2. $\mathbb{E} \parallel x_{it} \parallel^{2+k} < \infty$ and $\mathbb{E}|u_{it}|^{2+k} < \infty$ for some $k > 0$.

Assumption 4.3. $\mathbb{E} \parallel x_{it} \parallel^4 < \infty, \mathbb{E} \parallel z_{it} \parallel^4 < \infty, \mathbb{E} \parallel e_{it} \parallel^4 < \infty; \sigma_u^2 \left(p\lim \frac{X'QX}{n}\right)^{-1} = V_1,\ \sigma_u^2 \left(p\lim \frac{X'H_ZX}{n}\right)^{-1} = V_2,\ \left(p\lim \frac{X'\Omega^{-1}X}{n}\right)^{-1} = V_3,$ and $\left(\lim \frac{X'R_ZX}{n}\right)^{-1} = V_4,$ as $n \to \infty$.

Assumption 4.4. $\hat{\sigma}_u^2 = \sigma_u^2 + o_p\left(1\right)$, $\hat{\sigma}_\alpha^2 = \sigma_\alpha^2 + o_p\left(1\right).$

Assumption 4.5. $\text{rank}(\Pi) = q$.

Assumption 4.1-4.3 specify that the variables have finite fourth moments, so that central limit theory applies. Assumption 5 is the rank condition on Π to ensure that the coefficient β is identified.

Set $\Sigma = \mathbb{E}\left(e_{it}e_{it}'\right).$

Theorem 14 *Under Assumptions 4.1-4.5,*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{FE} - \beta \\ \hat{\beta}_{FE,2SLS} - \beta \end{pmatrix} \xrightarrow{d} h + \xi \tag{4.12}$$

where

$$h = \begin{pmatrix} \sigma_u^{-2} V_1 tr\,(Q\Sigma)\,\delta \\ 0 \end{pmatrix}$$

$\xi \sim N(0, V)$ *with*

$$V = \begin{pmatrix} V_1 & V_1 \\ V_1 & V_2 \end{pmatrix}$$

Furthermore,

$$\begin{aligned} H_n &\to (h+\xi)' \, B(h+\xi) && (4.13) \\ \sqrt{n}\left(\hat{\beta}_c - \beta\right) \xrightarrow{d} \Psi &= G_2'\xi - \left(\frac{\tau}{(h+\xi)' \, B(h+\xi)}\right)_1 G'\,(h+\xi) && (4.14) \end{aligned}$$

where $B = G\,(V_2 - V_1)^{-1}\,G'$. *And*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{RE} - \beta \\ \hat{\beta}_{RE,2SLS} - \beta \end{pmatrix} \xrightarrow{d} h^* + \xi^* \tag{4.15}$$

where

$$h^* = \begin{pmatrix} V_3 tr\,(\Omega^{-1}\Sigma)\,\delta \\ 0 \end{pmatrix}$$

$\xi^* \sim N(0, V^*)$ *with*

$$V^* = \begin{pmatrix} V_3 & V_3 \\ V_3 & V_4 \end{pmatrix}$$

Furthermore,

$$\begin{aligned} H_n^* &\to (h^*+\xi^*)' \, B^*(h^*+\xi^*) && (4.16) \\ \sqrt{n}\left(\hat{\beta}_c^* - \beta\right) \xrightarrow{d} \Psi^* &= G_2'\xi^* - \left(\frac{\tau}{(h^*+\xi^*)' \, B^*(h^*+\xi^*)}\right)_1 G'\,(h^*+\xi^*) && (4.17) \end{aligned}$$

where $B^* = G\,(V_4 - V_3)^{-1}\,G'$, $G = \begin{pmatrix} -I & I \end{pmatrix}'$, $G_2 = \begin{pmatrix} 0 & I \end{pmatrix}'$, *and* $(a)_1 = min\,[1, a]$.

Theorem 14 gives expression for the joint asymptotic distribution of $\hat{\beta}_{FE}$ and $\hat{\beta}_{FE,2SLS}$ estimators, the joint asymptotic distribution of $\hat{\beta}_{RE}$ and $\hat{\beta}_{FE,2SLS}$ estimators, the

Hausman statistic, and the two combined estimators under the local exogeneity assumption. The joint asymptotic distributions are normal. $\hat{\beta}_{FE}$ and $\hat{\beta}_{RE}$ have asymptotic bias when $\delta \neq 0$ but not the $\hat{\beta}_{FE,2SLS}$ and $\hat{\beta}_{RE,2SLS}$ estimators. The Hausman statistic controls the weight and thus the degree of shrinkage. It is an asymptotic non-central chi-square random variable with non-centrality parameter depending on the local endogeneity parameter δ. The asymptotic distributions of the combined estimators are nonlinear functions of the normal random vector and functions of the noncentrality parameter. $\hat{\beta}_c$ is written as random weight average of the asymptotic distributions of $\hat{\beta}_{FE}$ and $\hat{\beta}_{FE,2SLS}$, and $\hat{\beta}_c^*$ is random weight average of the asymptotic distributions of $\hat{\beta}_{RE}$ and $\hat{\beta}_{RE,2SLS}$.

4.4 Asymptotic Risk

The asymptotic risk of any sequence of estimators β_n of β is defined as

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

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

$$R(\beta_n) = \mathbb{E}(\psi' W \psi) = \operatorname{tr}(W \mathbb{E}(\psi\psi')) \tag{4.18}$$

This shows that for such estimators and loss function only the local properties of the loss function affect the asymptotic risk.

Define the largest eigenvalue of the matrix $W(V_2 - V_1)$ and $W(V_4 - V_3)$

$$\begin{aligned} \lambda_1 &= \lambda_{\max}(W(V_2 - V_1)) \\ \lambda_1^* &= \lambda_{\max}(W(V_4 - V_3)) \end{aligned}$$

and the ratio

$$
\begin{aligned}
d &= \frac{\operatorname{tr}\left(W\left(V_2 - V_1\right)\right)}{\lambda_1} \\
d^* &= \frac{\operatorname{tr}\left(W\left(V_4 - V_3\right)\right)}{\lambda_1^*}.
\end{aligned}
$$

Theorem 15 *Under Assumptions 4.1-4.5, if*

$$
d > 2 \text{ and } 0 < \tau \leq 2\left(d-2\right), \tag{4.19}
$$

then

$$
\begin{aligned}
R\left(\hat{\beta}_{FE,2SLS}\right) &= tr\left(WV_2\right), \\
R\left(\hat{\beta}_c\right) &< R\left(\hat{\beta}_{FE,2SLS}\right) - \frac{\tau\lambda_1\left(2\left(d-2\right)-\tau\right)}{\sigma_u^{-4}\delta' tr\left(Q\Sigma\right)V_1\left(V_2-V_1\right)^{-1}V_1 tr\left(Q\Sigma\right)\delta + q}
\end{aligned} \tag{4.20}
$$

and if

$$
d^* > 2 \text{ and } 0 < \tau^* \leq 2\left(d^*-2\right), \tag{4.21}
$$

then

$$
\begin{aligned}
R\left(\hat{\beta}_{RE,2SLS}\right) &= tr\left(WV_4\right), \\
R\left(\hat{\beta}_c^*\right) &< R\left(\hat{\beta}_{RE,2SLS}\right) - \frac{\tau^*\lambda_1^*\left(2\left(d^*-2\right)-\tau^*\right)}{\delta' tr\left(\Omega^{-1}\Sigma\right)V_3\left(V_4-V_3\right)^{-1}V_3 tr\left(\Omega^{-1}\Sigma\right)\delta + q}
\end{aligned} \tag{4.22}
$$

Equation (4.20) shows that the asymptotic risk of the combined FE and FE-2SLS estimator is strictly less than that of the FE-2SLS estimator, so long as the shrinkage parameter τ satisfies the condition (4.19). The assumption $d > 2$ is necessary in order for the right-hand-side of the inequality equation in (4.19) to be positive, which is necessary for the existence of τ. Similarly equation (4.22) shows that the combined RE and RE-2SLS estimator has strictly smaller risk than the RE-2SLS estimator, so long as τ^* satisfies the condition (4.21).

τ appears in the risk bound (4.20) as a quadratic expression, so there is an optimal choice $\tau_{opt} = \frac{\text{tr}(W(V_2 - V_1))}{\lambda_1} - 2$ which minimizes this bound. Similarly, an optimal choice $\tau^*_{opt} = \frac{\text{tr}(W(V_4 - V_3))}{\lambda^*_1} - 2$ will minimize the risk bound (4.22).

In the special case $W = (V_2 - V_1)^{-1}$, we find that condition (4.19) simplifies to $q > 2$ and $0 < \tau \leq 2(q-2)$. Similarly, in the case $W = (V_4 - V_3)^{-1}$, condition (4.21) simplifies to $q > 2$ and $0 < \tau^* \leq 2(q-2)$. The assumption $q > 2$ is Stein's (1956) classic condition for shrinkage. Stein (1956) shows that the shrinkage dimension must exceed 2 in order for shrinkage to achieve global reductions in risk relative to unrestricted estimation.

Corollary 16 *If* $d > 2$ *and* $0 < \tau \leq 2(d-2)$, $R\left(\hat{\beta}_c\right) - R\left(\hat{\beta}_{FE,2SLS}\right) < 0$; *If* $d^* > 2$ *and* $0 < \tau^* \leq 2(d^* - 2)$, $R\left(\hat{\beta}^*_c\right) - R\left(\hat{\beta}_{RE,2SLS}\right) < 0$.

Corollary 17 *Under the local exogeneity assumption,*

$$\begin{aligned} R\left(\hat{\beta}_{FE}\right) &= tr\left(WV_1\right) + \sigma_u^{-4}\delta' tr\left(Q\Sigma\right) V_1 W V_1 tr\left(Q\Sigma\right)\delta \\ R\left(\hat{\beta}_{RE}\right) &= tr\left(WV_3\right) + \delta' tr\left(\Omega^{-1}\Sigma\right) V_3 \left(V_4 - V_3\right)^{-1} V_3 tr\left(\Omega^{-1}\Sigma\right)\delta, \end{aligned}$$

and

$$\begin{cases} R\left(\hat{\beta}_{FE}\right) \leq R\left(\hat{\beta}_{FE,2SLS}\right) & \text{if } \sigma_u^{-4}\delta' tr\left(Q\Sigma\right) V_1 W V_1 tr\left(Q\Sigma\right)\delta \leq q \\ R\left(\hat{\beta}_{FE}\right) > R\left(\hat{\beta}_{FE,2SLS}\right) & \text{if otherwise.} \end{cases}$$

$$\begin{cases} R\left(\hat{\beta}_{RE}\right) \leq R\left(\hat{\beta}_{RE,2SLS}\right) & \text{if } \delta' tr\left(\Omega^{-1}\Sigma\right) V_3 \left(V_4 - V_3\right)^{-1} V_3 tr\left(\Omega^{-1}\Sigma\right)\delta \leq q \\ R\left(\hat{\beta}_{RE}\right) > R\left(\hat{\beta}_{RE,2SLS}\right) & \text{if otherwise.} \end{cases}$$

Corollary 17 indicates that when endogeneity is weak (ρ and hence δ is close to zero) the FE estimator may perform better than the FE-2SLS estimator, the RE estimator may perform better than the RE-2SLS estimator as well.

Corollary 18 *If* $q < \sigma_u^{-4}\delta' tr\left(Q\Sigma\right)V_1 W V_1 tr\left(Q\Sigma\right)\delta,\ d > 2,\ 0 < \tau \le 2\left(d-2\right),$ $R\left(\hat{\beta}_c\right) - R\left(\hat{\beta}_{FE}\right) < 0$*; If* $q < \delta' tr\left(\Omega^{-1}\Sigma\right)V_3\left(V_4 - V_3\right)^{-1}V_3 tr\left(\Omega^{-1}\Sigma\right)\delta,\ d^* > 2,$ $0 < \tau^* \le 2\left(d^*-2\right),\ R\left(\hat{\beta}_c^*\right) - R\left(\hat{\beta}_{RE}\right) < 0.$

Corollary 18 indicates that when endogeneity is strong, $d > 2,\ 0 < \tau \le 2\left(d-2\right),$ the combined FE and FE-2SLS estimator performs better than both the FE and FE-2SLS estimators. Similarly, the combined RE and RE-2SLS estimator performs better than both the RE and RE-2SLS estimators when endogeneity is strong, $d^* > 2,$ $0 < \tau^* \le 2\left(d^*-2\right).$

Remark 19 *If the weight* w *in FE case is non-stochastic, we can obtain the asymptotic optimal* w *by minimizing*

$$w^2 R\left(\hat{\beta}_{FE}\right) + \left(1-w\right)^2 R\left(\hat{\beta}_{FE,2SLS}\right) + w\left(1-w\right)E\left(\left(\hat{\beta}_{FE}-\beta\right)' W\left(\hat{\beta}_{FE,2SLS}-\beta\right)\right)$$

w *is given by*

$$w = tr\left(V_2 - V_1\right) / \left(tr\left(V_2 - V_1\right) + \sigma_u^{-4}\delta' tr\left(Q\Sigma\right)V_1 W V_1 tr\left(Q\Sigma\right)\delta\right)$$

which depends on the localizing parameter δ*.*

Similarly, if the weight w^* *in RE case is non-stochastic, the asymptotic optimal* w^* *is*

$$w^* = tr\left(V_4 - V_3\right) / \left(tr\left(V_4 - V_3\right) + \delta' tr\left(\Omega^{-1}\Sigma\right)V_3\left(V_4 - V_3\right)^{-1}V_3 tr\left(\Omega^{-1}\Sigma\right)\delta\right)$$

Remark 20 *If part of regressors are treated as endogenous, consider the following structural equation of a panel data model:*

$$y = Z\delta + D\alpha + u \tag{4.23}$$

where $Z = \left[Y_1,\quad X_1\right]$ *and* $\delta = \left[\gamma,\quad \beta\right].$ Y_1 *is the set of* m *right-hand side endogenous variables, and* X_1 *is the set of* q_1 *included exogenous variables. Let* $X = \left(X_1,\quad X_2\right)$

be the set of all exogenous variables. This equation is identified with q_2 the number of excluded exogenous variables from the equation (X_2) being larger than or equal to m. In this case, one can use QX as the set of instruments to get the FE-2SLS estimator as

$$\hat{\beta}_{FE-2SLS} = \left(Z'H_Z Z\right)^{-1} Z'H_Z y$$

with $H_Z = QX \left(X'QX\right)^{-1} X'Q$. Alternatively, one can use $\Omega^{-\frac{1}{2}}X$ as the set of instruments to get the RE-2SLS estimator in this case as

$$\hat{\beta}_{RE-2SLS} = \left(Z'R_Z Z\right)^{-1} Z'R_Z y$$

with $R_Z = \Omega^{-1}X \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}$.

4.5 Monte Carlo Simulation

Our simulation experiment uses a design similar to Hansen (2014). The observations are generated by the progress

$$\begin{aligned} y_{it} &= x_{it}\beta + \alpha_i + u_{it} \\ x_{it} &= \Pi z_{it} + e_{it} \\ u_{it} &= \rho e_{it} \frac{\iota}{\sqrt{q}} + \sqrt{1-\rho^2}\, \varepsilon_{it} \end{aligned}$$

The z_{it}, u_{it}, and reduced form errors e_{it} are all i.i.d $N\left(0,\ 1\right)$ across i, t, with the error u_{it} and e_{it} having correlation $\frac{\rho}{\sqrt{q}}$, but all other correlation zero. α_i are i.i.d $N\left(0,\ 1\right)$ independent of $\{x_{it},\ u_{it}\}$.

The results are not quantitatively sensitive to the value of β, so we set it to zero. We also set the $q \times q$ reduced form matrix as $\Pi = I_q d$ and the scale d set as $d = \sqrt{R^2/\left(1-R^2\right)}$, so that R^2 is the reduced form population R^2. Note that we set the dimension of z_{it} equal to that of x_{it}.

We vary $n = \{20, 100\}$, $T = 3$, $q = \{3, 6\}$, $R^2 = \{0.1,\ 0.4\}$ and ρ on a 40-point grid on $[0, 0.975]$. The parameter R^2 controls the strength of the instruments (small R^2 is the case of weak instruments; large R^2 is the case of strong instruments) and the parameter ρ controls the degree of endogeneity, ranging in $(0, 1)$ ($\rho = 0$ is the case of exogenous regressors; large ρ is the case of strong endogeneity).

We generated 50,000 samples on each calculated $\hat{\beta}_{FE}$, $\hat{\beta}_{FE,2SLS}$, $\hat{\beta}_{c,FE}$, $\hat{\beta}_{RE}$, $\hat{\beta}_{RE,2SLS}$, $\hat{\beta}_{c,RE}$. To compare the estimators we calculate the median squared error of each estimator, that is

$$R\left(\hat{\beta}\right) = median\left(\left(\hat{\beta} - \beta\right)'\left(\hat{\beta} - \beta\right)\right)$$

We present the results graphically. Figure 4.1 and 4.2 is the FE case, and Figure 4.3 and 4.4 is the RE case. The Figure 4.1 and 4.3, Figure 4.2 and 4.4 are remarkably similar across the choices $\{n,\ q,\ R^2\}$. All the plots show that the MSE of the combined estimators are uniformly smaller than that of the 2SLS estimator.

Consider the case of FE. Figure 4.1 are the cases $q = 3$, and Figure 4.2 are the cases $q = 6$. By contrasting the four plots in Figure 4.1 and the plots in Figure 4.2, we can see that they look similar, and the improvement in the combined estimator over FE-2SLS estimator with different values of ρ are greater in the cases of larger q. Figure 4.1(a), 4.1(c), 4.2(a), 4.2(c) are the cases $n = 20$, and Figure 4.1(b), 4.1(d), 4.2(b), 4.2(d) the cases $n = 100$. The reductions in MSE are much greater for the smaller values of n. Figure 4.1(a), 4.1(b), 4.2(a), 4.2(b) are the cases $R^2 = 0.4$, and Figure 4.1(c), 4.1(d), 4.2(c), 4.2(d) are the cases $R^2 = 0.1$. The region of dominance for the combined estimator over FE-2SLS estimator is greater for weak instruments.

Figure 4.2(b) plots the MSE for $n = 20$, $q = 6$, $R^2 = 0.4$. The figure shows that the combined estimator has much lower MSE than FE-2SLS estimator, regardless of the degree of the endogeneity. For small ρ the FE estimator has lower MSE than the combined estimator, but the ranking is reversed for larger value of ρ. Figure 4.2(d) plots the MSE for $n = 100$, $q = 6$, $R^2 = 0.1$. This is the case of a large sample and with weak identification. In this picture we again see that the combined estimator has uniformly smaller MSE than FE-2SLS estimator, with the MSE converging to that of FE-2SLS as ρ increases towards 1. The FE estimator achieves some reduction in MSE relative to FE-2SLS and combined estimators for small values of ρ, but has higher MSE for moderate values of ρ. Figure 1-6 plots the MSE for $n = 100$, $q = 6$, $R^2 = 0.4$. The general nature of the plot is the same, except that the gain from the combined estimator is not as strong as in the weak instruments case. Figure 4.2(c) plots the MSE for $n = 20$, $q = 6$, $R^2 = 0.1$. This is the case of a small sample and with weak identification. The reduction in risk achieved by shrinkage is dramatic. However, the MSE of the combined estimator is higher than that of the FE estimator.

Overall, the improvement in the combined estimators over individual effects estimators are greatest in the cases of small sample sizes, small degree of endogeneity and weak instruments. The plots for the RE case look similar to that in the FE cases. For very small ρ, individual effects estimators perform better than both 2SLS and combined estimators except when n and ρ are large. The combined estimators uniformly dominate 2SLS estimators. But for very small n and weak instruments, the combined estimators can be beaten by the individual effects estimator. In summary, the simulation results provide strong finite sample confirmation of Theorem 15.

4.6 Empirical Results

In this section, we use a panel data on 90 counties in North Carolina over the period 1981–1987 to revisit the effect of police on crime. Cornwell and Trumbull (1994), hereafter CT, analyze empirical evidence that the ability of the criminal justice system to deter crime. Baltagi (2006) replicates the CT estimation and confirms their conclusion. A large body of research has explored the police–crime relationship. Some theories support a negative impact of police on crime (e.g., Ehrlich, 1972; Marvell and Moody 1996; Levitt, 2002). Additional police presence deters crime by making criminals believe arrests and subsequent sanctions are more likely. However, many studies of police behavior believe that police have little or no impact on crime, see Bayley (1996). Another type of finding argues that some criminals, facing greater arrest risks, switch to less risky crime types and methods, resulting in a positive impact of police on crime, see Cook (1979).

The empirical model follows Cornwell and Trumbull (1994), and relates the crime rate to a set of explanatory variables. The explanatory variables consist of the probability of arrest (P_A), probability of conviction given arrest (P_C), probability of a prison sentence given a conviction (P_P); average prison sentence in days (S); the number of police per capita ($Police$); the population density ($Density$); a dummy variable ($Urban$) indicating whether the county is in the SMSA with population larger than 50000; the proportion of county population that is male and between the ages of 15 and 24 ($Percent\ Young\ Male$); the proportion that is minority or nonwhite ($Percent\ Minority$); regional dummies for western and central counties ($West$ and $Central$). Opportunities in the legal sector are captured by the average weekly wage in the county by industry. The industry categories are: construction ($WCON$); transportation, utilities and communications ($WTUC$); wholesale and retail trade ($WTRD$); finance, insurance and real estate ($WEIR$); services ($WSER$); manufacturing ($WMFG$); and federal, state and local government ($WFED$, $WSTA$ and $WLOC$). All variables are in logs except for the dummies. Details in this dataset can be found in CT.

CT worried about the endogeneity of police per capita and the probability of arrest. They used per capita tax revenue and offence mix as two instruments. Offence mix is the ratio of crimes involving "face-to-face" contact to those that do not. We also use the first lag of the regressors as instrumental variables.

The economic model of crime predicts that the estimated coefficients of P_A, P_C, P_P, and S will be negative since an increase in the probability or severity of punishment increases the expected cost, or decreases the expected utility, of crime. Furthermore, it is well known that under certain assumptions the economic model of crime implies an ordering of deterrent effects: the greatest impact on crime coming from P_A, followed by Pc and P_P. The estimated elasticities reported in table 4.1 are generally consistent with the predictions of the theoretical model. In all cases the estimated elasticities are negative.

FE results show that the probability of arrest, the probability of conviction given arrest and the probability of a prison sentence given a conviction all have a negative and significant effect on the crime rate with estimated elasticities of -0.3670, -0.3027, -0.1782. The sentence severity has a negative and insignificant effect on the crime rate crime. In both FE 2SLS and combined results, the estimated elasticities of the probability of arrest and the probability of conviction given arrest are statistically insignificant. The sentence severity has a positive and significant effect in the FE-2SLS results, and it is insignificant in the combined estimates. The MSE for FE, FE-2SLS and combined estimators are 0.7359, 0.3308 and 0.2957, respectively. The

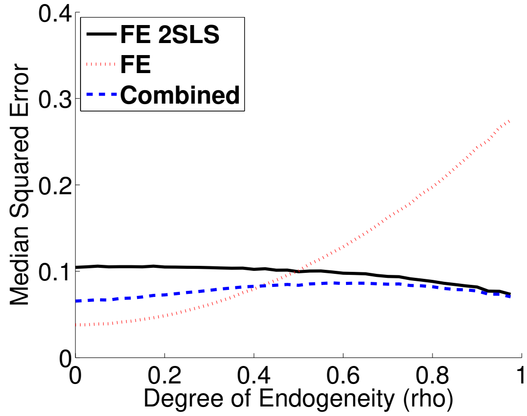
	$\hat{\beta}_{FE}$	$\hat{\beta}_{FE,2SLS}$	$\hat{\beta}_{c,FE}$
P_A	-0.3670	-1.0323	-0.9914
	(0.0426)	(0.8850)	(0.7623)
P_C	-0.3027	-0.8674	-0.8086
	(0.0263)	(0.5905)	(0.4558)
P_P	-0.1782	-1.9985	-1.2357
	(0.0313)	(0.2660)	(0.2437)
S	-0.0255	0.4599	0.1565
	(0.0341)	(0.0688)	(0.0878)
$Police$	0.3998	1.1023	1.0382
	(0.0202)	(0.7039)	(0.6730)

Table 4.1: Economics of Crime Estimates for North Carolina, 1981–1987 (standard errors in parentheses)

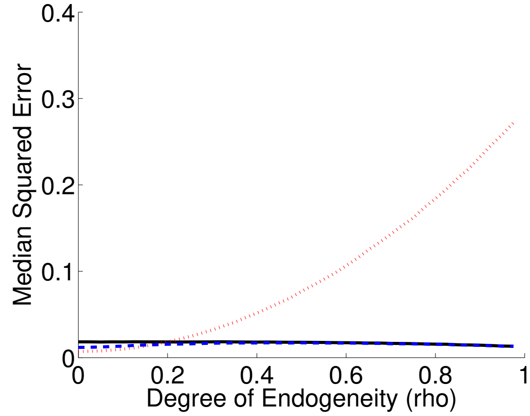
FE estimations on the elasticity of the crime rate with respect to the number of police are positive. However, accounting for the endogeneity, the respective FE-2SLS and the combined estimations are insignificant. Thus we conclude that police presence has an insignificant effect on the crime rate.

4.7 Conclusions

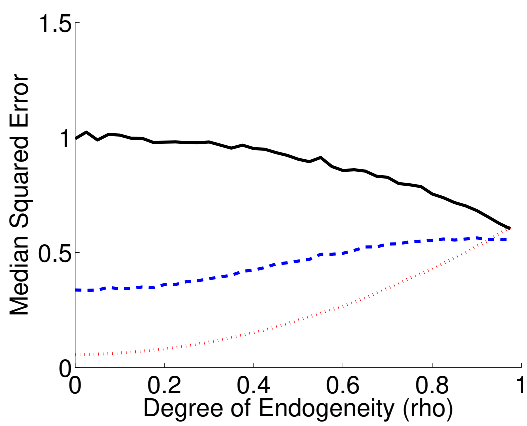
This chapter proposes two combined estimators which are weighted average of FE and FE-2SLS estimators, and weighted average of RE and RE-2SLS estimators, using the weights inversely proportional to the Hausman statistic. We show that under certain conditions, the combined FE and FE-2SLS estimator has strictly smaller asymptotic risk than the FE-2SLS estimator, and the combined RE and RE-2SLS estimator has strictly smaller asymptotic risk than the RE-2SLS estimator. Further, in the Monte Carlo study we show that the results also hold for small samples. The magnitude of efficiency of the combined estimator over FE and FE-2SLS estimators varies with respect to the degree of endogeneity and the strength of instruments. The combined estimator over RE and RE-2SLS estimators shares similar asymptotic and finite sample properties. We use the combined estimator over FE and FE-2SLS to revisit the effect of police on crime. We confirm the findings of Bayley (1996) that the police level has insignificant effect on crime. Based on the empirical results, the combined estimator gives smaller MSE than both FE and FE-2SLS estimators.



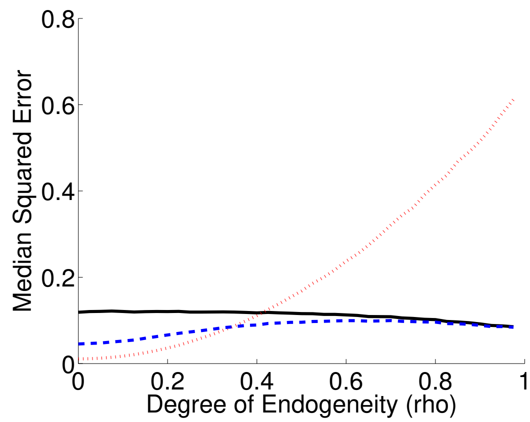
(a) $n = 20, T = 3, q = 3, R^2 = .4$



(b) $n = 100, T = 3, q = 3, R^2 = .4$

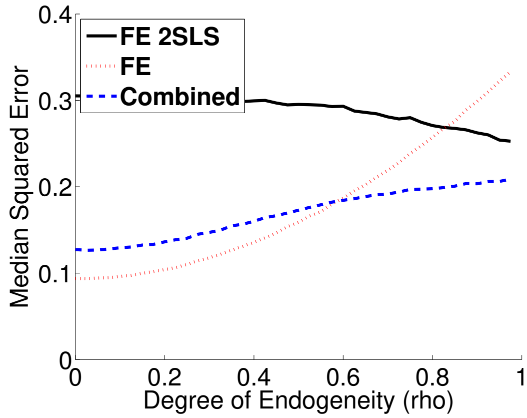


(c) $n = 20, T = 3, q = 3, R^2 = .1$

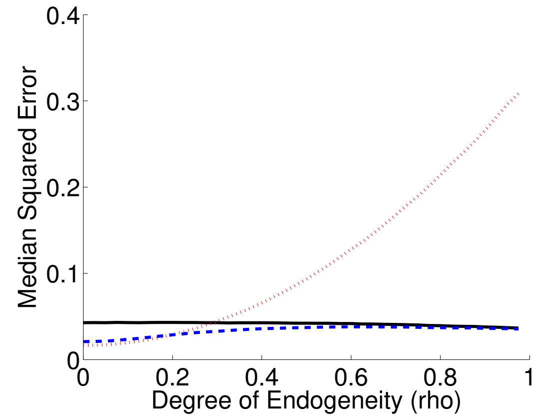


(d) $n = 100, T = 3, q = 3, R^2 = .1$

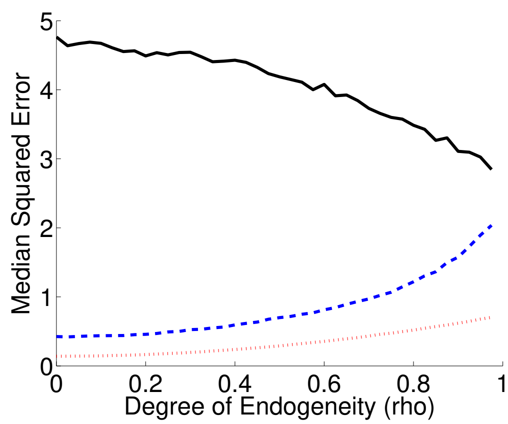
Figure 4.1: Relative MSE of FE, FE-2SLS and Combined Estimators, $n = \{20, 100\}$, $T = 3,\ q = 3,\ R^2 = \{.4, .1\}$



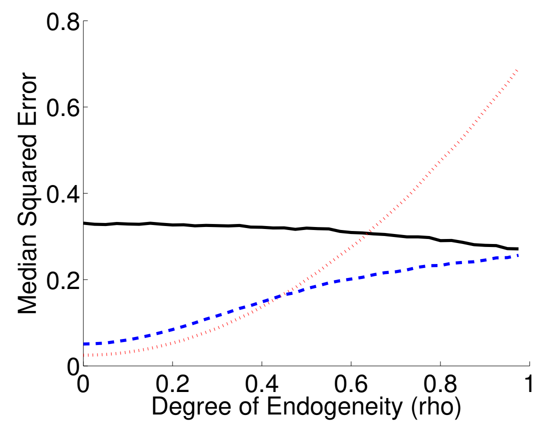
(a) $n = 20, T = 3, q = 6, R^2 = .4$



(b) $n = 100, T = 3, q = 6, R^2 = .4$

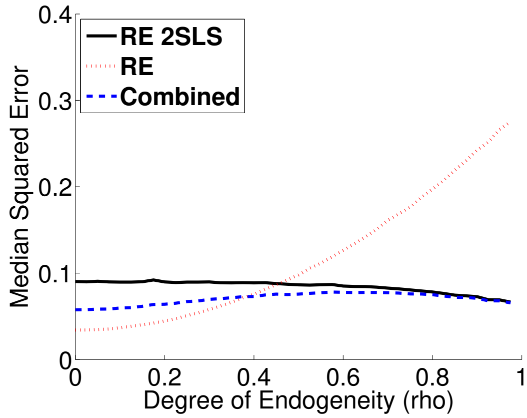


(c) $n = 20, T = 3, q = 6, R^2 = .1$

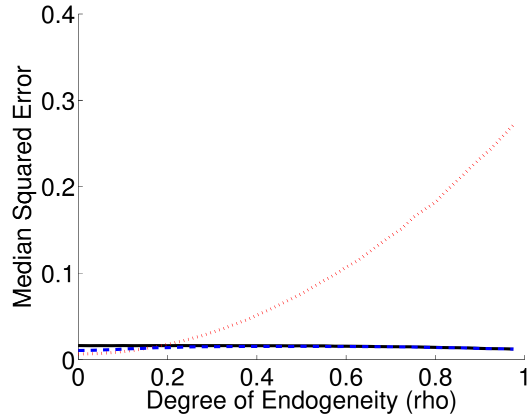


(d) $n = 100, T = 3, q = 6, R^2 = .1$

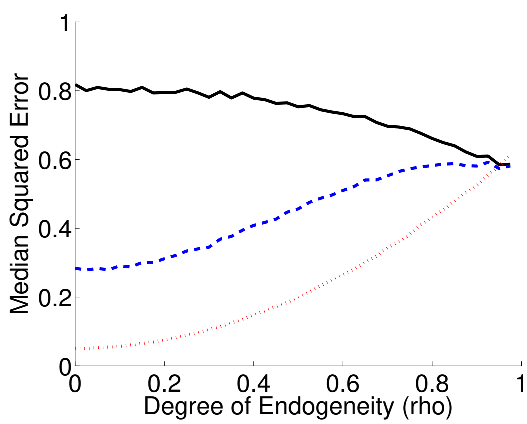
Figure 4.2: Relative MSE of FE, FE-2SLS and Combined Estimators, $n = \{20, 100\}$, $T = 3$, $q = 6$, $R^2 = \{.4, .1\}$



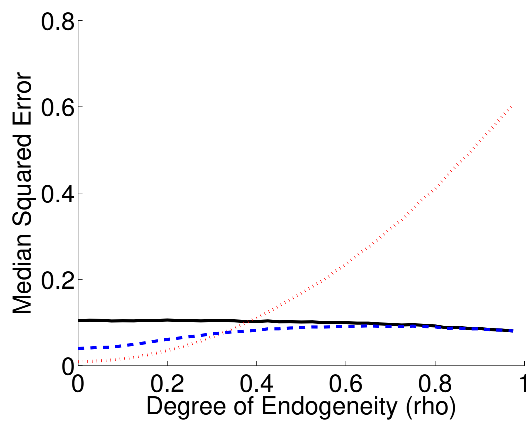
(a) $n = 20, T = 3, q = 3, R^2 = .4$



(b) $n = 100, T = 3, q = 3, R^2 = .4$

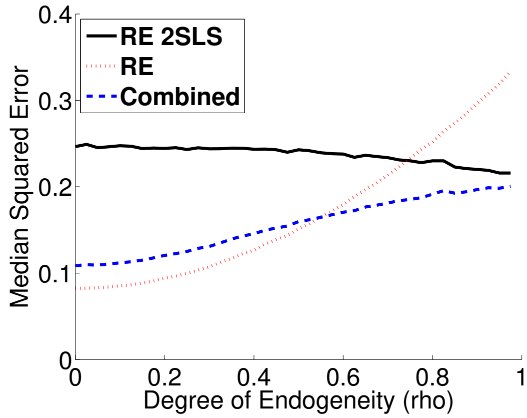


(c) $n = 20, T = 3, q = 3, R^2 = .1$

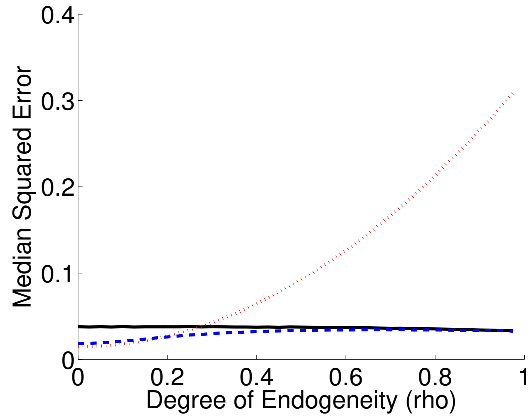


(d) $n = 100, T = 3, q = 3, R^2 = .1$

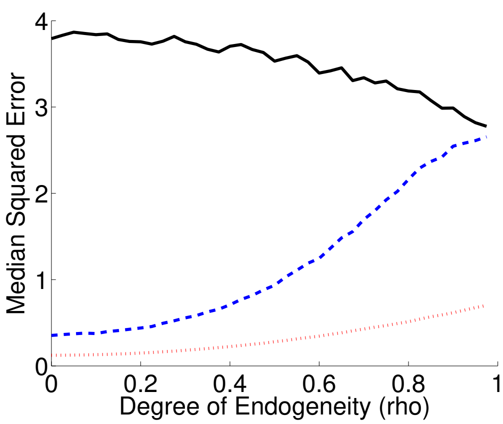
Figure 4.3: Relative MSE of RE, RE-2SLS and Combined Estimators, $n = \{20, 100\}$, $T = 3,\ q = 3,\ R^2 = \{.4, .1\}$



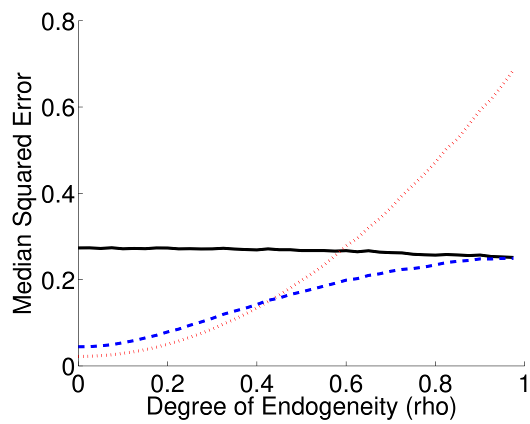
(a) $n = 20, T = 3, q = 6, R^2 = .4$



(b) $n = 100, T = 3, q = 6, R^2 = .4$



(c) $n = 20, T = 3, q = 6, R^2 = .1$



(d) $n = 100, T = 3, q = 6, R^2 = .1$

Figure 4.4: Relative MSE of RE, RE-2SLS and Combined Estimators, $n = \{20, 100\}$, $T = 3$, $q = 6$, $R^2 = \{.4, .1\}$

Chapter 5

A Combined Estimator for Large Heterogeneous Panels with Multifactor Error Structure

5.1 Introduction

Recently, there has been increased interest in the estimation of models with error cross-sectional dependence. 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 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 (CCE) 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. The pooled version is also provided, when the individual slope coefficients are homogenous.

There is a large body of literature on large panels with a multifactor error structure. The correlated common effects estimator based on cross-sectional averages has been developed by Pesaran (2006), Kapetanios, Pesaran, and Yamagata (2011), Pesaran and Tosetti (2011), Chudik, Pesaran, and Tosetti (2011), Pesaran, Smith, and Yamagata (2013), and Chudik and Pesaran (2015).

If error cross-sectional dependence exists in the model, the common correlated effects pooled (CCEP) model is consistent and the fixed effects (FE) model is inconsistent. On the other hand, if there is no error cross-sectional dependence, both CCEP and FE estimators are consistent and CCEP estimator is efficient. In this chapter, we propose a combined estimator which is a weighted combination of the standard fixed effects estimator and the common correlated effects pooled estimator of Pesaran (2006). 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 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 CCEP estimator for all degrees of endogeneity, as well as relative to FE estimator for moderate to large degrees of endogeneity.

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

The rest of this chapter is organized as follows. Section 2 presents the FE, CCEP and combined estimators. Section 3 presents the asymptotic distribution of the combined estimator. Section 4 derives the asymptotic risk of the combined estimator. Monte Carlo simulation is provided in Section 5. Empirical example is given in Section 6. Section 7 concludes.

5.2 The Model and Estimators

Consider a panel data regression model

$$y_{it} = x_{it}'\beta + \alpha_i + e_{it} \tag{5.1}$$

where $i = 1, 2, \ldots n$ and $t = 1, 2, \ldots T$. x_{it} is the ith observation on 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 remainder disturbances stochastic e_{it} has a multifactor structure

$$e_{it} = \gamma_i' f_t + \varepsilon_{it} \tag{5.2}$$

in which f_t is the $r \times 1$ vector of individual-invariant time-specific unobserved common effects, γ_i is a $1 \times r$ stochastic individual-specific factor loading vector, and ε_{it} are the idiosyncratic errors assumed to be independently distributed of x_{it}. To model the correlation between the individual specific regressor x_{it}, and the errors e_{it}, x_{it} can be written as

$$x_{it} = \Gamma_i' f_t + v_{it} \tag{5.3}$$

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

$$\begin{aligned} y_i &= X_i \beta + \alpha_i \iota_T + e_i \\ e_i &= F\gamma_i + \varepsilon_i \\ X_i &= F\Gamma_i + v_i \end{aligned} \tag{5.4}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$ is $T \times 1$, $X_i = (x_{i1}', \ldots, x_{iT}')'$ is $T \times q$, $e_i = (e_{i1}, \ldots, e_{iT})'$ 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 $T \times q$.

Make the following assumptions on the common factors, their loadings and the individual or unit specific errors:

Assumption 5.1: ε_{it} is independently and identically distributed (iid) across both i and t with $\mathbb{E}(\varepsilon_{it}) = 0$, $Var(\varepsilon_{it}) = \sigma_i^2 = \sigma^2 > 0$ and $\mathbb{E}(\| \varepsilon_{it} \|^4) < \infty$;

Assumption 5.2: v_{it} is iid across both i and t with $\mathbb{E}(v_{it}) = 0$, $Var(v_{it}) = \Sigma_i$ positive definite and $\mathbb{E}(\| v_{it} \|^4) < \infty$;

Assumption 5.3: f_t is covariance stationary with absolute summable autocovariances, such that $\mathbb{E}(\| f_t \|^4) < \infty$;

Assumption 5.4: γ_i and Γ_i are iid across i and of ε_{jt} and v_{jt}, f_t for all i and t with fixed means γ and Γ, and finite variances. In particular,

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

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

Assumption 5.5: ε_{it}, v_{it} and f_t are mutually independent.

5.2.1 The Fixed Effects Estimator

Define $Q_T \equiv I_T - \iota_T \left(\iota_T'\iota_T\right)^{-1} \iota_T'$, which is a $T \times T$ symmetric, idempotent matrix. Further, $Q_T\iota_T = 0$, and so for ith unit, premultiplying (5.4) by Q_T gives

$$Q_T y_i = Q_T X_i \beta + Q_T e_i$$

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

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

and

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

where $R^* = \text{plim}(\sigma^2 \sum_{i=1}^{n} X_i' Q_T X_i + \sum_{i=1}^{n} X_i' Q_T F \Omega_\eta F' Q_T X_i)$, $\Psi^{*-1} = \text{plim}(\sum_{i=1}^{n} X_i' Q_T X_i)$.

If error cross-sectional dependence exists in the model ($\gamma_i \neq 0$), FE estimator is inconsistent. If $\gamma_i = 0$, $\hat{\beta}_{FE}$ is consistent and has the following asymptotic distribution, as $n \to \infty$

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

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

Assumption 5.6: $\sum_{i=1}^{n} \left(\frac{X_i' Q_T X_i}{n}\right)$ is bounded and nonsingular.

5.2.2 The Common Correlated Effects Pooled Estimator

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.

As a way to illustrate this result, write (5.1) and (5.3) more compactly as

$$z_{it} = \begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = B_i + C_i' f_t + u_{it} \tag{5.5}$$

where

$$\begin{aligned} u_{it} &= \begin{pmatrix} \beta' v_{it} + \varepsilon_{it} \\ v_{it} \end{pmatrix} \\ B_i &= \begin{pmatrix} \alpha_i \\ \mathbf{0} \end{pmatrix}, \; C_i = \begin{pmatrix} \gamma_i & \Gamma_i \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \beta & I_q \end{pmatrix} \end{aligned}$$

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 u_{it} is given by

$$\mathbb{E}\left(u_{it} u_{it}'\right) = \Sigma_{u,i} = \begin{bmatrix} \beta' \Sigma_i \beta + \sigma_i^2 & \beta' \Sigma_i \\ \Sigma_i \beta & \Sigma_i \end{bmatrix}$$

Then the cross section average

$$\bar{z}_t = \bar{B} + \bar{C}' f_t + \bar{u}_t$$

where

$$\bar{z}_t = \frac{1}{n} \sum_{i=1}^{n} z_{it}, \; \bar{B} = \frac{1}{n} \sum_{i=1}^{n} B_i, \; \bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i, \; \bar{u}_t = \frac{1}{n} \sum_{i=1}^{n} u_{it}.$$

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

If assume

$$\text{Rank}\left(\bar{C}\right) = r \leq q+1, \text{ for all } n \tag{5.6}$$

it follows that

$$f_t = \left(\bar{C}\bar{C}'\right)^{-1} \bar{C} \left(\bar{z}_t - \bar{B} - \bar{u}_t\right).$$

Therefore, f_t can be approximated by a linear combination of $\{\bar{z}_t,\ 1\}$, if $\bar{u}_t \overset{q.m.}{\to} 0$, as $n \to \infty$, see Lemma 1 in Pesaran (2006). In such a case, we obtain

$$f_t - (CC')^{-1} C \left(\bar{z}_t - \bar{B}\right) \overset{p}{\to} 0,\ \text{as } n \to \infty$$

where

$$\bar{C} \overset{p}{\to} C = \tilde{\Gamma} \begin{pmatrix} 1 & 0 \\ \beta & I_k \end{pmatrix},\ \text{as } n \to \infty$$

$\tilde{\Gamma} = \begin{pmatrix} E(\gamma_i) & E(\Gamma_i) \end{pmatrix} = (\gamma,\ \Gamma)$.

From (5.1) and (5.2), y_{it} is generated as

$$y_{it} = x'_{it}\beta + \alpha_i + \gamma'_i f_t + \varepsilon_{it} \tag{5.7}$$

Next, substitute $f_t = (CC')^{-1} C \left(\bar{z}_t - \bar{B}\right)$ into eq(5.7),

$$\begin{aligned} y_{it} &= x'_{it}\beta + \alpha_i + \gamma'_i (CC')^{-1} C \left(\bar{z}_t - \bar{B}\right) + \varepsilon_{it} \qquad (5.8) \\ &= x'_{it}\beta + \left(\alpha_i - \gamma'_i (CC')^{-1} C\bar{B}\right) + \gamma'_i (CC')^{-1} C\bar{z}_t + \varepsilon_{it} \\ &= x'_{it}\beta + \bar{h}'_t c_i + \varepsilon_{it} \end{aligned}$$

where $c_i = \begin{pmatrix} \alpha_i - \gamma'_i (CC')^{-1} C\bar{B} & \gamma'_i (CC')^{-1} C \end{pmatrix}'$ is $(q+2) \times 1$, $\bar{h}_t = (1,\ \bar{z}'_t)'$ is $(q+2) \times 1$. This suggest using $\bar{h} = (1,\ \bar{z}'_t)'$ as observable proxies for f_t. In vector notation,

$$y_i = X_i\beta + \bar{H}c_i + \epsilon_i \tag{5.9}$$

where $\bar{H} = \left(\iota_T,\ \bar{Z}\right)$ is $T \times (q+2)$, $\bar{Z} = (\bar{z}_1, \ldots, \bar{z}_T)$ is $T \times (q+1)$.

Let

$$\bar{M} = I_T - \bar{H}\left(\bar{H}'\bar{H}\right)^{-1} \bar{H}'$$

where $\bar{H} = \left(\iota_T,\ \bar{Z}\right)$, $\bar{Z} = (\bar{z}_1, \ldots, \bar{z}_T)$ is $T \times (q+1)$. $\bar{M}\bar{H} = 0$. Then

$$\bar{M}y_i = \bar{M}X_i\beta + \bar{M}\epsilon_i$$

Now we make the following assumption:

Assumption 5.7: $\sum_{i=1}^{n}\left(\frac{X_i'\bar{M}X_i}{n}\right)$ is bounded and nonsingular.

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

$$\hat{\beta}_{CCEP} = \left(\sum_{i=1}^{n} X_i'\bar{M}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'\bar{M}y_i\right)$$

Following Pesaran (2006), for fixed T, and $n \to \infty$, the asymptotic for CCEP estimator still holds. Under *Assumptions* 1-5 and 7, and the rank condition (5.6) satisfied

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

where $\Sigma_p = \Psi^{-1}R\Psi^{-1}$, $R = \text{plim}\big(\sigma^2 \frac{1}{n}\sum_{i=1}^{n} X_i'M_gX_i + \frac{1}{n}\sum_{i=1}^{n} X_i'M_gF\Omega_\eta F'M_gX_i\big)$, $\Psi^{-1} = \text{plim}\big(\frac{1}{n}\sum_{i=1}^{n} X_i'M_gX_i\big)$, $M_g = I - G\left(G'G\right)^{-1}G'$, $G = \left(\iota_T,\ F\right)$.

Remark 21 *The above DGP has many similarities with the DGP considered by Pesaran (2006); however, they are not the same. The idea is to have a simple and transparent DGP that is consistent with the typical assumptions of the literature.*

5.2.3 The Combined Estimator

In the panel data models, when the standard assumption of cross-sectionally uncorrelated errors is violated, the usual FE model does not produce consistent estimates of the coefficients of interest. Pesaran (2006) suggests the CCEP approach that yields consistent estimates in the presence of correlated unobserved common effects. We propose the following combined estimator of β, which is a weighted combined FE and CCEP estimator with weights depending on Hausman statistic:

$$\hat{\beta}_c = w\hat{\beta}_{FE} + (1-w)\hat{\beta}_{CCEP} \tag{5.10}$$

where

$$
\begin{aligned}
w &= \begin{cases} \frac{\tau}{H_n} & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases} \\
H_n &= (\hat{\beta}_{CCEP} - \hat{\beta}_{FE})'\hat{V}(\hat{\beta}_{CCEP} - \hat{\beta}_{FE})^{-1}(\hat{\beta}_{CCEP} - \hat{\beta}_{FE})
\end{aligned}
$$

where τ is a shrinkage parameter. The degree of shrinkage depends on the ratio τ/H_n. When $H_n < \tau$ then $\hat{w} = \hat{\beta}_{FE}$, 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 $\hat{\beta}_{FE}$ when τ/H_n is large.

5.3 Asymptotic Distribution

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

$$\gamma_i = \frac{1}{\sqrt{n}}\delta_i \tag{5.11}$$

where δ_i is a $r \times 1$ localizing parameter, which is the degree of correlation between x_i and e_i. When $\delta_i = 0$, x_{it} are exogenous. When $\delta_i \neq 0$, x_{it} are endogenous. δ_i controls the degree of endogeneity.

Set $S = \mathbb{E}\left(X_i'Q_T F\delta_i\right)$ and further assume,

Assumption 5.8. X_i, $i = 1, \ldots, n$, are iid. over i. $E \parallel x_{it} \parallel^{2+K} < \infty$ for some $K > 0$. $\mathbb{E} \parallel x_{it} \parallel^4 < \infty$; $\sigma^2 \left(\text{plim} \sum_{i=1}^{n} \frac{X_i'Q_T X_i}{n}\right)^{-1} = V_1$, $\left(\text{plim} \frac{\Psi^{-1}R\Psi^{-1}}{n}\right)^{-1} = V_2$, as $n \to \infty$.

Theorem 22 *Under Assumptions 5.1-5.8,*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_{FE} - \beta \\ \hat{\beta}_{CCEP} - \beta \end{pmatrix} \to h + \xi \tag{5.12}$$

where

$$h = \begin{pmatrix} \sigma^{-2}V_1 S \\ 0 \end{pmatrix} \tag{5.13}$$

and $\xi \sim N(0, V)$ *with*

$$V = \begin{pmatrix} V_1 & V_{21}' \\ V_{21} & V_2 \end{pmatrix}$$

Furthermore,

$$H_n \to (h+\xi)' B(h+\xi) \tag{5.14}$$

and

$$\sqrt{n}\left(\hat{\beta}_c - \beta\right) \xrightarrow{d} \tilde{\Psi} = G_2'\xi - \left(\frac{\tau}{(h+\xi)' B(h+\xi)}\right)_1 G'(h+\xi) \tag{5.15}$$

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

Theorem 22 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. The asymptotic distribution of $\hat{\beta}_c$ is written as a random weighted average of the asymptotic distribution of $\hat{\beta}_{FE}$ and $\hat{\beta}_{CCEP}$.

5.4 Asymptotic Risk

The asymptotic risk of any sequence of estimators β_n of β can be defined as

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

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

$$\begin{aligned} \lambda_1 &= \lambda_{\max}\left(\frac{A+A'}{2}\right) \\ \lambda_1^* &= \lambda_{\max}\left(\frac{A^*+A^{*\prime}}{2}\right) \end{aligned}$$

where

$$
\begin{aligned}
A &= \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}} W \left(V_2 - V_{21}\right) \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-\frac{1}{2}}, \\
A^* &= \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}} W \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{\frac{1}{2}}.
\end{aligned}
$$

Let

$$d = \frac{tr\left(W\left(V_2 - V_{21}\right)\right)}{\lambda_1}.$$

Theorem 23 *Under Assumptions 5.1-5.8, if*

$$d > 2 \tag{5.16}$$

and

$$0 < \tau \leq \frac{2\lambda_1 (d-2)}{\lambda_1^*}, \tag{5.17}$$

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

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

and

$$R\left(\hat{\beta}_c\right) < R\left(\hat{\beta}_{CCEP}\right) - \frac{\tau\left(2\lambda_1(d-2) - \lambda_1^*\tau\right)}{\sigma^{-4} S' V_1 \left(V_1 + V_2 - \left(V_{21} + V_{21}'\right)\right)^{-1} V_1 S + q}. \tag{5.18}$$

Equation (5.18) 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 (5.17). τ appears in the risk bound (5.18) 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 (5.17) to be positive, which is necessary for the existence of τ satisfying (5.17).

Corollary 24 $R\left(\hat{\beta}_c\right) - R\left(\hat{\beta}_{CCEP}\right) < 0$, *for* $d > 2$ *and* $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$. *In the case* $W = \left(V_2 - V_{21}\right)^{-1}$, $0 < \tau \leq 2\left(\frac{q-2}{\lambda_1^*}\right)$ *and* $q > 2$ *which is Stein's (1956) classic condition for shrinkage.*

Corollary 25 $R\left(\hat{\beta}_{FE}\right) = tr(WV_1) + \sigma^{-4}S'V_1WV_1S$; $R\left(\hat{\beta}_{FE}\right) \leq R\left(\hat{\beta}_{CCEP}\right)$ *when* $\sigma^{-4}S'V_1WV_1S \leq q$, *and* $R\left(\hat{\beta}_{RE}\right) > R\left(\hat{\beta}_{FE}\right)$ *otherwise.*

The result in Corollary 25 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 26 $R\left(\hat{\beta}_c\right) - R\left(\hat{\beta}_{FE}\right) < 0$, *for* $q < \sigma^{-4}S'V_1WV_1S$, $d > 2$, *and* $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$.

The result in Corollary 26 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.

5.5 Monte Carlo Simulation

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

$$\begin{aligned} y_{it} &= \alpha_i + \beta' x_{it} + \gamma_i' f_t + \varepsilon_{it} \\ x_{it} &= \Gamma_i' f_t + v_{it} \end{aligned}$$

where α_i is drawn from $N(0,\ 1)$, $\varepsilon_{it} \sim \text{IID} N(0,\ 1)$, $v_{it} \sim \text{IID} N(0,\ 1)$.

We vary $n = \{20, 100\}$, $T = 8$, $\beta = 0$. Set $r = 1$ and $q = 3$. $\tilde{\gamma}_i \sim \text{IID} N(1,\ 0.2)$.
$\Gamma_i = \begin{pmatrix} \Gamma_{i11} & \Gamma_{i21} & \Gamma_{i31} \end{pmatrix} \sim \text{IID} \begin{pmatrix} N(0.5,\ 0.5) & N(0,\ 0.5) & N(0,\ 0.5) \end{pmatrix}$.

We consider ρ on a 40-point grid on $[0, 0.975]$. ρ controls the degree of endogeneity, ranging in $(0,1)$ ($\rho = 0$ is the case of exogenous regressors; large ρ is the case of strong endogeneity). $\gamma_i = \tilde{\gamma}_i \rho$.

We compare the estimator by MSE

$$MSE\left(\hat{\beta}\right) = \mathbb{E}\left(\hat{\beta} - \beta\right)'\left(\hat{\beta} - \beta\right)$$

We normalize the MSE of the estimators by that of the CCEP estimator. Thus value less than one indicate improved precision relative to CCEP estimator, and values greater than one indicate worse performance, larger MSE than the CCEP estimator. We generated 100,000 samples on each calculated $\hat{\beta}_{CCEP}$, $\hat{\beta}_{FE}$, $\hat{\beta}_{c}$.

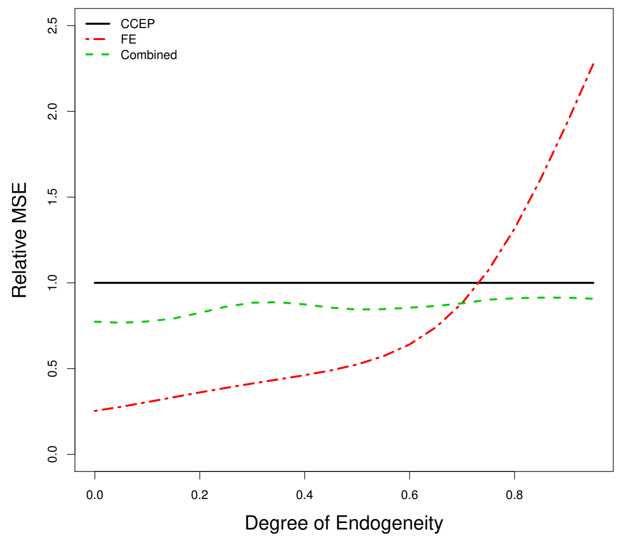
We do a bootstrap pairs procedure that resample with replacement over i and uses all observed time periods for a given individual. For data $\{(y_i,\ X_i)\,,\ i = 1, \ldots, n\}$ this yields B pseduo-samples and for each pseudo-sample we perform regression, yielding B estimates, $b = 1, \ldots, B$. The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

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

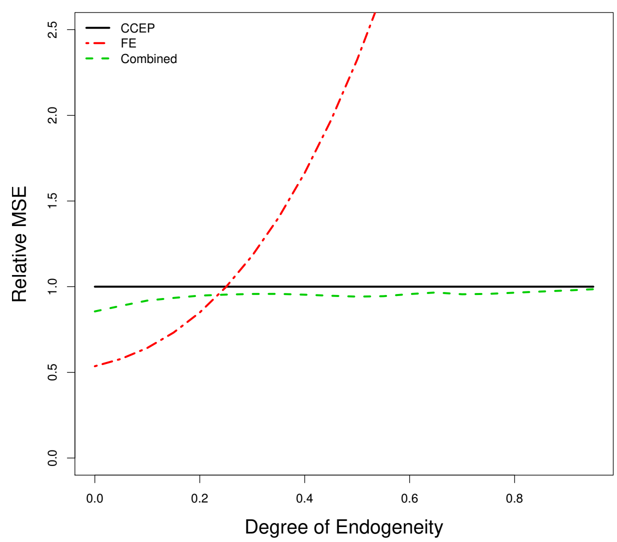
b denotes the bth of B bootstrap replications, and $\hat{\theta} = \hat{\beta}_{CCEP} - \hat{\beta}_{FE}$, $\bar{\hat{\theta}} = B^{-1}\sum_b \hat{\theta}_b$.

Figure 5.1(a) is the case $n = 20$, and Figure 5.1(b) is the case $n = 100$. The solid line is the normalized MSE of the CCEP estimator, the short dashed line is the normalized MSE of the combined estimator, the longer dashed line is the normalized MSE of the FE estimator.

By contrasting Figure 5.1(a) and 5.1(b), we can see that the region of dominance for the combined estimator over CCEP estimator is greater for small n. 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. Generally, the dominance for combined estimator over CCEP estimator is greater for small sample size. For very small ρ, the combined estimator can be beaten by CCEP estimator. In summary, the simulation results provide strong finite sample



(a)



(b)

Figure 5.1: (a) Relative MSE of CCEP, FE and Combined Estimators, $n = 100, T = 8, q = 3, r = 1$. (b) Relative MSE of CCEP, FE and Combined Estimators, $n = 100, T = 8, q = 3, r = 1$.

confirmation of Theorem 23.

5.6 Empirical Results

Real house prices can vary between States because real incomes differ, they can also differ because of scarcity of land or other idiosyncratic factors. The effects of common shocks on house prices such as changes in interest rates, oil prices and technological change, could also differ across States. Holly, Pesaran, and Yamagata (2010), here after HPY, examine the extent to which real house prices at the State level are driven by fundamentals such as real per capita disposable income, as well as by common shocks. Baltagi and Li (2002) replicate the results of HPY, using a slightly different dataset. They extend the period of study to 2011, incorporating the information reflected by the housing market crash in 2007. Using housing price indexes for 381 metropolitan statistical areas and over the period 1975–2011, they find that the HPY results are fairly robust. In this section, we use the panel of 49 states over the period 1975–2011, and following HPY, consider the following panel data model for US states

$$p_{i,t} = \beta_0 + \beta_y y_{i,t} + \beta_g g_{i,t-1} + \beta_c c_{i,t-1} + \alpha_i + e_{it}$$

where $i = 1, \ldots, 48, t = 1, \ldots, 17$, $p_{i,t}$ 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

	FE	CCEP	Combined
Log(real per capita income)	0.5804	1.2705	1.2151
	(0.3013)	(0.2990)	(0.2986)
Population growth rate	1.3286	1.6367	1.6120
	(1.8132)	(1.5217)	(1.4237)
Real cost of borrowing	-0.5088	-0.1781	-0.2047
	(0.1546)	(0.1541)	(0.1530)

Table 5.1: Economics of Real House Prices Estimates for 49 U.S. states, 1975–2011 (standard errors in parentheses)

coefficient for $c_{i,t-1}$. The effect of population growth on real house prices is expected to be positive.

Table 5.1 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 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 columns report the FE and CCEP estimates. The estimates of the combined estimator lies 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.

5.7 Conclusion

This chapter proposes a combined estimator from combining the fixed effects estimator (FE) and the common correlated effects pooled (CCEP)estimator, using weights inversely proportional to the Hausman statistic. We show that the combined estimator has strictly smaller asymptotic 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 CCEP estimator for all degrees of endogeneity, as well as relative to FE estimator for moderate to large degrees of endogeneity. The use of the combined estimator allows researchers to implement robust inference in the error cross-sectional dependence framework. Following Holly, Pesaran, and Yamagata (2010), we use the combined estimator to analyze US real house prices. 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 theirs. Based on the empirical results, the combined estimator gives smaller MSE than both FE and CCEP estimators.

Chapter 6

A Combined Forecasting Approach with Some Empirical Evidence from US Electricity and Natural-gas Consumption

6.1 Introduction

Over the past few years, considerable progress has been made in the area of econometric forecasting. The assessment of forecasts and their uncertainty is of particular interest for policy makers confronted with the task of exploiting all the available information and evaluating the relative accuracy and relevance of forecasts from different sources. This chapter contributes to the literature on forecast uncertainty by investigating the forecast model combination in the panel data model.

Bates and Granger (1969) made the econometric profession aware of the benefits of forecast combination when a limited number of forecasts is considered. 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 et al. (2000, 2005). Pooling forecasts is related to forecast combination and operates a reduction on the space of forecasts.

In this chapter, we propose a combined approach to econometric forecast within a panel data framework. First, we calculate the coefficients based combination weights depending on Hausman test statistic. Second, we show that under endogenity, the forecast combining both fixed effects and random effects models using the weights from step one outperforms forecast with fixed effects 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. It also can reduce finite sample MSFE relative to the random effects forecast for moderate to large degrees of endogeneity and heterogeneity.

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 this study. Maddala et al. (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 et al. (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 fixed effects, random effects and the proposed combined forecasting procedures 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 chapter is organized as follows. Section 2 presents the model and estimators. Monte Carlo simulation is provided in Section 3. Empirical example is given in Section 4. Section 5 concludes.

6.2 Models and Estimators

6.2.1 Forecasting with Random Effects

Consider a panel data regression model

$$y_{it} = x_{it}\beta + \alpha_i + u_{it} \tag{6.1}$$

where $i = 1, 2, \ldots n$ and $t = 1, 2, \ldots T$. x_{it} is the ith observation on q explanatory variables, β is a $q \times 1$ unknown parameter, α_i is known as the individual effect and u_{it} is the random error.

The random effects (RE) model assumes $\alpha_i \sim$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 x_{it} are independent of the α_i and u_{it} for all i and t. Under this assumption, we can write

$$y_{it} = x_{it}\beta + v_{it},\ \mathbb{E}\left(v_{it}|x_i\right) = 0 \tag{6.2}$$

where $v_{it} = \alpha_i + u_{it}$. Write the model (6.2) in matrix form

$$y = X\beta + v \tag{6.3}$$

where $y = (y_{11}, \ldots, y_{1T}, y_{21}, \ldots, y_{nT})'$ 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 be a vector of ones. $J_T = \iota_T \iota_T'$. $P = I_n \otimes \bar{J}_T$ where $\bar{J}_T = J_T/T$. and $Q = I_{nT} - P$ is a matrix which obtains the deviations from individual means. The variance-covariance matrix of v is given by

$$\Omega = \sigma_\alpha^2 (I_n \otimes J_T) + \sigma_u^2 (I_n \otimes I_T) = \sigma_1^2 P + \sigma_u^2 Q$$

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

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

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 $\bar{y}_i = \sum_{t=1}^{T} y_{it}/T$, we get

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

Noting that $\hat{\sigma}_1^2$ and $\hat{\sigma}_u^2$ estimate $\hat{\sigma}_\alpha^2 = \frac{1}{T}\left(\hat{\sigma}_1^2 - \hat{\sigma}_u^2\right)$. With these estimates, one can obtain the generalized least squares (GLS) of β based on (6.3) is

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

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

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

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

Suppose we want to predict S periods ahead for the ith individual. First, by minimizing

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

we can obtain

$$\hat{\alpha}_i = \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_1^2} T \bar{\hat{u}}_{i(RE)}$$

where $\bar{\hat{u}}_{i(RE)} = \frac{1}{T}\sum_t \hat{u}_{it(RE)}$. Then the S period ahead forecast for the ith individual is

$$\hat{y}_{i,T+S,RE} = x_{i,T+S}\hat{\beta}_{RE} + \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_1^2} T \bar{\hat{u}}_{i(RE)} \tag{6.4}$$

where $\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_1^2}\sum_t \hat{u}_{it(RE)}$ can be treated as $\hat{\alpha}_{i,RE}$. Goldberger (1962) showed that the best linear unbiased predictor of $y_{i,T+S}$ is

$$\hat{y}_{i,T+S,RE} = x_{i,T+S}\hat{\beta}_{RE} + w'\Omega^{-1}\hat{v}_{RE}$$

where $\hat{v}_{RE} = y - X\hat{\beta}_{RE}$ and $w = E\left(v_{i,T+1}v\right)$. Note that for period $T+S$

$$v_{i,T+S} = \alpha_i + u_{i,T+S}$$

and $w = \sigma_\alpha^2\left(l_i \otimes \iota_T\right)$ where l_i is the ith column of I_N. In this case

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

since $\left(l_i' \otimes \iota_T'\right)P = \left(l_i' \otimes \iota_T'\right)$ and $\left(l_i' \otimes \iota_T'\right)Q = 0$. The typical element of $w'\Omega^{-1}\hat{v}_{RE}$ becomes $\left(\frac{T\hat{\sigma}_\alpha^2}{\hat{\sigma}_1^2}\bar{\hat{u}}_{i(RE)}\right)$ which is $\bar{\hat{u}}_{i(RE)}$.

6.2.2 Forecasting with Fixed Effects

For the fixed effects (FE) case, the α_i are assumed to be fixed parameters to be estimated and remainder disturbances stochastic with $u_{it} \sim$i.i.d.$(0, \sigma_u^2)$. One can premultiply the model by Q and perform OLS on the resulting transformed model:

$$Qy = QX\beta + Qu$$

The resulting OLS estimators are

$$\begin{aligned} \hat{\beta}_{FE} &= \left(X'QX\right)^{-1} X'Qy, \\ \hat{\alpha}_{FE} &= \left(D'D\right)^{-1} D' \left(y - X\hat{\beta}_{FE}\right). \end{aligned} \tag{6.5}$$

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

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

where $V_{FE} = \sigma_u^2 \left(\text{plim}\frac{X'QX}{n}\right)^{-1}$. From (6.5), we know that for the ith individual, $\hat{\alpha}_{i,FE} = \bar{y}_i - \bar{x}_i\hat{\beta}_{FE}$. Thus, the S period ahead forecast for the ith individual is

$$\hat{y}_{i,T+S,FE} = \bar{y}_i + \left(x_{i,T+S} - \bar{x}_i\right)\hat{\beta}_{FE} \tag{6.6}$$

Alternatively,

$$\hat{y}_{i,T+S,FE} = x_{i,T+S}\hat{\beta}_{FE} + \bar{\hat{u}}_{i(FE)}$$

6.2.3 The Combined Forecast

We have showed that the combined estimator from combining the fixed and random effects estimators with weights depending on Hausman test statistic, performs better than the fixed effects estimator, as well as compared to the random effects estimator under certain conditions. Motivated by this observation, we would like to see if the use of both procedures can result in an improved forecast. We combine $\hat{y}_{i,T+S,RE}$ and $\hat{y}_{i,T+S,FE}$ using the weight inversely proportional to the Hausman statistic for exogeneity.

$$\hat{y}_{i,T+S,c} = w_c\hat{y}_{i,T+S,RE} + \left(1 - w_c\right)\hat{y}_{i,T+S,FE} \tag{6.7}$$

where

$$\begin{aligned}
w_c &= \begin{cases} \frac{\tau}{H_n} & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases} \\
H_n &= n\left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)' \left(\hat{V}_{FE} - \hat{V}_{RE}\right)^{-1} \left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)
\end{aligned}$$

τ is a shrinkage parameter which controls the degree of shrinkage. Set $\tau = q - 2$ when $q > 2$.

6.3 Monte Carlo Simulation

In this section, we summarize the main results from a small Monte Carlo experiment designed to illustrate the finite sample properties of the combined forecast defined in section 2. To compare the prediction procedures, we calculate the 1-step ahead out of sample mean squared forecast error (MSFE) of each approach. The $T+1$ forecast error is defined as

$$e_{T+1} = Y_{T+1} - \hat{Y}_{T+1}$$

and

$$MSFE\left(e_{T+1}\right) = \mathbb{E}\left(e'_{T+1} e_{T+1}\right)$$

We consider the following data generating process

$$\begin{aligned}
y_{it} &= x_{it}\beta + \alpha_i + u_{it} \\
\alpha_i &= \rho\sqrt{T}\bar{x}_i \frac{\iota}{q} + \sqrt{1-\rho^2}\ \epsilon_i
\end{aligned}$$

where $\{x_{it},\ u_{it}\}$ are i.i.d $N\left(0,\ I_{q+1}\right)$ across $i,\ t$. ϵ_i are i.i.d $N\left(0,\ 1\right)$ independent of $\{x_{it},\ u_{it}\}$.

We use a portion of the available data for testing, and use the rest of the data for estimating (or "training") the model. Then the testing data can be used to measure how well the model is likely to forecast on new data. The process works as follows: First, use the observations at times $1, 2, \ldots, T-1$ to estimate the forecasting model. Compute the 1-step error on the forecast for time T. Second, compute the forecast accuracy measures based on the errors obtained.

We set $\beta = 0$, $T = 5$ and $q = 4$. We also vary $n = \{20, 100\}$ and ρ on a 40-point grid on $[0, 0.975]$. ρ controls the degree of endogeneity, ranging in $(0, 1)$. We also set $\sqrt{\theta} = \frac{\sigma_\alpha}{\sigma_u} \in \left\{\frac{1}{4}, 1, 4\right\}$ so $\rho^* = \frac{\theta}{1+\theta} = \{.06,\ .05,\ .94\}$. ρ^* controls the degree of heterogeneity.

We generated 10,000 samples on each calculated $\hat{Y}_{T+1,FE}$, $\hat{Y}_{T+1,RE}$, $\hat{Y}_{T+1,C}$ and plot the relative MSFE, that is

$$\frac{\mathbb{E}\left[\left(\hat{Y}_{T+1} - Y\right)'\left(\hat{Y}_{T+1} - Y\right)\right]}{\mathbb{E}\left[\left(\hat{Y}_{T+1,FE} - Y\right)'\left(\hat{Y}_{T+1,FE} - Y\right)\right]}$$

Thus value less than one indicate improved precision relative to FE forecast, and values greater than one indicate worse performance, larger MSFE than the FE forecast.

Figure 6.1(a), 6.1(c), 6.1(e) are the cases $n = 20$, and Figure 6.1(b), 6.1(d), 6.1(f) are the cases $n = 100$. The region of dominance for the combined forecast over FE forecast is greater for small n.

Figure 6.1(a) and Figure 6.1(b) are the cases $\rho^* = 0.5$. For small ρ the RE forecast has lower MSFE than the combined forecast, but the ranking is reversed for larger values of ρ. Figure 6.1(c) and Figure 6.1(d) are the cases $\rho^* = 0.06$. The MSFE of the RE forecast is smaller than that of the combined forecast for all the values of ρ. The MSFE of the RE and combined forecasts are uniformly smaller than that of the FE forecast. Figure 6.1(e) and Figure 6.1(f) are the cases $\rho^* = 0.94$. The

FE and combined forecasts are near equivalents. RE forecast has similar MSFE to FE and combined forecasts for small ρ, but the MSFE of RE forecast increases dramatically after intermediate values of ρ. In all the cases, the combined forecast uniformly dominates the FE forecast.

In summary, The improvements in combined forecast over FE forecast are greater for smaller heterogeneity ρ^*. For very small ρ^*, RE forecast tends to be better than both FE and combined forecasts. For moderate to large ρ^* and higher ρ, or moderate to large ρ and higher ρ^*, the combined forecast is better than RE forecast. For very large ρ^* and low ρ, the combined forecast is close to RE forecast.

6.4 Empirical Results

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. 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 a 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 per capital personal income 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:

$$\hat{y}_{i,T+1} = \beta_0 + \hat{\beta}_{i,1}x_{1i,T+1} + \hat{\beta}_{i,2}x_{2i,T+1} + \hat{\beta}_{i,3}x_{3i,T+1} + \hat{\alpha}_i$$

where $i = 1, \ldots, 51$, $t = 1, \ldots, 14$.

The variables for the electricity regression are $y_{i,t} = log$(residential electricity per capita consumption), $x_{1i,t} = log$(real per capita personal income), $x_{2i,t} = log$(real residential electricity price), $x_{3i,t} = log$(real residential natural-gas price). For the natural-gas regression, we have $y_{i,t} = log$(residential natural gas per capita consumption), $x_{1i,t} = log$(real per capita personal income), $x_{2i,t} = log$(real residential natural gas price), $x_{3i,t} = log$(real residential electricity price).

We use the prediction performance criteria to help us choose among alternative estimators. 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. Table 6.1 gives a comparison of forecasts using the root mean square errors criterion criterion (RMSE) for residential electricity demand while Table 6.2 does the same for residential natural-gas demand. Because of the ability of an estimator to characterize long-run as well as short-run responses is at issue, the RMSE is calculated across the 51 states at different forecast horizons. The relative forecast rankings are reported in Tables 6.1 and 6.2 after 1 and 3 years.

For electricity demand (Table 6.1), the relatively weak performance of RE estimator arises because of the endogeneity problem of regression. Similarly, the weak forecast performance of the FE estimator relative to the RE estimator can be attributed to the weak endogeneity issue. Thus, it seems an advantage to use the combined

Ranking	1st year		3th year	
	Approach	RMSE	Approach	RMSE
1	Combined	5.2837	Combined	6.1467
2	FE	5.5261	FE	6.4135
3	RE	6.6025	RE	7.9441

Table 6.1: Comparison of Forecast Performance for US Electricity Demand (standard errors in parentheses)

Ranking	1st year		3th year	
	Approach	RMSE	Approach	RMSE
1	Combined	5.7310	Combined	7.4145
2	FE	6.1593	FE	8.1922
3	RE	7.1044	RE	9.0036

Table 6.2: Comparison of Forecast Performance for US Natural-gas Demand (standard errors in parentheses)

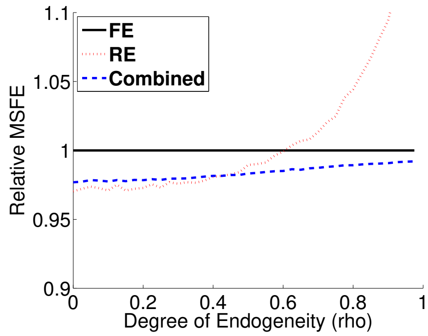
forecast. The overall RMSE forecast rankings offer a strong endorsement for the combined forecast. For natural-gas demand (Table 6.2), the top ranked estimators is the combined estimator whether it is for the 1 year or 3 year forecast performance. The combined estimator ranks first, followed by the FE and RE estimator.

Both residential electricity and natural gas demand RMSE forecast rankings offer an endorsement for the combined forecast based on their out-of-sample forecast performance.

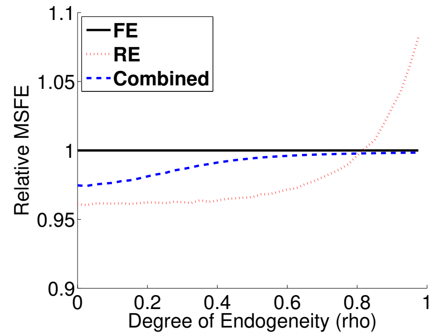
6.5 Conclusion

This chapter provides a combined forecasting approach from combining fixed effects and random effects forecasts, with the weights depending inversely on the Hausman statistic. We show that forecasting can benefit from the use of both procedures.

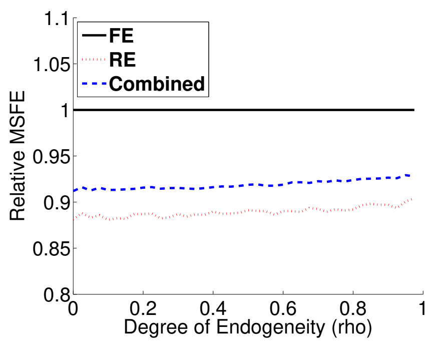
Our simulation experiment shows that the combined forecast can uniformly dominate the fixed effects forecast for all degrees of endogeneity. It also can reduce finite sample MSFE relative to the random effecys forecast for moderate to large degrees of endogeneity and heterogeneity. We examine the applicability of the combined forecasting approach using US panel data sets on residential electricity and natural-gas demand across 51 states (include DC) over the period 1997–2012. Our results show that the combined forecast offers the best out-of-sample forecasts.



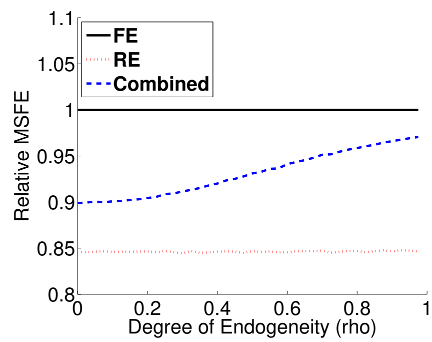
(a) $n = 20$, $T = 5$, $q = 4$, $\rho^* = .5$



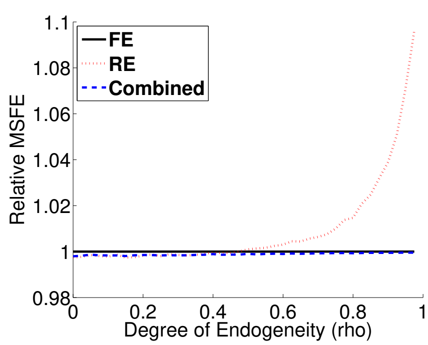
(b) $n = 100$, $T = 5$, $q = 4$, $\rho^* = .5$



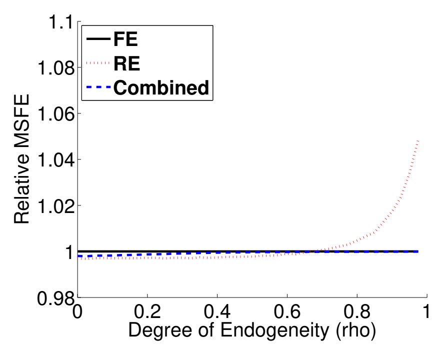
(c) $n = 20$, $T = 5$, $q = 4$, $\rho^* = .06$



(d) $n = 100$, $T = 5$, $q = 4$, $\rho^* = .06$



(e) $n = 20$, $T = 5$, $q = 4$, $\rho^* = .94$



(f) $n = 100$, $T = 5$, $q = 4$, $\rho^* = .94$

Figure 6.1: Relative One-Step MSFE of FE, RE and Combined Estimators

Bibliography

[1] Amemiya, T. and MaCurdy, T. E. (1986). "Instrumental-variable estimation of an error-components model. Econometrica: *Journal of the Econometric*, 869-880.

[2] Balestra, P. and Nerlove, M. (1966). "Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas." *Econometrica*, 34, 585-612.

[3] Baltagi, B. H. (2006). Estimating an economic model of crime using panel data from North Carolina. *Journal of Applied Econometrics*, 21(4), 543-547.

[4] Baltagi, B. (2008). *Econometric analysis of panel data*. John Wiley & Sons.

[5] Baltagi, B.H., Bresson, G. and 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.

[6] Baltagi, B.H. and Li, D.(2002). "Series estimation of partially linear panel data models with fixed effects. " *Annals of economics and finance*, 3(1), 103-116.

[7] Baltagi, B. H. and 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.

[8] Baltagi, B.H. and Pinnoi N. (1995). "Public capital stock and state productivity growth: further evidence from an error components model." *Empirical Economics*, 20(2), 351–359.

[9] Bates, J.M. and Granger, C.W. (1969) "The combination of forecasts." *Journal of the Operational Research Society*, 20(4), 451-468.

[10] Bayley, D. H. (1996). *Police for the Future*. Oxford University Press.

[11] Breusch, T. S., Mizon, G. E., & Schmidt, P. (1989). Efficient estimation using panel data. Econometrica: *Journal of the Econometric*, 695-700.

[12] Chamberlain, G. (1984). "Panel data." *Handbook of econometrics*, 2, 1247-1318.

[13] Chudik, A., M.H. Pesaran, and E. Tosetti (2011) "Weak and Strong Cross-Section Dependence and Estimation of Large Panels." *The Econometrics Journal*, 14, C45–C90.

[14] Chudik, A. and M.H. Pesaran (2015) "Common Correlated Effects Estimation of Heterogeneous Dynamic Panel Data Models with Weakly Exogenous Regressors." *Journal of Econometrics*, 188, 393–420, heterogeneity in Panel Data and in Nonparametric Analysis in honor of Professor Cheng Hsiao.

[15] Cook, P. J. (1979). The clearance rate as a measure of criminal justice system effectiveness. *Journal of Public Economics*, 11(1), 135-142.

[16] Cornwell, C., & Trumbull, W. N. (1994). Estimating the economic model of crime with panel data. *The Review of economics and Statistics*, 360-366.

[17] Ehrlich, I. (1972). The deterrent effect of criminal law enforcement. The Journal of Legal Studies, 1(2), 259-276.

[18] Evans, P. and Karras G. (1994). "Are government activities productive? Evidence from a panel of US states." *The Review of Economics and Statistics,* 76(1), 1–11.

[19] Forni, M., Hallim, M. and Lippi, M., Reichlin, L. (2000) "The generalized dynamic factor model: Identification and estimation." *Review of Economics and Statistics* 82, 540 554.

[20] Forni, M., Hallim, M. and Lippi, M., Reichlin, L. (2005) "The generalized dynamic factor model: One-sided estimation and forecasting." *Journal of the American Statistical Association* 100, 830 840.

[21] Hansen, B.E. (2014). "A Stein-Like 2SLS Estimator." *University of Wisconsin, Madison.*

[22] Hansen, B.E. (2016). "Efficient shrinkage in parametric models." *Journal of Econometrics,* 190(1), 115-132.

[23] Hausman, J.A. (1978). "Specification tests in econometrics."*Econometrica*, 46, 1251-1271.

[24] Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. Econometrica: *Journal of the Econometric*, 1377-1398.

[25] Henderson, D. and Ullah A. (2008). "Nonparametric estimation in a one-way error component model: a Monte Carlo analysis." *Riverside: University of California.*

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

[27] Holtz-Eakin, D. (1994). "Public-sector capital and the productivity puzzle." *The Review of Economics and Statistics*, 76(1), 12-21.

[28] Horowitz, J.L. and Markatou, M. (1996). "Semiparametric estimation of regression models for panel data. " *The Review of Economic Studies*, 63(1), 145-168.

[29] Hsiao, C. (2014). *Analysis of panel data*, Vol. 54. 1, Cambridge university Press.

[30] Hyslop, Dean R. (1999). "State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women." *Econometrica* 67(6), 1255-1294.

[31] Huang, B. (2015). "A Combined fixed and random effects estimator for parametric panel model" *University of California, Riverside*.

[32] James, W. and Stein, C. (1961). "Estimation with quadratic loss." *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, 1, 361-379.

[33] Kapetanios, G., M.H. Pesaran, and T. Yamagata (2011) "Panels with Non-Stationary Multifactor Error Structures." *Journal of Econometrics*, 160, 32 –348.

[34] Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. *The American Economic Review*, 92(4), 1244-1250.

[35] Li, Q. and Hsiao, C. (1998). "Testing serial correlation in semiparametric panel data models." *Journal of econometrics*, 87(2), 207-237.

[36] Li, Q. and Ullah, A. (1998). "Estimating partially linear panel data models with one-way error components." *Econometric reviews*, 17(2), 145-166.

[37] Maddala, G. S. (1971). "The use of variance components models in pooling cross section and time series data.." *Econometrica*, 39, 341-358.

[38] Maddala, G.S., Trost, R.P., Li, H. and Joutz, F. (1997) "Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators." J*ournal of Business and Economic Statistics,* 15, 90–100.

[39] Marvell, T. B., & Moody, C. E. (1996). Specification problems, police levels, and crime rates. *Criminology*, 34(4), 609-646.

[40] Mundlak, Y. (1961). "Empirical production function free of management bias." *Journal of Farm Economics*, 43(1), 44-56.

[41] Munnell, A.H. (1990). "How does public infrastructure affect regional economic performance?" *New England Economic Review,* 11-33.

[42] Nerlove, M. (1971). "Further evidence on the estimation of dynamic economic relations from a time series of cross sections." *Econometrica*, 39, 359-382.

[43] Olsen, M. K. and Schafer, J. L. (2001). "A two-part random-effects model for semicontinuous longitudinal data." *Journal of the American Statistical Association*, 96(454), 730-745.

[44] Pedroni, P. (2007). "Social capital, barriers to production and capital shares: implications for the importance of parameter heterogeneity from a nonstationary panel approach." *Journal of Applied Econometrics*, 22(2), 429-451.

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

[46] Pesaran, M.H., L.V. Smith and T. Yamagata (2013) "A panel unit root test in the presence of a multifactor error structure." *Journal of Econometrics* 175, 94–115.

[47] Pesaran, M.H. and E. Tosetti (2011) "Large panels with common factors and spatial correlations." *Journal of Econometrics* 161, 182–202.

[48] Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005). "Teachers, schools, and academic achievement." *Econometrica*, 73(2), 417-458.

[49] Rockoff, J. E. (2004). "The impact of individual teachers on student achievement: Evidence from panel data." *The American Economic Review*, 94(2), 247-252.

[50] Scheike, T. H., Sun, Y., Zhang, M. J., and Jensen, T. K. (2010). "A semiparametric random effects model for multivariate competing risks data." *Biometrika*, 97(1), 133-145.

[51] 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* Vol.1, 399.

[52] Stock, J. and Watson, M. (1999) "Forecasting inflation." *Journal of Monetary Economics*, 44, 293 335.

[53] Stock, J. and Watson, M. (2002a) "Macroeconomic forecasting using diffusion indexes." *Journal of Business and Economic Statistics*, 20, 147 162.

[54] Stock, J. and Watson, M. (2002b) "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association*, 460, 1167 1179.

[55] Su, L. and Ullah, A. (2006). "Profile likelihood estimation of partially linear panel data models with fixed effects." *Economics Letters*, 92(1), 75-81.

[56] Su, L. and Ullah, A. (2007). "More efficient estimation of nonparametric panel data models with random effects." *Economics Letters*, 96(3), 375-380.

[57] Ullah, A. and Roy, N. (1998). "Nonparametric and semiparametric econometrics of panel data." *STATISTICS TEXTBOOKS AND MONOGRAPHS*, 155, 579-604.

[58] Wallace, T. D. and Hussain, A. (1969). "The use of error components models in combining cross section with time series data." *Econometrica*, 37, 55-72.

[59] Wang, Y., Zhang, Y. and Zhou, Q. (2016). "A Stein-like estimator for linear panel data models." *Economics Letters*, 141, 156-161.

Appendix A

Appendix for Chapter 1

Proof of Theorem 2: The proof technique is based on the arguments in Theorem 2 of Hansen (2014).

Noting that $\sqrt{n}\left(\hat{\beta}_{FE}-\beta\right)\xrightarrow{d} G_2'\xi \sim N\left(0,\ V_2\right)$, then

$$R\left(\hat{\beta}_{FE}\right)=\mathbb{E}\left(\xi' G_2' W G_2' \xi\right)=\operatorname{tr}\left(W V_2\right)$$

Define Ψ^* as a random variable without positive part trimming in (2.12)

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

Then using (2.13) and the fact that the pointwise quadric risk of Ψ is strictly smaller than that of Ψ^*, then we have

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

We can calculate that

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

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

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

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

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

Therefore,

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

Since

$$GWG_2'V = WG_2'VG = W\left(V_2 - V_1\right)$$

and

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

Then

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

Thus

$$\begin{aligned}
\mathbb{E}\left(\psi^{*\prime}W\psi^*\right) \;=\; & R\left(\hat{\beta}_{FE}\right) + \tau^2\mathbb{E}\left(\frac{(h+\xi)' GWG'(h+\xi)}{\left((h+\xi)' B(h+\xi)\right)^2}\right) \\
& +4\tau\mathbb{E}\mathrm{tr}\left(\frac{(h+\xi)'GWG'(h+\xi)}{\left((h+\xi)' B(h+\xi)\right)^2}\right) \\
& -2\tau\mathbb{E}\mathrm{tr}\left(\frac{\left(W\left(V_2 - V_1\right)\right)}{(h+\xi)' B(h+\xi)}\right)
\end{aligned}$$

Define

$$\begin{aligned} B_1 &= (V_2 - V_1)^{-\frac{1}{2}} G' \\ A^* &= (V_2 - V_1)^{\frac{1}{2}} W (V_2 - V_1)^{\frac{1}{2}} \end{aligned}$$

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

$$\begin{aligned} \operatorname{tr}(B(h+\xi)(h+\xi)'GWG_2'V) &= (h+\xi)'B_1'A^*B_1(h+\xi) \qquad \text{(A.1)} \\ &\leq (h+\xi)'B(h+\xi)\lambda_1 \end{aligned}$$

Using equation (A.1) and Jensen's inequality, we have

$$\begin{aligned} \mathbb{E}(\psi^{*\prime}W\psi^*) &\leq R\left(\hat{\beta}_{FE}\right) + \left(\tau^2 + 4\tau\right)\mathbb{E}\left(\frac{\lambda_1}{(h+\xi)'B(h+\xi)}\right) \\ &\quad -2\tau\mathbb{E}\mathrm{tr}\left(\frac{(W(V_2 - V_1))}{(h+\xi)'B(h+\xi)}\right) \qquad \text{(A.2)} \\ &= R\left(\hat{\beta}_{FE}\right) - \mathbb{E}\left(\frac{\tau\left(2\left(\operatorname{tr}(W(V_2 - V_1)) - 2\lambda_1\right) - \tau\lambda_1\right)}{(h+\xi)'B(h+\xi)}\right) \\ &\leq R\left(\hat{\beta}_{FE}\right) - \frac{\tau\left(2\left(\operatorname{tr}(W(V_2 - V_1)) - 2\lambda_1\right) - \tau\lambda_1\right)}{\mathbb{E}\left((h+\xi)'B(h+\xi)\right)} \qquad \text{(A.3)} \end{aligned}$$

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

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

Substituted into (A.3) we have

$$R\left(\hat{\beta}_c\right) < R\left(\hat{\beta}_{FE}\right) - \frac{\tau\left(2\left(\operatorname{tr}(W(V_2 - V_1)) - 2\lambda_1\right) - \tau\lambda_1\right)}{\sigma_1^{-4}\delta'\bar{X}'\bar{X}V_1(V_2 - V_1)^{-1}V_1\bar{X}'\bar{X}\delta + q}$$

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

Appendix B

Appendix for Chapter 2

Proof of Theorem 7:

The randome effects estimator is given as:

$$\hat{\beta}_{SP,RE} = \left(X^{*\prime}\Omega^{-1}X^{*}\right)^{-1} X^{*\prime}\Omega^{-1}y^{*}$$

Since

$$\Omega^{-1} = \sigma_u^{-2}\left(\lambda P + Q\right) = \frac{P}{\sigma_1^2} + \frac{Q}{\sigma_u^2}, \quad \text{with } \lambda = \frac{\sigma_u^2}{\sigma_1^2}$$

we can write $\hat{\beta}_{SP,RE}$ as

$$\hat{\beta}_{SP,RE} = \left(X^{*\prime}\left(\lambda P + Q\right)X^{*}\right)^{-1} X^{*\prime}\left(\lambda P + Q\right)y^{*} = Ay^{*}$$

where $A \equiv \left(X^{*\prime}\left(\lambda P + Q\right)X^{*}\right)^{-1} X^{*\prime}\left(\lambda P + Q\right)$, then

$$\begin{aligned}
\hat{\beta}_{SP,RE} &= A\left(X^{*}\beta + D\alpha + u\right) \\
&= \beta + A\left(D\alpha + u\right)
\end{aligned}$$

From (3.13),

$$\alpha = \bar{X}\frac{\delta}{\sqrt{n}} + \epsilon$$

Then

$$
\begin{aligned}
\hat{\beta}_{SP,RE} - \beta &= A\left(D\bar{X}\frac{\delta}{\sqrt{n}} + D\epsilon + u\right) \\
\sqrt{n}\left(\hat{\beta}_{SP,RE} - \beta\right) &= AD\bar{X}\delta + \left(\frac{X^{*\prime}(\lambda P + Q)X^{*}}{n}\right)^{-1}\frac{1}{\sqrt{n}}X^{*\prime}(\lambda P + Q)(D\epsilon + u) \\
&\to \xi = \left(\text{plim}\frac{X^{*\prime}(\lambda P + Q)X^{*}}{n}\right)^{-1}\left(\text{plim}\frac{X^{*\prime}(\lambda P + Q)D\bar{X}}{n}\right)\delta \\
&\quad + \left(\text{plim}\frac{X^{*\prime}(\lambda P + Q)X^{*}}{n}\right)^{-1} z
\end{aligned}
$$

where

$$
z = \frac{1}{\sqrt{n}}X^{*\prime}(\lambda P + Q)(D\epsilon + u) \sim N\left(0,\ \sigma_u^2\left(\text{plim}\frac{X^{*\prime}(\lambda P + Q)X^{*}}{n}\right)\right)
$$

And

$$
\begin{aligned}
\xi &\to \left(C,\ \sigma_u^2\text{plim}\left(\frac{X^{*\prime}(\lambda P + Q)X^{*}}{n}\right)^{-1}\right) \\
\text{with } C &\equiv \left(\text{plim}\frac{X^{*\prime}(\lambda P + Q)X^{*}}{n}\right)^{-1}\left(\text{plim}\frac{X^{*\prime}(\lambda P + Q)D\bar{X}}{n}\right)\delta
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\text{plim}\sqrt{n}E\left(\hat{\beta}_{SP,RE} - \beta\right) &\to \left(\text{plim}\frac{X^{*\prime}(\lambda P + Q)X^{*}}{n}\right)^{-1}\left(\text{plim}\frac{\lambda\bar{X}^{*\prime}\bar{X}}{n}\right)\delta \\
&= \sigma_1^{-2}V_1\bar{X}^{*\prime}\bar{X}\delta \\
\text{AV}\left(\sqrt{n}\left(\hat{\beta}_{SP,RE}\right)\right) &= \sigma_u^2\text{plim}\left(\frac{X^{*\prime}(\lambda P + Q)X^{*}}{n}\right)^{-1} = \left(\text{plim}\frac{X^{*\prime}\Omega^{-1}X^{*}}{n}\right)^{-1}
\end{aligned}
$$

Finally, we get

$$
\sqrt{n}\left(\hat{\beta}_{SP,RE} - \beta\right) \to N\left(\sigma_1^{-2}V_1\bar{X}^{*\prime}\bar{X}\delta,\ \left(\text{plim}\frac{X^{*\prime}\Omega^{-1}X^{*}}{n}\right)^{-1}\right)
$$

Next, the FE estimator is given as:

$$
\hat{\beta}_{SP,FE} = \left(X^{*\prime}Q^{*}X^{*}\right)^{-1}X^{*\prime}Q^{*}\left(X^{*}\beta + D^{*}\alpha + u\right)
$$

From (3.13), we can obtain

$$\hat{\beta}_{SP,FE} - \beta = \left(X^{*\prime}Q^{*}X^{*}\right)^{-1} X^{*\prime}Q^{*}\left(D^{*}\bar{X}\frac{\delta}{\sqrt{n}} + u\right) = \left(X^{*\prime}Q^{*}X^{*}\right)^{-1} X^{*\prime}Q^{*}u$$

and therefore

$$\sqrt{n}\left(\hat{\beta}_{SP,FE} - \beta\right) \rightarrow N\left(0,\ \sigma_u^2\left(\text{plim}\frac{X^{*\prime}Q^{*}X^{*}}{n}\right)^{-1}\right)$$

Also

$$\begin{aligned}
& n\left(\hat{\beta}_{SP,RE} - \beta\right)'\left(\hat{\beta}_{SP,FE} - \beta\right) \\
= \quad & \left(\frac{X^{*\prime}\left(\lambda P + Q\right)X^{*}}{n}\right)^{-1}\frac{1}{n}X^{*\prime}\left(\lambda P + Q\right)uu'Q^{*}X^{*}\left(\frac{X^{*\prime}Q^{*}X^{*}}{n}\right)^{-1} \\
\rightarrow \quad & \sigma_u^2\left(\text{plim}\frac{\left(X^{*\prime}\Omega^{-1}X^{*}\right)^{-1}X^{*\prime}\Omega^{-1}Q^{*}X^{*}\left(X^{*\prime}Q^{*}X^{*}\right)^{-1}}{n}\right)
\end{aligned}$$

(3.15) and (3.16) follow by the continuous mapping theorem.

Proof of Theorem 8: The proof technique is based on the arguments in Theorem 2 of Hansen (2014) with the main difference that we relax that RE estimator may not be fully efficient.

Noting that $\sqrt{n}\left(\hat{\beta}_{SP,FE} - \beta\right) \xrightarrow{d} G_2'\xi \sim N\left(0,\ V_2\right)$, then

$$R\left(\hat{\beta}_{SP,FE}\right) = \mathbb{E}\left(\xi' G_2' W G_2' \xi\right) = \text{tr}\left(WV_2\right)$$

Define Ψ^{*} as a random variable without positive part trimming in (3.16)

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

Then using (3.16) and the fact that the pointwise quadric risk of Ψ is strictly smaller than that of Ψ^{*}

$$R\left(\hat{\beta}_{SP,c}\right) = \mathbb{E}\left(\Psi' W \Psi\right) < \mathbb{E}\left(\Psi^{*\prime} W \Psi^{*}\right)$$

we can calculate that

$$\mathbb{E}\left(\Psi^{*\prime} W \Psi^{*}\right)=R\left(\hat{\beta}_{SP,FE}\right)+\tau^{2}\mathbb{E}\left(\frac{(h+\xi)^{\prime}GWG^{\prime}(h+\xi)}{\left((h+\xi)^{\prime}B(h+\xi)\right)^{2}}\right)-2\tau\mathbb{E}\left(\frac{(h+\xi)^{\prime}GWG_{2}^{\prime}\xi}{(h+\xi)^{\prime}B(h+\xi)}\right)$$

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

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

$\eta(x)=x/(x^{\prime}Bx)$, and

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

Therefore

$$\begin{aligned}
&\mathbb{E}\left(\frac{(h+\xi)^{\prime}GWG_{2}^{\prime}\xi}{(h+\xi)^{\prime}B(h+\xi)}\right)\\
=\;&\mathbb{E}\mathrm{tr}\left(\frac{GWG_{2}^{\prime}V}{(h+\xi)^{\prime}B(h+\xi)}-\frac{2GWG_{2}^{\prime}V}{\left((h+\xi)^{\prime}B(h+\xi)\right)^{2}}B(h+\xi)(h+\xi)^{\prime}\right)\\
=\;&\mathbb{E}\left(\frac{\mathrm{tr}\left(GWG_{2}^{\prime}V\right)}{(h+\xi)^{\prime}B(h+\xi)}\right)-2\mathbb{E}\mathrm{tr}\left(\frac{GWG_{2}^{\prime}V}{\left((h+\xi)^{\prime}B(h+\xi)\right)^{2}}B(h+\xi)(h+\xi)^{\prime}\right)
\end{aligned}$$

Since

$$GWG_{2}^{\prime}V=WG_{2}^{\prime}VG=W\left(V_{2}-V_{21}\right)$$

and

$$\begin{aligned}
GWG_{2}^{\prime}VB &= GWG_{2}^{\prime}VG\left(V_{1}+V_{2}-(V_{21}+V_{21}^{\prime})\right)^{-1}G^{\prime}\\
&= GW\left(V_{2}-V_{21}\right)\left(V_{1}+V_{2}-(V_{21}+V_{21}^{\prime})\right)^{-1}G^{\prime}
\end{aligned}$$

set

$$C\equiv W\left(V_{2}-V_{21}\right)\left(V_{1}+V_{2}-(V_{21}+V_{21}^{\prime})\right)^{-1}$$

then

$$\mathbb{E}\mathrm{tr}\left(\frac{GWG_{2}^{\prime}V}{\left((h+\xi)^{\prime}B(h+\xi)\right)^{2}}B(h+\xi)(h+\xi)^{\prime}\right)=\mathbb{E}\mathrm{tr}\left(\frac{(h+\xi)^{\prime}GCG^{\prime}(h+\xi)}{\left((h+\xi)^{\prime}B(h+\xi)\right)^{2}}\right)$$

Thus

$$\begin{aligned}
\mathbb{E}\left(\Psi^{*\prime}W\Psi^{*}\right) &= R\left(\hat{\beta}_{FE}\right)+\tau^{2}\mathbb{E}\left(\frac{(h+\xi)'\,GWG'(h+\xi)}{\left((h+\xi)'\,B(h+\xi)\right)^{2}}\right) \\
&\quad +4\tau\mathbb{E}\mathrm{tr}\left(\frac{(h+\xi)'GCG'(h+\xi)}{\left((h+\xi)'\,B(h+\xi)\right)^{2}}\right) \\
&\quad -2\tau\mathbb{E}\mathrm{tr}\left(\frac{\left(W\left(V_{2}-V_{21}\right)\right)}{(h+\xi)'\,B(h+\xi)}\right) \qquad \text{(B.1)}
\end{aligned}$$

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 $b'ab\leq\left(b'b\right)\lambda_{\max}\left(a\right)$ for symmetric a, and let

$$\lambda_{\max}\left(a\right)=\lambda_{\max}\left(\frac{A+A'}{2}\right)=\lambda_{1}$$

Then

$$\begin{aligned}
\mathrm{tr}\left(B(h+\xi)(h+\xi)'GWG_{2}'V\right) &= \frac{(h+\xi)'B_{1}'\left(A+A'\right)B_{1}(h+\xi)}{2} \\
&\leq (h+\xi)'B(h+\xi)\lambda_{1} \qquad \text{(B.2)}
\end{aligned}$$

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 $GWG'=B_{1}'A^{*}B_{1},\ B_{1}'B_{1}=B$, and let

$$\lambda_{\max}\left(a\right)=\lambda_{\max}\left(\frac{A^{*}+A^{*\prime}}{2}\right)=\lambda_{1}^{*}$$

then we have

$$\begin{aligned}
\mathrm{tr}\left((h+\xi)'GWG'(h+\xi)\right) &= \frac{(h+\xi)'B_{1}'\left(A^{*}+A^{*\prime}\right)B_{1}(h+\xi)}{2} \\
&\leq (h+\xi)'B(h+\xi)\lambda_{1}^{*} \qquad \text{(B.3)}
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\left(\Psi^{*\prime} W \Psi^{*}\right) &\leq R\left(\hat{\beta}_{SP,FE}\right)+\tau^{2}\mathbb{E}\left(\frac{\lambda_{1}^{*}}{(h+\xi)^{\prime} B(h+\xi)}\right)+4\tau\mathbb{E}\left(\frac{\lambda_{1}}{(h+\xi)^{\prime} B(h+\xi)}\right) \\
&\quad -2\tau\mathbb{E}\mathrm{tr}\left(\frac{\left(W\left(V_{2}-V_{12}\right)\right)}{(h+\xi)^{\prime} B(h+\xi)}\right) \\
&= R\left(\hat{\beta}_{SP,FE}\right)-\mathbb{E}\left(\frac{\tau\left(2\left(\mathrm{tr}W\left(V_{2}-V_{21}\right)-2\lambda_{1}\right)-\lambda_{1}^{*}\tau\right)}{(h+\xi)^{\prime} B(h+\xi)}\right) \\
&\leq R\left(\hat{\beta}_{SP,FE}\right)-\frac{\tau\left(2\left(\mathrm{tr}W\left(V_{2}-V_{21}\right)-2\lambda_{1}\right)-\lambda_{1}^{*}\tau\right)}{\mathbb{E}\left((h+\xi)^{\prime} B(h+\xi)\right)} \qquad \text{(B.4)}
\end{aligned}$$

Since $\mathrm{tr}(BV)=\mathrm{tr}\left(G\left(V_{1}+V_{2}-\left(V_{21}+V_{21}^{\prime}\right)\right)^{-1} G^{\prime} V\right)=q$. We have

$$\begin{aligned}
\mathbb{E}\left((h+\xi)^{\prime} B(h+\xi)\right) &= h^{\prime} B h+\mathrm{tr}(BV) \\
&= \sigma_{1}^{-4} \delta^{\prime} \bar{X}^{\prime} \bar{X}^{*} V_{1}\left(V_{1}+V_{2}-\left(V_{21}+V_{21}^{\prime}\right)\right)^{-1} V_{1} \bar{X}^{*\prime} \bar{X} \delta+q
\end{aligned}$$

Substitute into (B.4), finally we obtain

$$R\left(\hat{\beta}_{SP,c}\right) \leq R\left(\hat{\beta}_{SP,FE}\right)-\frac{\tau\left(2\left(\mathrm{tr}W\left(V_{2}-V_{21}\right)-2\lambda_{1}\right)-\lambda_{1}^{*}\tau\right)}{\sigma_{1}^{-4} \delta^{\prime} \bar{X}^{\prime} \bar{X}^{*} V_{1}\left(V_{1}+V_{2}-\left(V_{21}+V_{21}^{\prime}\right)\right)^{-1} V_{1} \bar{X}^{*\prime} \bar{X} \delta+q}$$

Appendix C

Appendix for Chapter 5

Proof of equation (5.13)

Fixed effects estimator is given as

$$\begin{aligned}\hat{\beta}_{FE} &= \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i' Q_T y_i\right) \\ \hat{\beta}_{FE} - \beta &= \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i' Q_T e_i\right)\end{aligned}$$

Given that

$$e_i = F\gamma_i + \varepsilon_i$$

Then

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

From (5.11) that

$$\gamma_i = \frac{1}{\sqrt{n}}\delta_i$$

Then we have

$$
\begin{aligned}
\hat{\beta}_{FE}-\beta &= \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1}\left(\sum_{i=1}^{n}\left(X_i' Q_T F \frac{\delta_i}{\sqrt{n}} + X_i' Q_T \varepsilon_i\right)\right) \\
\sqrt{n}\left(\hat{\beta}_{FE}-\beta\right) &= \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i' Q_T F \delta_i\right) \\
&\quad + \left(\sum_{i=1}^{n} X_i' Q_T X_i\right)^{-1}\left(\sum_{i=1}^{n} \frac{1}{\sqrt{n}} X_i' Q_T \varepsilon_i\right) \\
&\to \xi = \left(\text{plim}\frac{\sum_{i=1}^{n} X_i' Q_T X_i}{n}\right)^{-1}\left(\text{plim}\sum_{i=1}^{n}\frac{X_i' Q_T F \delta_i}{n}\right) \\
&\quad + \left(\text{plim}\frac{\sum_{i=1}^{n} X_i' Q_T X_i}{n}\right)^{-1} z
\end{aligned}
$$

where

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

And

$$
\xi \to \left(\left(\text{plim}\frac{\sum_{i=1}^{n} X_i' Q_T X_i}{n}\right)^{-1}\left(\text{plim}\sum_{i=1}^{n}\frac{X_i' Q_T F \delta_i}{n}\right),\ \sigma^2\left(\text{plim}\sum_{i=1}^{n}\frac{X_i' Q_T X_i}{n}\right)\right)
$$

Set $S=\mathbb{E}\left(X_i' Q_T F \delta_i\right)$, so

$$
\text{plim}\sqrt{n}\mathbb{E}\left(\hat{\beta}_{FE}-\beta\right) \to \left(\text{plim}\frac{\sum_{i=1}^{n} X_i' Q_T X_i}{n}\right)^{-1}\left(\text{plim}\sum_{i=1}^{n}\frac{X_i' Q_T F \delta_i}{n}\right) = \sigma^{-2} V_1 S
$$

Finally, we obtain

$$
\sqrt{n}\left(\hat{\beta}_{FE}-\beta\right) \to N\left(\sigma^{-2} V_1 S,\ V_1\right)
$$

where $V_1=\sigma^2\left(\text{plim}\sum_{i=1}^{n}\frac{X_i' Q_T X_i}{n}\right)$

Proof of equation (5.18), eq (5.18) can be obtained similar to the Semiparametric case.