

Combined Estimation of Semiparametric Panel Data Models[☆]

Bai Huang^{a,*}, Tae-Hwy Lee^b, Aman Ullah^b

^a*School of Statistics and Mathematics, Central University of Finance & Economics, Beijing, China*

^b*Department of Economics, University of California, Riverside, Riverside, CA, United States*

Abstract

The combined estimation for the semiparametric panel data models is proposed. The properties of estimators for the semiparametric panel data models with random effects (RE) and fixed effects (FE) are examined. When the RE estimator suffers from endogeneity due to the individual effects correlated with the regressors, the semiparametric RE and FE estimators may be adaptively combined, with the combining weights depending on the degree of endogeneity. The asymptotic distributions of these three estimators (RE, FE, and combined estimators) for the semiparametric panel data models are derived using a local asymptotic framework. These three estimators are then compared in asymptotic risk. The semiparametric combined estimator has strictly smaller asymptotic risk than the semiparametric fixed effect estimator. The Monte Carlo study shows that the semiparametric combined estimator outperforms semiparametric FE and RE estimators except when the degrees of endogeneity and heterogeneity of the individual effects are very small. Also presented is an empirical application where the effect of public sector capital in the private economy production function is examined using the US state level panel data.

Keywords: Endogeneity, Panel Data, Semiparametric FE estimator, Semiparametric RE estimator, Semiparametric Combined Estimator, Local Asymptotics, Hausman Test

[☆]The authors are thankful to the co-editor, the guest-editor, three anonymous referees, and the seminar participants (at EcoSta2018 at City University of Hong Kong, Monash University, CUF, UCR), for many helpful comments and suggestions. We are also grateful to Liangjun Su for discussions on the subject matter of the paper. .

*Corresponding author

URL: huangbai@cufe.edu.cn (Bai Huang), tae.lee@ucr.edu (Tae-Hwy Lee), aman.ullah@ucr.edu (Aman Ullah)

1. Introduction

When the functional form of the parametric regression function is in question, semi-parametric (SP) and nonparametric (NP) kernel methods have become widely accepted statistical techniques. In panel data model, when the functional form is in question due to little prior knowledge on a particular specification, such SP and NP kernel methods may also be very useful to capture nonlinear structure in the panel data regression functions. In this paper we adopt an SP approach to modelling a general partially linear panel data model. There is a rich literature on SP estimation of panel data model. See Horowitz and Markatou (1996)[1], Ullah and Roy (1998)[2] and Li and Hisao (1998)[3], to mention only a few. Li and Ullah (1998)[4] discuss the partially linear panel data model with random effects. Baltagi and Li (2002)[5], Henderson, Carroll, and Li (2008)[6], and Su and Ullah (2006)[7] consider the estimation of the partially linear panel data models with fixed effects.

When the exogeneity of the regressors in regression functions is in question, the model may suffer from inconsistent estimation. Various consistent estimation methods (e.g., using instruments or control functions) have become widely accepted econometric techniques. In panel data model, when the exogeneity of the regressors is in question due to possible correlation of the individual effects with the regressors, such consistent estimation may be achieved by taking the individual effects as fixed rather than random. If the individual effects are correlated with the regressors in the model, the random-effects (RE) estimation is inconsistent while the fixed-effects (FE) estimation is still consistent. 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 inconsistent. On the other hand, if the individual effects are not correlated with the regressors in the model, both the random- and fixed-effects estimators are consistent while the random-effects estimator is efficient. Therefore, there is a trade-off between inefficient fixed-effects estimation and inconsistent random-effects estimation. Under this scenario, when the exogeneity assumption fails or falls in doubt, i.e., when the endogeneity exists but is weak, the inefficient fixed-effects estimation and the inconsistent random-effects estimation may be combined to take advantage of each of the two estimators and to obtain a combined estimator.

In this paper, we adopt the combined estimation approach for the semiparametric panel data models. The combined estimator (denoted SPCombined) is a weighted combination of the SP random-effects (SPRE) estimator and the SP fixed-effects (SPFE) estimator. The combining weights depend on the Hausman
 30 statistic that measures the degrees of the endogeneity. The asymptotic distributions of the SPRE estimator, the SPFE estimator, and their combined estimator, are derived using a local asymptotic framework (Theorem 1). We show that under certain conditions, the SPCombined estimator has strictly smaller asymptotic risk than the SPFE estimator (Theorem 2). The parametrically combined estimator can be viewed as a special case of the semi-parametric combined estimator (Remark 2). Further, in the Monte Carlo study we show
 35 that the SPCombined estimator outperforms the SPFE and SPRE estimators except when the degrees of “endogeneity” and “heterogeneity” of the individual effects are very small. Finally, in an application, we re-examine the role for public sector in affecting private sector economic performance, using these three estimators for the SP panel data model.

This paper is organized as follows. Section 2 presents three estimators for the semiparametric panel data
 40 model – the SPRE, SPFE, and SPCombined estimators. Section 3 presents the asymptotic distributions of these three estimators. Section 4 compares them in the asymptotic risk. The proofs are collected in Appendix (Section 8). Section 5 gives Monte Carlo simulation. An empirical application is presented in Section 6. Section 7 concludes.

2. Estimation

45 In this section, we introduce the three estimators – the SPRE, SPFE, and SPCombined estimators, for the semiparametric panel data models.

First, we consider the following semi-parametric regression model with fixed effects (SPFE):

$$y_{it} = x'_{it}\beta + m(z_{it}) + \alpha_i + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (1)$$

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

for a fixed T . In matrix notation, (1) can be written as

$$\mathbf{y} = \mathbf{X}\beta + m(\mathbf{Z}) + \mathbf{D}\alpha + u.$$

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

A local linear approximation of the model (1) can be written as

$$\begin{aligned} y_{it} &\approx x'_{it}\beta + m(z) + (z_{it} - z)\dot{m}(z) + \alpha_i + u_{it}, \\ &= x'_{it}\beta + \mathbf{Z}'_{it}(z)\delta(z) + \alpha_i + u_{it}. \end{aligned} \tag{2}$$

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

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}(z)\delta(z) + \mathbf{D}\alpha + \mathbf{u},$$

where $\mathbf{y} = (y_{11}, \dots, y_{1T}, \dots, y_{n1}, \dots, y_{nT})'$, $\mathbf{X} = (x_{11}, \dots, x_{1T}, \dots, x_{n1}, \dots, x_{nT})'$ and

$$\mathbf{Z}(z) = (Z_{11}(z), \dots, Z_{1T}(z), \dots, Z_{n1}(z), \dots, Z_{nT}(z))'.$$

Let K denote a kernel function on \mathbb{R}^p and $\mathbf{H} = \text{diag}(h_1, \dots, h_p)$, a matrix of bandwidth sequences.

(In our application $p = 1$. Nevertheless, our theory is general to take $p > 1$. See Assumption 5. Many applications with $p > 1$ can be found in, e.g., Henderson and Parmeter, 2015[8]; Rilston and Ullah, 1989[9]; Pagan and Ullah, 1999[10]; Ullah 1988[11]; Su, Ullah and Wang, 2013[12]; Su, Murtazashvili, and Ullah, 2013[13]. Our Monte Carlo simulation considers $p = 1$ and $p = 2$.)

Set $K_H(z) = |H|^{-1}K(H^{-1}z)$, where $|H|$ is the determinant of H . Further denote that $\mathbf{K}(H^{-1}z) = \text{diag}(K(H^{-1}(z_{11} - z)), \dots, K(H^{-1}(z_{1T} - z)), \dots, K(H^{-1}(z_{nT} - z)))$. Given $(\alpha', \beta)'$, Su and Ullah (2006)[7] estimate $\delta(z)$ by minimizing the following criterion function

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

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

$$\delta(z) = \mathbf{S}(z) (\mathbf{y} - \mathbf{D}\alpha - \mathbf{X}\beta).$$

In particular, $m(z)$ is given by

$$m(z) = \mathbf{s}(z)'(\mathbf{y} - \mathbf{X}\beta - \mathbf{D}\alpha), \quad (3)$$

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

$$(\mathbf{y} - \mathbf{X}\beta - \mathbf{D}\alpha - m(\mathbf{Z}))'(\mathbf{y} - \mathbf{X}\beta - \mathbf{D}\alpha - m(\mathbf{Z})), \quad (4)$$

where $m(\mathbf{Z}) = (m(z_{11}), \dots, m(z_{1T}), \dots, m(z_{nT}))'$. Plugging (3) into (4) and using the formula for partitioned regression, we can obtain

$$\hat{\beta}_{SPFE} = (\mathbf{X}^*\mathbf{Q}^*\mathbf{X}^*)^{-1} \mathbf{X}^*\mathbf{Q}^*\mathbf{y}^* \quad (5)$$

and

$$\hat{\alpha}_{SPFE} = (\mathbf{D}^*\mathbf{D}^*)^{-1} \mathbf{D}^*\mathbf{y}^*,$$

where $\mathbf{D}^* = (\mathbf{I}_{nT} - \mathbf{S})\mathbf{D}$, $\mathbf{y}^* = (\mathbf{I}_{nT} - \mathbf{S})\mathbf{y}$, $\mathbf{X}^* = (\mathbf{I}_{nT} - \mathbf{S})\mathbf{X}$, $\mathbf{Q}^* = \mathbf{I}_{nT} - \mathbf{D}^*(\mathbf{D}^*\mathbf{D}^*)^{-1}\mathbf{D}^*$, $\mathbf{S} = (s_{11}, \dots, s_{1T}, s_{21}, \dots, s_{nT})'$, and $s_{it} = s(z_{it})$. The profile likelihood estimator for $\delta(z)$ is given by

$$\hat{\delta}(z) = \mathbf{S}(z) \left(\mathbf{y} - \mathbf{X}\hat{\beta}_{SPFE} - \mathbf{D}\hat{\alpha}_{SPFE} \right).$$

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

$$\hat{m}(z) = \mathbf{s}(z)' \left(\mathbf{y} - \mathbf{X}\hat{\beta}_{SPFE} - \mathbf{D}\hat{\alpha}_{SPFE} \right).$$

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

$$\sqrt{n} \left(\hat{\beta}_{SPFE} - \beta \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{SPFE}), \quad (6)$$

where $\mathbf{V}_{SPFE} = \sigma_u^2 \left(\text{plim} \frac{\mathbf{X}^*\mathbf{Q}^*\mathbf{X}^*}{n} \right)^{-1}$.

Second, we present the semi-parametric regression model with random effects (SPRE):

$$y_{it} = x'_{it}\beta + m(z_{it}) + \alpha_i + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (7)$$

in matrix notation, (7) can be written as

$$\mathbf{y} = \mathbf{X}\beta + m(Z) + \mathbf{v}. \quad (8)$$

The error \mathbf{v} in (8) follows an one-way error components structure:

$$\mathbf{v} = \mathbf{D}\alpha + \mathbf{u},$$

where $\alpha \sim (0, \sigma_\alpha^2 \mathbf{I}_n)$, $\mathbf{u} \sim (0, \sigma_u^2 \mathbf{I}_{nT})$, $\mathbf{v} \sim (0, \mathbf{\Omega})$. The variance-covariance matrix of \mathbf{v} is given by $\mathbf{\Omega} = \sigma_u^2 \mathbf{I}_{nT} + \sigma_\alpha^2 \mathbf{D}\mathbf{D}' = \sigma_1^2 \mathbf{P} + \sigma_u^2 \mathbf{Q}$, where $\mathbf{Q} = \mathbf{I}_{nT} - \mathbf{P}$, $\mathbf{P} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1} \mathbf{D}'$, $\sigma_1^2 = T\sigma_\alpha^2 + \sigma_u^2$. The inverse matrix of $\mathbf{\Omega}$ is given by $\mathbf{\Omega}^{-1} = \frac{1}{\sigma_1^2} \mathbf{P} + \frac{1}{\sigma_u^2} \mathbf{Q}$. By taking expectation of (7) conditional on z_{it} , obtain

$$E(y_{it}|z_{it}) = E(x'_{it}|z_{it})\beta + m(z_{it}). \quad (9)$$

Then one can eliminate the unknown function $m(\cdot)$ by subtracting (9) from (7) to get

$$y_{it} - E(y_{it}|z_{it}) = (x_{it} - E(x'_{it}|z_{it}))\beta + v_{it}.$$

In vector notation,

$$\begin{aligned} \mathbf{y} - E(\mathbf{y}|\mathbf{Z}) &= (\mathbf{X} - E(\mathbf{X}|\mathbf{Z}))\beta + \mathbf{v}, \\ \mathbf{y}^* &= \mathbf{X}^*\beta + \mathbf{v}, \end{aligned} \quad (10)$$

where the conditional expectations $E(\mathbf{y}|\mathbf{Z})$ and $E(\mathbf{X}|\mathbf{Z})$ can be estimated by local linear least squares estimators (LLLS). Therefore, \mathbf{y}^* and \mathbf{X}^* in (10) are

$$\mathbf{y}^* = (\mathbf{I}_{nT} - \mathbf{S})\mathbf{y}$$

and

$$\mathbf{X}^* = (\mathbf{I}_{nT} - \mathbf{S})\mathbf{X}.$$

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

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

and

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

where $\hat{\mathbf{v}} = \mathbf{y}^* - \mathbf{X}^* \hat{\beta}_{OLS}$ is the OLS residual and $\hat{\beta}_{OLS} = (\mathbf{X}^{*\prime} \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{y}^*$. 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 (10) is

$$\hat{\beta}_{SPRE} = \left(\mathbf{X}^{*\prime} \hat{\mathbf{\Omega}}^{-1} \mathbf{X}^* \right)^{-1} \mathbf{X}^{*\prime} \hat{\mathbf{\Omega}}^{-1} \mathbf{y}^*. \quad (11)$$

Following the appendix of Li and Ullah (1998)[4], it can be easily verified that $\hat{\sigma}_u^2 = \sigma_u^2 + o_p(1)$ and $\hat{\sigma}_\alpha^2 = \sigma_\alpha^2 + o_p(1)$, and $\hat{\mathbf{\Omega}}^{-1} = \mathbf{\Omega}^{-1} + o_p(1)$. $\hat{\beta}_{SPRE}$ has an asymptotic distribution as

$$\sqrt{n} \left(\hat{\beta}_{SPRE} - \beta \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{SPRE}), \quad (12)$$

where $\mathbf{V}_{SPRE} = \left(\text{plim} \frac{\mathbf{X}^{*\prime} \mathbf{\Omega}^{-1} \mathbf{X}^*}{n} \right)^{-1}$.

Third, we now introduce the semiparametric combined estimator (SPCombined) from combining the SPRE and SPFE estimators. Under the RE specification, the SPRE estimator is the asymptotically efficient estimator while the SPFE estimator is unbiased and consistent but not efficient. If $E(\alpha_i x_{it}) \neq 0$, the SPRE estimator is biased and inconsistent while the SPFE estimator is not affected. Motivated by this observation, we would like to see if the combination of $\hat{\beta}_{SPRE}$ and $\hat{\beta}_{SPFE}$ can result in an improved estimation. As in Hansen (2017)[14] for combining 2SLS and OLS, we consider the following SPCombined estimator of β , which is a weighted combination of $\hat{\beta}_{SPRE}$ and $\hat{\beta}_{SPFE}$ with weights depending on the Hausman statistic [15]:

$$\hat{\beta}_{SP,c} = w \hat{\beta}_{SPRE} + (1-w) \hat{\beta}_{SPFE}, \quad (13)$$

where

$$w = \begin{cases} \frac{\tau}{H_n} & \text{if } H_n \geq \tau \\ 1 & \text{if } H_n < \tau \end{cases} \quad (14)$$

and

$$H_n = (\hat{\beta}_{SPFE} - \hat{\beta}_{SPRE})' \left[\hat{\mathbf{V}}(\hat{\beta}_{SPFE} - \hat{\beta}_{SPRE}) \right]^{-1} (\hat{\beta}_{SPFE} - \hat{\beta}_{SPRE}), \quad (15)$$

where τ is a shrinkage parameter and H_n is the Hausman statistic. The notation H_n here should not

60 be confused with the notation $\mathbf{H} = \text{diag}(h_1, \dots, h_p)$ introduced earlier for a matrix of bandwidths. Also, the notation h_1, \dots, h_p for the bandwidths should not be confused with the non-centrality parameter h in Equation (19) for Theorem 1.

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

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

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

3. Asymptotic Distributions

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

$$\alpha_i = \bar{\mathbf{x}}_i' \rho + \epsilon_i, \tag{16}$$

where $E(\bar{\mathbf{x}}_i \epsilon_i) = 0$. The variable x_{it} is exogenous if α_i and x_{it} are uncorrelated, or equivalently if the coefficient ρ is zero. We use the local asymptotic approach. For fixed T , ρ is local to zero,

$$\rho = \frac{1}{\sqrt{n}} \delta. \tag{17}$$

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

Now, we make the following assumptions:

Assumption 1. $(\alpha_i, \mathbf{u}_i, \mathbf{x}_i, \mathbf{z}_i)$, $i = 1, \dots, n$, are i.i.d. over i , u_{it} is i.i.d over t , where $\mathbf{u}_i = (u_{i1}, \dots, u_{iT})'$ and \mathbf{x}_i and \mathbf{z}_i are similarly defined. $E(u_{it}|x_{it}, \alpha_i) = 0$ and $E(u_{it}^4|x_{it}, \alpha_i) < \infty$.

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

Assumption 3. Let $x_{it}^* \equiv x_{it} - E(x_{it}|z_{it})$, $\sum_t E\{x_{it}^* [x_{it}^* - \sum_s x_{is}^*/T]'\}$ is positive definite.

Assumption 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 5. As $n \rightarrow \infty$, $\|\mathbf{H}\| \rightarrow 0$, $n|\mathbf{H}|^2 \rightarrow \infty$, $\|\mathbf{H}\|^4 |\mathbf{H}|^{-1} \rightarrow 0$ and $n|\mathbf{H}| \|\mathbf{H}\|^4 \rightarrow c \in [0, \infty)$, where

$|\mathbf{H}|$ is the determinant of \mathbf{H} .

Assumption 6. $E \|x_{it}\|^4 < \infty$; $\sigma_u^2 \left(\text{plim} \frac{\mathbf{X}^{*'} \mathbf{Q}^* \mathbf{X}^*}{n} \right)^{-1} = \mathbf{V}_2$, $\text{plim} \left(\frac{\mathbf{X}^{*'} \boldsymbol{\Omega}^{-1} \mathbf{X}^*}{n} \right)^{-1} = \mathbf{V}_1$ and $\sigma_u^2 \left(\text{plim} \frac{(\mathbf{X}^{*'} \mathbf{Q}^* \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{Q}^* \boldsymbol{\Omega}^{-1} \mathbf{X}^* (\mathbf{X}^{*'} \boldsymbol{\Omega}^{-1} \mathbf{X}^*)^{-1}}{n} \right) = \mathbf{V}_{21}$ as $n \rightarrow \infty$.

Assumptions 1 and 2 are standard in the literature. Assumption 3 rules out time-invariant terms x_{it} .

Assumption 4 requires that K be compactly supported. From Su and Ullah (2006)[7], Assumption 5 is easily

satisfied by considering $\mathbf{H} = \text{diag}(h_1, \dots, h_p)$ 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 practical application of the partially linear semiparametric panel data models. In

Assumption 6, we simplify the notation by $\mathbf{V}_{SPRE} =: \mathbf{V}_1$ and $\mathbf{V}_{SPFE} =: \mathbf{V}_2$. It is important to note that

the SPRE estimator may not be fully efficient for the SP panel models and thus $\mathbf{V}_1 \neq \mathbf{V}_{21}$, which makes

Theorem 1 and Theorem 2 of this paper different from those of Hansen (2017)[14].

Theorem 1. *Under Assumptions 1–6,*

$$\sqrt{n} \begin{pmatrix} \hat{\beta}_{SPRE} - \beta \\ \hat{\beta}_{SPFE} - \beta \end{pmatrix} \xrightarrow{d} \mathbf{h} + \boldsymbol{\xi}, \quad (18)$$

where

$$\mathbf{h} = \begin{pmatrix} \sigma_1^{-2} \mathbf{V}_1 \boldsymbol{\Sigma} \delta \\ 0 \end{pmatrix}, \quad \text{with } \boldsymbol{\Sigma} = \text{plim} \frac{\bar{\mathbf{X}}^{*'} \bar{\mathbf{X}}}{n}, \quad (19)$$

and

$$\boldsymbol{\xi} \sim N(0, \mathbf{V}), \quad \text{with } \mathbf{V} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}'_{21} \\ \mathbf{V}_{21} & \mathbf{V}_2 \end{pmatrix}.$$

Furthermore,

$$H_n \xrightarrow{d} (\mathbf{h} + \boldsymbol{\xi})' \mathbf{B} (\mathbf{h} + \boldsymbol{\xi}), \quad (20)$$

and

$$\sqrt{n} \left(\hat{\beta}_{SP,c} - \beta \right) \xrightarrow{d} \boldsymbol{\Psi} = \mathbf{G}'_2 \boldsymbol{\xi} - \left(\frac{\tau}{(\mathbf{h} + \boldsymbol{\xi})' \mathbf{B} (\mathbf{h} + \boldsymbol{\xi})} \right)_1 \mathbf{G}' (\mathbf{h} + \boldsymbol{\xi}), \quad (21)$$

where $\bar{\mathbf{X}}$ is $n \times q$ with $\bar{\mathbf{x}}_i = \sum_t x_{it}/T$ in its i th row, $\bar{\mathbf{X}}^*$ is $n \times q$ with $\bar{\mathbf{x}}^*_i = \sum_t x^*_{it}/T$ in its i th row,

$\mathbf{B} = \mathbf{G} (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-1} \mathbf{G}'$, $\mathbf{G} = \begin{pmatrix} -\mathbf{I} & \mathbf{I} \end{pmatrix}'$, $\mathbf{G}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{I} \end{pmatrix}'$, and $(a)_1 = \min[1, a]$. \square

Theorem 1 presents the joint asymptotic distribution of $\hat{\beta}_{SPRE}$ and $\hat{\beta}_{SPFE}$, the Hausman statistic, and $\hat{\beta}_{SP,c}$ under the local exogeneity assumption. The joint asymptotic distribution of $\hat{\beta}_{SPRE}$ and $\hat{\beta}_{SPFE}$ is normal. $\hat{\beta}_{SPRE}$ has an asymptotic bias when $\delta \neq 0$ but $\hat{\beta}_{SPFE}$ does not have an asymptotic bias. The Hausman statistic has an asymptotic non-central chi-square distribution, with non-centrality parameter \mathbf{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 \mathbf{h} .

Remark 1. Theorem 1 extends Hansen (2017) for the panel data models and generalizes his results by allowing $\mathbf{V}_1 \neq \mathbf{V}_{21}$ and $\mathbf{B} = \mathbf{G}(\mathbf{V}_1 + \mathbf{V}_2 - \mathbf{V}_{21} - \mathbf{V}'_{21})^{-1} \mathbf{G}'$ to be asymmetric. If $\hat{\beta}_{SPRE}$ were fully efficient, then $\mathbf{V}_1 = \mathbf{V}_{21}$ and $\mathbf{B} = \mathbf{G}(\mathbf{V}_2 - \mathbf{V}_1)^{-1} \mathbf{G}'$ is symmetric as in the case of Hansen (2017). In general $\hat{\beta}_{SPRE}$ is not fully efficient and so $\mathbf{V}_1 \neq \mathbf{V}_{21}$. Therefore, the computation of \mathbf{V}_{21} and $\mathbf{B} = \mathbf{G}(\mathbf{V}_1 + \mathbf{V}_2 - \mathbf{V}_{21} - \mathbf{V}'_{21})^{-1} \mathbf{G}'$ is needed. This also affects the calculation of the asymptotic risk as shown in the next section.

4. Asymptotic Risk

Based on Theorem 1, we now compare $\hat{\beta}_{SPRE}$, $\hat{\beta}_{SPFE}$, $\hat{\beta}_c$ in the asymptotic risk. The asymptotic risk of any sequence of estimators β_n of β is defined as

$$R(\beta_n, \beta, \mathbf{W}) = \lim_{n \rightarrow \infty} E [n(\beta_n - \beta)' \mathbf{W}(\beta_n - \beta)], \quad (22)$$

so long as the estimator has an asymptotic distribution

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

for some random variable ψ . Denote $R(\beta_n, \beta, \mathbf{W}) =: R(\beta_n)$ for short. The asymptotic risk can be calculated using

$$R(\beta_n) = E(\psi' \mathbf{W} \psi) = \text{tr}(\mathbf{W} E(\psi \psi')). \quad (23)$$

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

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

and

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

where

$$\mathbf{A} = (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{\frac{1}{2}} \mathbf{W} (\mathbf{V}_2 - \mathbf{V}_{21}) (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-\frac{1}{2}}$$

and

$$\mathbf{A}^* = (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{\frac{1}{2}} \mathbf{W} (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{\frac{1}{2}}.$$

Let

$$d = \frac{\text{tr}(\mathbf{W}(\mathbf{V}_2 - \mathbf{V}_{21}))}{\lambda_1}.$$

Theorem 2. *Under Assumptions 1–6, if*

$$d > 2 \tag{24}$$

and

$$0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}, \tag{25}$$

then

$$R(\hat{\beta}_{SP,c}) = \text{tr}[\mathbf{W}E(\boldsymbol{\Psi}\boldsymbol{\Psi}')],$$

$$R(\hat{\beta}_{SPFE}) = \text{tr}(\mathbf{W}\mathbf{V}_2)$$

and

$$R(\hat{\beta}_{SP,c}) < R(\hat{\beta}_{SPFE}) - \frac{\tau(2\lambda_1(d-2) - \lambda_1^*\tau)}{\sigma_1^{-4}\delta'\boldsymbol{\Sigma}'\mathbf{V}_1(\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-1}\mathbf{V}_1\boldsymbol{\Sigma}\delta + q}. \tag{26}$$

100 \square

Proof: See Appendix.

Equation (26) shows that the asymptotic risk of $\hat{\beta}_{SP,c}$ is strictly less than that of $\hat{\beta}_{SPFE}$, so long as τ satisfies the condition (25). The assumption $d > 2$ is the critical condition needed to ensure that $\hat{\beta}_{SP,c}$ can have smaller asymptotic risk than that of $\hat{\beta}_{SPFE}$. It is necessary in order for the right-hand-side of (25) to be positive, which is necessary for the existence of τ satisfying (25). τ appears in the risk bound (26) as a

quadratic expression, so there is an optimal choice

$$\tau^* = \frac{\lambda_1}{\lambda_1^*} (d-2), \quad (27)$$

which minimizes this bound.

Corollary 1. $R(\hat{\beta}_{SP,c}) - R(\hat{\beta}_{SPFE}) < 0$, for $d > 2$ and $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$. When $\mathbf{W} = (\mathbf{V}_2 - \mathbf{V}_{21})^{-1}$, $0 < \tau \leq 2\left(\frac{q-2}{\lambda_1^*}\right)$ and $q > 2$, which is Stein's (1956)[16] classic condition for shrinkage. \square

105 The following two corollaries are obtained with $\mathbf{W} = (\mathbf{V}_2 - \mathbf{V}_1)^{-1}$.

Corollary 2. $R(\hat{\beta}_{SPRE}) = \text{tr}(\mathbf{W}\mathbf{V}_1) + \sigma_1^{-4}\delta'\Sigma'\mathbf{V}_1\mathbf{W}\mathbf{V}_1\Sigma\delta$; $R(\hat{\beta}_{SPRE}) \leq R(\hat{\beta}_{SPFE})$ when $\sigma_1^{-4}\delta'\Sigma'\mathbf{V}_1\mathbf{W}\mathbf{V}_1\Sigma\delta \leq q$, and $R(\hat{\beta}_{SPRE}) > R(\hat{\beta}_{SPFE})$ otherwise. \square

Corollary 3. $R(\hat{\beta}_{SP,c}) - R(\hat{\beta}_{SPRE}) < 0$, for $q < \sigma_1^{-4}\delta'\Sigma'\mathbf{V}_1\mathbf{W}\mathbf{V}_1\Sigma\delta$, $d > 2$, and $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$. \square

Corollary 2 indicates that when endogeneity is weak (ρ and hence δ is close to zero), $\hat{\beta}_{SPRE}$ may perform
 110 better than $\hat{\beta}_{SPFE}$. Corollary 3 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 2. A parametric combined estimator can be viewed as a special case of the semi-parametric combined estimator. Write (2) as

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

where $\alpha(z) = m(z) - z\dot{m}(z)$, $\mathbf{Z}_{it} = (1, z_{it})$, $\delta(z) = (\alpha(z), (\dot{m}(z))')'$. As $h_l \rightarrow \infty$, $1 \leq l \leq p$, the weighted function $\mathbf{K}_H(z) \rightarrow \mathbf{K}(0)$ and local minimization becomes global minimization:

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

In this case, dropping \mathbf{Z} or assuming it is in \mathbf{X} , one can obtain the combined estimator for the parametric linear panel data model as

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

where $\hat{\beta}_{SPFE} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}\mathbf{y}$ and $\hat{\beta}_{SPRE} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{y}$. Noting that if \mathbf{Z} is included in
115 regression, the only difference is that \mathbf{X} will be replaced by $\mathbf{M}\mathbf{X}$ and \mathbf{y} will be replaced by $\mathbf{M}\mathbf{y}$ where \mathbf{M} is
the same as \mathbf{Q} but based on \mathbf{Z} .

5. Monte Carlo

The observations are generated by the progress

$$y_{it} = x'_{it}\beta + m(z_{it}) + \alpha_i + u_{it}, \quad (28)$$

$$\alpha_i = \rho\sqrt{T}\bar{\mathbf{x}}'_i \frac{\iota}{\sqrt{q}} + \sqrt{1-\rho^2}\epsilon_i, \quad (29)$$

where x_{it} are i.i.d. $N(0, \mathbf{I}_q)$, u_{it} are i.i.d. $N(0, \sigma_u^2)$, and z_{it} are i.i.d. $N(0, \mathbf{I}_p)$, all across i, t . ϵ_i are i.i.d. $N(0, 1)$ independent of $\{x_{it}, u_{it}\}$. Then α_i and x_{it} have correlation $\frac{\rho}{\sqrt{q}}$, but all other correlations are zero. We allow ρ to vary in $(-1, 1)$. As the distribution is invariant to β , we set it to zero, $\beta = 0$, without loss of generality. Let ι be an $n \times 1$ vector of ones. We consider $n \in \{20, 100\}$, $T = 3$, $q \in \{4, 8\}$, $p \in \{1, 2\}$ and

$$m(z) = 2z + e^{-4(z-0.5)^2} - 1, \text{ for } p = 1 \quad (30)$$

and

$$m(z) = 2(z_1 + z_2) + e^{-4(z_1+z_2-0.5)^2} - 1, \text{ for } p = 2. \quad (31)$$

We use the Gaussian kernel. For bandwidths, we use the least-squares cross validation to choose $\mathbf{H} =$
120 $\text{diag}(h_1, \dots, h_p)$. We consider different values for $\sigma_u \in \{.4, .6, .8, 1, 1.2, 1.4, 1.6\}$, with fixing $\text{Var}(\alpha_i) =$
 $\sigma_\alpha^2 = 1$. Let $\sqrt{\theta} \equiv \frac{\sigma_\alpha}{\sigma_u}$, which measures the heterogeneity σ_α of the individual effects (relative to σ_u). Then
 $\sqrt{\theta} \in \{\frac{5}{2}, \frac{5}{3}, \frac{5}{4}, 1, \frac{5}{6}, \frac{5}{7}, \frac{5}{8}\}$. We then normalize θ on the unit interval $(0, 1)$ by defining $\rho^* \equiv \frac{\theta}{1+\theta}$. Then
 $\rho^* \in \{.28, .34, .41, .50, .61, .74, .86\}$. We report the simulation results in figures for these seven values of
125 ρ^* to examine the effect of the different degrees of heterogeneity, which also measures the temporal correlation
between $(\alpha_i + u_{it})$ and $(\alpha_i + u_{it'})$.

Having generated 100,000 samples on each of the estimators $\hat{\beta}_{SP,RE}$, $\hat{\beta}_{SP,FE}$, $\hat{\beta}_{SP,c}$, the median squared error (MedSE) of each estimator is calculated. To compare the estimators we plot the relative MedSE of

each estimator relative to MedSE of the SPFE estimator, that is

$$\frac{\text{median} \left[\left(\hat{\beta} - \beta \right)' \left(\hat{\beta} - \beta \right) \right]}{\text{median} \left[\left(\hat{\beta}_{SPFE} - \beta \right)' \left(\hat{\beta}_{SPFE} - \beta \right) \right]}. \quad (32)$$

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

Figures for $q = 4$ were similar to the plots in Figures with $q = 8$ and thus not reported. The improvement
 130 in the combined estimator over the SPFE estimator with different values of ρ^* are greater in the case of larger number (q) of regressors.

In Figure 1, (a), (c) and (e) plot the relative MedSE for $n = 20$; (b), (d) and (f) plot the relative MedSE for $n = 100$. We see that the region of dominance for the combined estimator over the SPFE and SPRE estimators is greater for small n . Figure 1 plots the relative MedSE for $\rho^* = \{.86, .74, .61\}$. These are the
 135 cases of large degree of heterogeneity. The combined estimator has uniformly smaller MedSE than SPFE. SPRE has similar MedSE to the combined estimator for small ρ , but the MedSE of the SPRE estimator increases dramatically after intermediate values of ρ .

Figure 2 plots the relative MedSE of $\hat{\beta}_{SPRE}$, $\hat{\beta}_{SPFE}$, $\hat{\beta}_{SP,c}$ with $\rho^* = \{.50, .41, .34\}$. These are the cases of moderate to small degree of heterogeneity. In Figure 2: (a), (c) and (e) plot the relative MedSE for
 140 $n = 20$; (b), (d) and (f) plot the relative MedSE for $n = 100$. We see again that the region of dominance for the combined estimator over SPFE is greater for small n . Figure 2(a) and Figure 2(b) are the cases with $\rho^* = .50$. We see that the gains from the combined estimator are strong for small ρ , with the MedSE converging to that of SPFE as ρ increases towards 1. This is consistent with Theorem 2, which shows that the improvement are asymptotically local to $\rho = 0$. The SPRE estimator has lower MedSE than the combined
 145 estimator, but the ranking is reversed for larger values of ρ .

Figure 3 plots the relative MedSE of $\hat{\beta}_{SPRE}$, $\hat{\beta}_{SPFE}$, $\hat{\beta}_{SP,c}$ with $\rho^* = .28$ for the small degree of heterogeneity. The combined estimator has much lower MedSE than SPFE, regardless of the degree of endogeneity. The combined estimator uniformly dominates SPFE. For small ρ , SPRE has lower MedSE than

the combined estimator, but the ranking is reversed for larger values of ρ .

150 Figures 4, 5, and 6 plot the relative MedSE for $p = 2$, which are similar to the plots in Figures 1, 2, and 3 for $p = 1$. The general nature is the same for both $p = 1$ and $p = 2$. With $p = 2$, we see again that the combined estimator has uniformly smaller MedSE than SPFE and we also see that the gains are stronger for smaller values of ρ and for smaller values of ρ^* . This is general finding in the simulation experiment for $p = 1$ and also for $p = 2$, which is consistent with Theorem 2.

155 Note that the improvement in the combined estimator over the SPFE estimator is greater for smaller heterogeneity ρ^* . For very small ρ^* , the SPRE estimator tends to be better than both the SPFE and the combined estimators. For moderate ρ^* and higher ρ , or moderate ρ and higher ρ^* , the combined estimator is better than SPRE estimator. For very large ρ^* and very low ρ , the combined estimator is close to the SPRE estimator.

160 Generally, the dominance of the combined estimator over the SPFE estimator is greater for small sample size n . For very small ρ^* , the SPRE estimator performs better than both the SPFE and combined estimators except when n and ρ are large. For moderate ρ^* and larger ρ , the combined estimator performs better than SPRE estimator, but for very small ρ the combined estimator can be beaten by the SPRE estimator. For very large ρ^* and very small ρ , the combined estimator is close to both SPRE and SPFE estimators, while
 165 both the combined and SPFE estimators have smaller risk than the SPRE estimator for larger values of ρ .

In summary, the simulation results provide strong finite sample confirmation of Theorem 2 and its Corollaries 1, 2, 3.

Remark 3. The risk function (22) with a general weight matrix \mathbf{W} includes many special cases. For example, the unweighted MSE is obtained by setting $\mathbf{W} = \mathbf{I}_q$, in which case the coefficients are of equal
 170 importance. The canonical case is motivated by ease of use and simplicity, which is obtained by setting $\mathbf{W} = (\mathbf{V}_{SPFE} - \mathbf{V}_{SPRE})^{-1}$. This choice simplifies many formulae, e.g., equation (24) has the simplification $d = q$, and the optimal choice of τ , τ^* in (27) is $q - 2$. Following Hansen (2017), we set $\tau = q - 2$. However, when when $\mathbf{W} = \mathbf{I}_q$ is used instead of $\mathbf{W} = (\mathbf{V}_{SPFE} - \mathbf{V}_{SPRE})^{-1}$, it is possible that the condition $0 < \tau \leq \frac{2\lambda_1(d-2)}{\lambda_1^*}$ in equation (25) may not hold if $\tau = q - 2$ is used, especially when the dimension q of \mathbf{X}

175 is large. In that case we should use the theoretical optimal $\tau = d - 2 = \frac{\text{tr}(\mathbf{W}(\mathbf{V}_2 - \mathbf{V}_{21}))}{\lambda_1} - 2$.

Remark 4. The optimal choice of τ , τ^* in (27), is obtained from minimizing the “bound” of the risk of $R(\hat{\beta}_{SP,c})$ in Theorem 2. The bound is the RHS term in equation (26). However, the optimal choice of $\tau = \tau^*$ is too small when ρ is small because the probability that the Hausman statistic is smaller than $\tau = \tau^*$ will be too small. Hence one can increase it to $2\tau^*$ as this choice still satisfies the classic James-Stein
 180 conditions in equations (24) and (25). This makes the MSE of the SP-Combined estimator closer to the MSE of SPFE when the degree of endogeneity is small. Hence, this indicates that the optimal choice of $\tau = \tau^*$ in (27) obtained from minimizing the “bound” of the risk of $R(\hat{\beta}_{SP,c})$ is not “optimal” in the sense of minimizing the risk of $R(\hat{\beta}_{SP,c})$ itself. To our knowledge, there is no result on this yet and thus we leave this for a future work.

185 6. Application

In this section, we use 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)[17]. A large body of research has explored the public-sector capital and private economic performance
 190 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)[18]]. 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) [19] and Baltagi and Pinnoi
 195 (1995) [20].

The following SP panel data model is estimated:

$$\log(Y_{it}) = \beta_0 + \beta_1 \log(KG_{i,t-2}) + \beta_2 \log(KPR_{i,t-2}) + \beta_3 \log(L_{i,t-2}) + m(UNEMP_{i,t-2}) + \alpha_i + u_{it},$$

where $i = 1, \dots, 48, t = 1, \dots, 17$. Y_{it} denotes the gross private non-agricultural 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 non-agricultural payrolls, and $UNEMP$ 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 SPFE, SPRE and SPCombined estimators are 0.0148, 0.0133 and 0.0122, respectively. The SP combined estimator gives smaller MSE than both SP fixed effects and SP random effects estimators. In order to obtain these results, we use the Gaussian kernel and choose the bandwidth of $m(UNEMP)$ according to leave-one-out cross-validation.

Table 1 suggests that the elasticities of gross private non-agricultural product for SPFE estimator with respect to KG , KPR , and L are -0.0207 , 0.3501 and 0.5237 , respectively; The elasticities for SPRE 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 the SPFE and combined estimators report that the public capital is counter productive and insignificantly in the state private production. In contrast, the SPRE estimator finds that the public capital is productive and insignificantly.

Note that the Hausman statistic is 48.9985. Thus, the null hypothesis of exogeneity is easily rejected at the one percent level of significance. This indicates that there exists a huge problem of endogeneity. In this circumstance, the SPFE estimator solves the problem. Thus, it would be more appropriate to treat α_i as fixed. The SPRE estimator seems overwhelming that public capital has a positive impact on private sector output. The combined estimator result is consistent with SPFE estimator. Our empirical analysis

225 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 SPFE estimator. As a result, the combined estimator is more reliable under this scenario. This is also supported by the smaller MSE of the SPCombined estimator than both SPFE and SPRE estimators.

230 Therefore, we conclude that the public sector output has an insignificant effect on the private sector productivity. The public infrastructure has insignificant effects on private sector across states.

Table 1. Economics of Private Sector Output: Semi-parametric Models

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

48 U.S. States. 1970-1986. Reported are parameter estimates with the standard errors in parentheses.

7. Conclusions

This paper considers the combined estimation for the semiparametric panel data model with weak endogeneity (i.e., local to exogeneity). We introduce the combined estimation of the SPRE and SPFE estimators for the semi-parametric model, when the semiparametric RE estimator suffers from inconsistency due to the individual effects being random and correlated with the regressors. The combined estimator uniformly dominates SPFE, while the combined estimator also dominates SPRE except when endogeneity and heterogeneity of the individual effects are weak. The use of the combined estimation allows applied researchers to implement efficient estimation under the presence of possible endogeneity without having to select the consistent SPFE estimator or the efficient SPRE estimator.

8. Appendix

Proof of Theorem 1: Let $\mathbf{h}_I = \mathbf{G}_1 \mathbf{h}$ and $\xi_I = \mathbf{G}_1 \xi$ with $\mathbf{G}_1 = (\mathbf{I} \ \mathbf{0})'$, and let $\mathbf{h}_{II} = \mathbf{G}_2 \mathbf{h}$ and $\xi_{II} = \mathbf{G}_2 \xi$ with $\mathbf{G}_2 = (\mathbf{0} \ \mathbf{I})'$.

First, the semiparametric random effect (SPRE) estimator is

$$\begin{aligned}\hat{\beta}_{SPRE} &= (\mathbf{X}^{*'} \boldsymbol{\Omega}^{-1} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \boldsymbol{\Omega}^{-1} \mathbf{y}^*, \\ &= (\mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{X}^*)^{-1} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{y}^* = \mathbf{A} \mathbf{y}^*\end{aligned}$$

where $\boldsymbol{\Omega}^{-1} = \frac{1}{\sigma_u^2} (\lambda \mathbf{P} + \mathbf{Q}) = \frac{\mathbf{P}}{\sigma_1^2} + \frac{\mathbf{Q}}{\sigma_u^2}$, $\lambda = \frac{\sigma_u^2}{\sigma_1^2}$

$$\hat{\beta}_{SPRE} = \mathbf{A} (\mathbf{X}^* \beta + \mathbf{D} \alpha + \mathbf{u}) = \beta + \mathbf{A} (\mathbf{D} \alpha + \mathbf{u}),$$

where α is correlated with $\bar{\mathbf{X}}$ with the correlation $\rho = \frac{\delta}{\sqrt{n}}$

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

Then

$$\begin{aligned}\hat{\beta}_{SPRE} - \beta &= \mathbf{A} \left(\mathbf{D} \bar{\mathbf{X}} \frac{\delta}{\sqrt{n}} + \mathbf{D} \epsilon + \mathbf{u} \right), \\ \sqrt{n} (\hat{\beta}_{SPRE} - \beta) &= \mathbf{A} \mathbf{D} \bar{\mathbf{X}} \delta + \left(\frac{1}{n} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{X}^* \right)^{-1} \frac{1}{\sqrt{n}} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) (\mathbf{D} \epsilon + \mathbf{u}) \xrightarrow{d} \mathbf{h}_I + \xi_I,\end{aligned}$$

where

$$\begin{aligned}\mathbf{h}_I &= \left(\text{plim} \frac{1}{n} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{X}^* \right)^{-1} \left(\text{plim} \frac{1}{n} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{D} \bar{\mathbf{X}} \right) \delta, \\ &= \left(\text{plim} \frac{1}{n} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{X}^* \right)^{-1} \left(\text{plim} \lambda \frac{1}{n} \bar{\mathbf{X}}^{*'} \bar{\mathbf{X}} \right) \delta, \\ &= \sigma_1^{-2} \mathbf{V}_1 \boldsymbol{\Sigma} \delta,\end{aligned}$$

with $\boldsymbol{\Sigma} = \text{plim} \frac{1}{n} \bar{\mathbf{X}}^{*'} \bar{\mathbf{X}}$, and

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

with

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

Hence,

$$\xi_I \sim N \left(\mathbf{0}, \sigma_u^2 \left(\text{plim} \frac{1}{n} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{X}^* \right)^{-1} \right) = N(\mathbf{0}, \mathbf{V}_1),$$

and

$$\sqrt{n} \left(\hat{\beta}_{SPRE} - \beta \right) \rightarrow N(\mathbf{h}_I, \mathbf{V}_1).$$

Next, the semiparametric fixed effect (SPFE) estimator is

$$\begin{aligned} \hat{\beta}_{SPFE} &= (\mathbf{X}^{*'} \mathbf{Q}^* \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{Q}^* (\mathbf{X}^* \beta + \mathbf{D}^* \alpha + \mathbf{u}), \\ \hat{\beta}_{SPFE} - \beta &= (\mathbf{X}^{*'} \mathbf{Q}^* \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{Q}^* \left(\mathbf{D}^* \bar{\mathbf{X}} \frac{\delta}{\sqrt{n}} + \mathbf{u} \right) = \left(\frac{1}{n} \mathbf{X}^{*'} \mathbf{Q}^* \mathbf{X}^* \right)^{-1} \left(\frac{1}{n} \mathbf{X}^{*'} \mathbf{Q}^* \mathbf{u} \right), \end{aligned}$$

and

$$\sqrt{n} \left(\hat{\beta}_{SPFE} - \beta \right) \xrightarrow{d} \xi_{II} \sim N \left(\mathbf{0}, \sigma_u^2 \left(\text{plim} \frac{1}{n} \mathbf{X}^{*'} \mathbf{Q}^* \mathbf{X}^* \right)^{-1} \right).$$

Finally, to obtain the joint asymptotic distribution of the RE and FE estimators, the asymptotic covari-
250 ance matrix of the RE and FE estimators is

$$\begin{aligned} n \left(\hat{\beta}_{SPRE} - \beta \right)' \left(\hat{\beta}_{SPFE} - \beta \right) &= \left(\frac{1}{n} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{X}^* \right)^{-1} \frac{1}{n} \mathbf{X}^{*'} (\lambda \mathbf{P} + \mathbf{Q}) \mathbf{u} \mathbf{u}' \mathbf{Q}^* \mathbf{X}^* \left(\frac{1}{n} \mathbf{X}^{*'} \mathbf{Q}^* \mathbf{X}^* \right)^{-1} \\ &\rightarrow \sigma_u^2 \left(\text{plim} \frac{1}{n} (\mathbf{X}^{*'} \boldsymbol{\Omega}^{-1} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \boldsymbol{\Omega}^{-1} \mathbf{Q}^* \mathbf{X}^* (\mathbf{X}^{*'} \mathbf{Q}^* \mathbf{X}^*)^{-1} \right) \end{aligned}$$

The rest follows by the continuous mapping theorem, as in theorem 1 of Hansen (2017).

Proof of Theorem 2: The proof is based on the arguments in Theorem 2 of Hansen (2017). The important difference in this paper is that the SPRE estimator may not be fully efficient and thus the proof is not the same as that of Hansen (2017) as shown below.

Noting that $\sqrt{n} \left(\hat{\beta}_{SPFE} - \beta \right) \rightarrow_d \mathbf{G}'_2 \xi \sim N(\mathbf{0}, \mathbf{V}_2)$, then

$$R \left(\hat{\beta}_{SPFE} \right) = E \left(\xi' \mathbf{G}'_2 \mathbf{W} \mathbf{G}'_2 \xi \right) = \text{tr}(\mathbf{W} \mathbf{V}_2).$$

Define $\boldsymbol{\Psi}^*$ as a random variable without positive part trimming

$$\boldsymbol{\Psi}^* = \mathbf{G}'_2 \xi - \left(\frac{\tau}{(\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi)} \right) \mathbf{G}'(\mathbf{h} + \xi).$$

Then using the fact that the pointwise quadric risk of $\boldsymbol{\Psi}$ is strictly smaller than that of $\boldsymbol{\Psi}^*$

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

we can calculate that

$$E(\Psi^{*'}\mathbf{W}\Psi^*) = R(\hat{\beta}_{SPFE}) + \tau^2 E\left(\frac{(\mathbf{h} + \xi)' \mathbf{G}\mathbf{W}\mathbf{G}'(\mathbf{h} + \xi)}{((\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi))^2}\right) - 2\tau E\left(\frac{(\mathbf{h} + \xi)' \mathbf{G}\mathbf{W}\mathbf{G}'_2 \xi}{(\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi)}\right).$$

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

$$E(\eta(\mathbf{Z} + \mathbf{h})' \mathbf{K}\mathbf{Z}) = E\text{tr}\left(\frac{\partial}{\partial x} \eta(\mathbf{Z} + \mathbf{h})' \mathbf{K}\mathbf{V}\right),$$

$\eta(\mathbf{x}) = \mathbf{x} / (\mathbf{x}' \mathbf{B}\mathbf{x})$, and

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

255 Therefore

$$\begin{aligned} E\left(\frac{(\mathbf{h} + \xi)' \mathbf{G}\mathbf{W}\mathbf{G}'_2 \xi}{(\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi)}\right) &= E\text{tr}\left(\frac{\mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V}}{(\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi)} - \frac{2\mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V}}{((\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi))^2} \mathbf{B}(\mathbf{h} + \xi) (\mathbf{h} + \xi)'\right) \\ &= E\left(\frac{\text{tr}(\mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V})}{(\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi)}\right) - 2E\text{tr}\left(\frac{\mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V}}{((\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi))^2} \mathbf{B}(\mathbf{h} + \xi) (\mathbf{h} + \xi)'\right). \end{aligned}$$

Since

$$\mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V} = \mathbf{W}\mathbf{G}'_2 \mathbf{V}\mathbf{G} = \mathbf{W}(\mathbf{V}_2 - \mathbf{V}_{21}),$$

and

$$\mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V}\mathbf{B} = \mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V}\mathbf{G}(\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-1} \mathbf{G}' = \mathbf{G}\mathbf{W}(\mathbf{V}_2 - \mathbf{V}_{21})(\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-1} \mathbf{G}',$$

set $\mathbf{W}(\mathbf{V}_2 - \mathbf{V}_{21})(\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-1} = \mathbf{C}$, then

$$E\text{tr}\left(\frac{\mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V}}{((\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi))^2} \mathbf{B}(\mathbf{h} + \xi) (\mathbf{h} + \xi)'\right) = E\text{tr}\left(\frac{(\mathbf{h} + \xi)' \mathbf{G}\mathbf{C}\mathbf{G}'(\mathbf{h} + \xi)}{((\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi))^2}\right).$$

Thus

$$\begin{aligned} E(\Psi^{*'}\mathbf{W}\Psi^*) &= R(\hat{\beta}_{FE}) + \tau^2 E\left(\frac{(\mathbf{h} + \xi)' \mathbf{G}\mathbf{W}\mathbf{G}'(\mathbf{h} + \xi)}{((\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi))^2}\right) + 4\tau E\text{tr}\left(\frac{(\mathbf{h} + \xi)' \mathbf{G}\mathbf{C}\mathbf{G}'(\mathbf{h} + \xi)}{((\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi))^2}\right) \\ &\quad - 2\tau E\text{tr}\left(\frac{\mathbf{W}(\mathbf{V}_2 - \mathbf{V}_{21})}{(\mathbf{h} + \xi)' \mathbf{B}(\mathbf{h} + \xi)}\right). \end{aligned} \quad (33)$$

Define $\mathbf{B}_1 = (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-\frac{1}{2}} \mathbf{G}'$ and $\mathbf{A} = (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{\frac{1}{2}} \mathbf{C}(\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{\frac{1}{2}}$

Note that $\mathbf{G}\mathbf{W}\mathbf{G}'_2 \mathbf{V}\mathbf{B} = \mathbf{G}\mathbf{C}\mathbf{G}' = \mathbf{B}'_1 \mathbf{A}\mathbf{B}_1$, $\mathbf{B}'_1 \mathbf{B}_1 = \mathbf{B}$. Using the inequality $b'ab \leq (b'b)\lambda_{\max}(a)$ for

symmetric a , and let

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

Then

$$\begin{aligned} \text{tr}(\mathbf{B}(\mathbf{h} + \xi)(\mathbf{h} + \xi)' \mathbf{G} \mathbf{W} \mathbf{G}'_2 \mathbf{V}) &= \frac{(\mathbf{h} + \xi)' \mathbf{B}'_1 (\mathbf{A} + \mathbf{A}') \mathbf{B}_1 (\mathbf{h} + \xi)}{2} \\ &\leq (\mathbf{h} + \xi)' \mathbf{B} (\mathbf{h} + \xi) \lambda_1. \end{aligned} \quad (34)$$

Define $A^* = (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{\frac{1}{2}} \mathbf{W} (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{\frac{1}{2}}$. Note that $\mathbf{G} \mathbf{W} \mathbf{G}' = \mathbf{B}'_1 \mathbf{A}^* \mathbf{B}_1$, $\mathbf{B}'_1 \mathbf{B}_1 = \mathbf{B}$, and let

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

we have

$$\begin{aligned} \text{tr}((\mathbf{h} + \xi)' \mathbf{G} \mathbf{W} \mathbf{G}' (\mathbf{h} + \xi)) &= \frac{(\mathbf{h} + \xi)' \mathbf{B}'_1 (\mathbf{A}^* + \mathbf{A}^{*'}) \mathbf{B}_1 (\mathbf{h} + \xi)}{2} \\ &\leq (\mathbf{h} + \xi)' \mathbf{B} (\mathbf{h} + \xi) \lambda_1^*. \end{aligned} \quad (35)$$

260 Plug (34) and (35) into (33) and use Jensen's inequality, then we have

$$\begin{aligned} E(\Psi^{*'} \mathbf{W} \Psi^*) &\leq R(\hat{\beta}_{SPFE}) + \tau^2 E\left(\frac{\lambda_1^*}{(\mathbf{h} + \xi)' \mathbf{B} (\mathbf{h} + \xi)}\right) + 4\tau E\left(\frac{\lambda_1}{(\mathbf{h} + \xi)' \mathbf{B} (\mathbf{h} + \xi)}\right) \\ &\quad - 2\tau E \text{tr}\left(\frac{(\mathbf{W} (\mathbf{V}_2 - \mathbf{V}_{21}))}{(\mathbf{h} + \xi)' \mathbf{B} (\mathbf{h} + \xi)}\right) \\ &= R(\hat{\beta}_{SPFE}) - E\left(\frac{\tau(2(\text{tr} \mathbf{W} (\mathbf{V}_2 - \mathbf{V}_{21}) - 2\lambda_1) - \lambda_1^* \tau)}{(\mathbf{h} + \xi)' \mathbf{B} (\mathbf{h} + \xi)}\right) \\ &\leq R(\hat{\beta}_{SPFE}) - \frac{\tau(2(\text{tr} \mathbf{W} (\mathbf{V}_2 - \mathbf{V}_{21}) - 2\lambda_1) - \lambda_1^* \tau)}{E((\mathbf{h} + \xi)' \mathbf{B} (\mathbf{h} + \xi))}. \end{aligned} \quad (36)$$

Since $\text{tr}(\mathbf{B} \mathbf{V}) = \text{tr}(G(\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-1} \mathbf{G}' \mathbf{V}) = q$. We have

$$\begin{aligned} E((\mathbf{h} + \xi)' \mathbf{B} (\mathbf{h} + \xi)) &= \mathbf{h}' \mathbf{B} \mathbf{h} + \text{tr}(\mathbf{B} \mathbf{V}) \\ &= \sigma_1^{-4} \delta' \Sigma' \mathbf{V}_1 (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-1} \mathbf{V}_1 \Sigma \delta + q. \end{aligned}$$

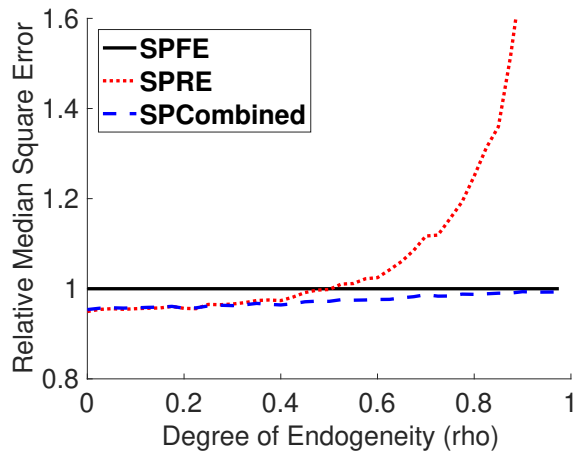
Substitute into (36), then we have

$$R(\hat{\beta}_{SP,c}) \leq R(\hat{\beta}_{SPFE}) - \frac{\tau(2(\text{tr} \mathbf{W} (\mathbf{V}_2 - \mathbf{V}_{21}) - 2\lambda_1) - \lambda_1^* \tau)}{\sigma_1^{-4} \delta' \Sigma' \mathbf{V}_1 (\mathbf{V}_1 + \mathbf{V}_2 - (\mathbf{V}_{21} + \mathbf{V}'_{21}))^{-1} \mathbf{V}_1 \Sigma \delta + q}.$$

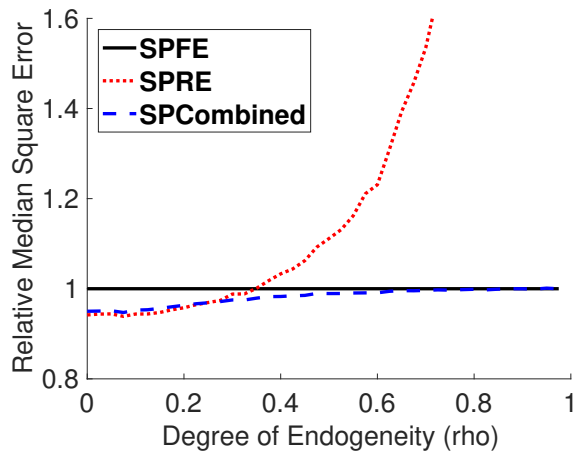
References

- [1] J. Horowitz, M. Markatou, Semiparametric estimation of regression models for panel data, *Review of Economic Studies* 63 (1) (1996) 145–168.
- 265 [2] A. Ullah, N. Roy, Nonparametric and semiparametric econometrics of panel data, *Handbook of Applied Economic Statistics*.
- [3] Q. Li, C. Hsiao, Testing serial correlation in semiparametric panel data models, *Review of Economic Studies* 87 (2) (1998) 207–237.
- [4] Q. Li, A. Ullah, Estimating partially linear panel data models with one-way error components, *Econometric Reviews* 17 (2) (1998) 145–166.
- 270 [5] B. Baltagi, D. Li, Series estimation of partially linear panel data models with fixed effects, *Annals of Economics and Finance* 3 (1) (2002) 103–116.
- [6] R. C. D.J. Henderson, Q. Li, Nonparametric estimation and testing of fixed effects panel data models, *Journal of Econometrics* 144 (2008) 257–275.
- 275 [7] L. Su, A. Ullah, Profile likelihood estimation of partially linear panel data models with fixed effects, *Economics Letters* 92 (1) (2006) 75–81.
- [8] D. Henderson, C. Parmeter, *Applied nonparametric econometrics*, Cambridge University Press.
- [9] P. Rilston, A. Ullah, Nonparametric estimation of response coefficients, *Communications in Statistics-Theory and Methods* 18 (7) (1989) 2615–2627.
- 280 [10] A. Pagan, A. Ullah, *Nonparametric econometrics*, Cambridge University Press.
- [11] A. Ullah, Non-parametric estimation of econometric functionals, *Canadian Journal of Economics* 21 (1988) 625–658.
- [12] A. U. L. Su, Y. Wang, Nonparametric regression estimation with general parametric error covariance: a more efficient two-step estimator, *Empirical economics* 45 (2) (2013) 1009–1024.

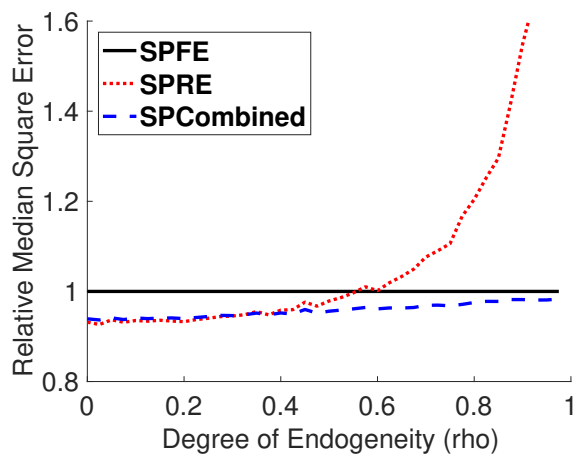
- 285 [13] I. M. L. Su, A. Ullah, Local linear gmm estimation of functional coefficient iv models with an application to estimating the rate of return to schooling, *Journal of Business and Economic Statistics* 31 (2) (2013) 184–207.
- [14] B. Hansen, A stein-like 2sls estimator, *Econometric Reviews* 36 (2017) 840–852.
- [15] J. Hausman, Specification tests in econometrics, *Econometrica* 46 (1994) 1251–1271.
- 290 [16] C. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1 (1956) 399.
- [17] A. Munnell, How does public infrastructure affect regional economic performance?, *New England Economic Review* (1990) 11–33.
- 295 [18] P. Evans, G. Karras, Are government activities productive? evidence from a panel of us states, *Review of Economics and Statistics* 76 (1) (1994) 1–11.
- [19] D. Holtz-Eakin, Public-sector capital and the productivity puzzle, *Review of Economics and Statistics* 76 (1) (1994) 12–21.
- [20] B. Baltagi, N. Pinnoi, Public capital stock and state productivity growth: further evidence from an error components model, *Empirical Economics* 20 (2) (1995) 351–359.
- 300



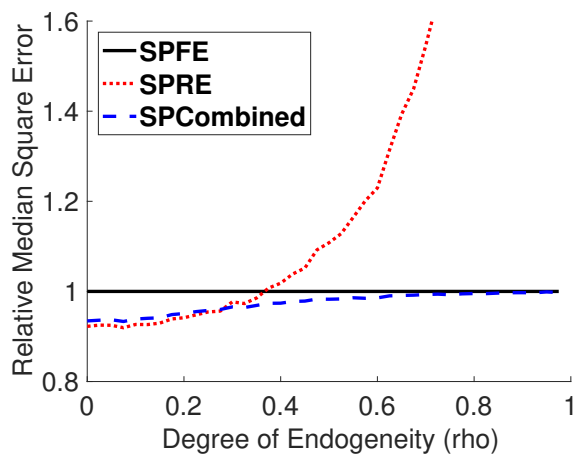
(a) $n = 20, T = 3, q = 8, p = 1, \rho^* = .86$



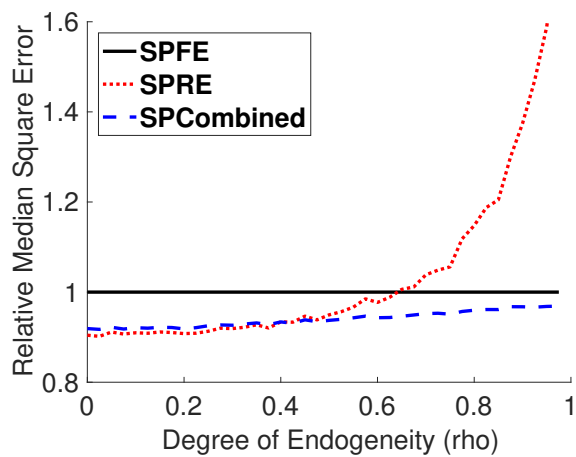
(b) $n = 100, T = 3, q = 8, p = 1, \rho^* = .86$



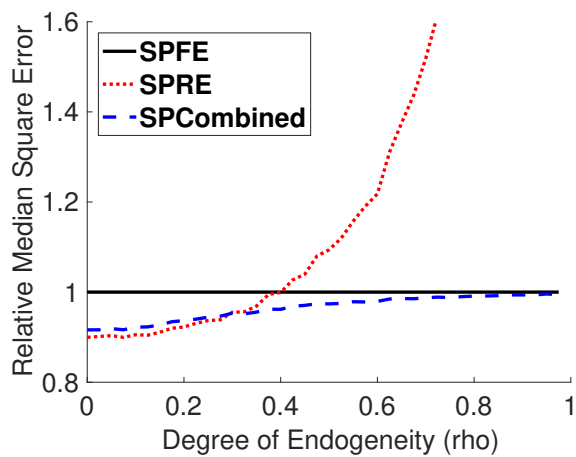
(c) $n = 20, T = 3, q = 8, p = 1, \rho^* = .74$



(d) $n = 100, T = 3, q = 8, p = 1, \rho^* = .74$

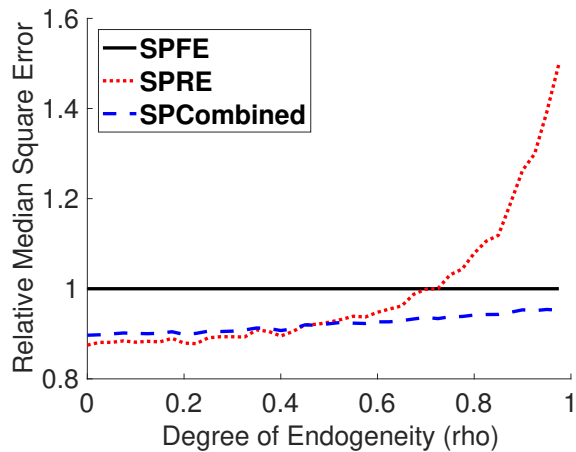


(e) $n = 20, T = 3, q = 8, p = 1, \rho^* = .61$

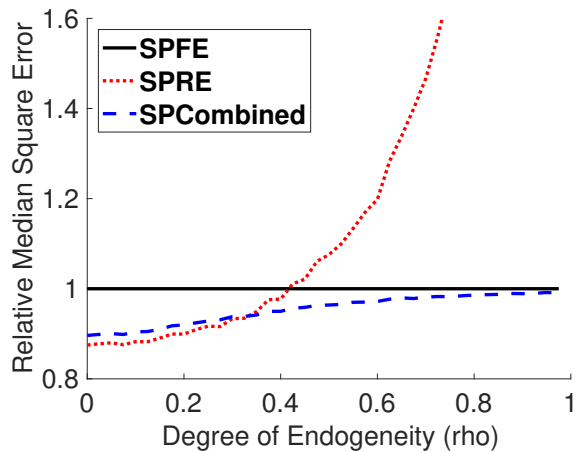


(f) $n = 100, T = 3, q = 8, p = 1, \rho^* = .61$

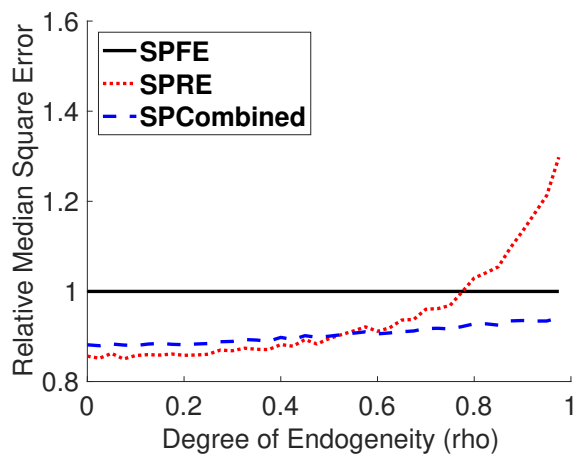
Figure 1: Median Squared Error of SPFE, SPRE and SPCombined Estimators, $n = \{20, 100\}$, $T = 3, q = 8, p = 1, \rho^* = \{.61, .74, .86\}$.



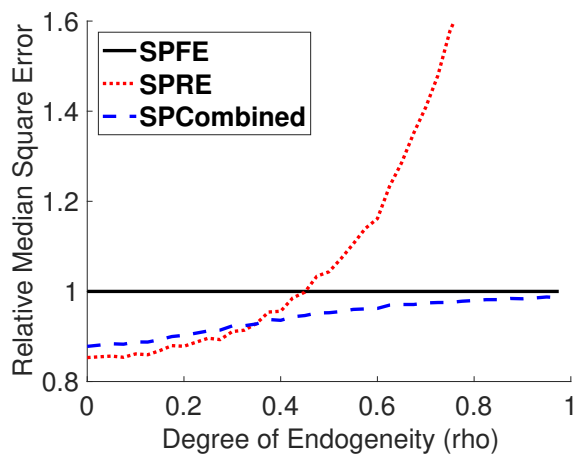
(a) $n = 20, T = 3, q = 8, p = 1, \rho^* = .50$



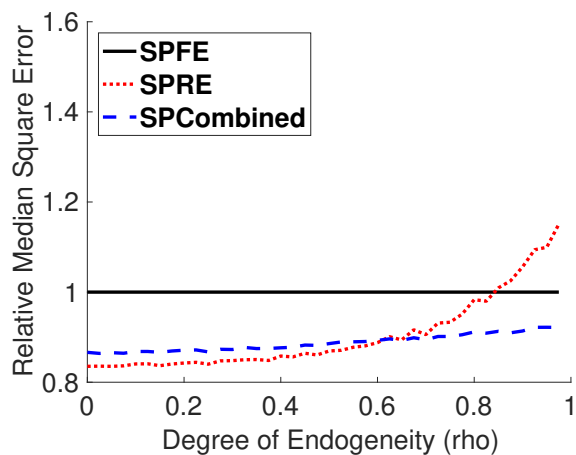
(b) $n = 100, T = 3, q = 8, p = 1, \rho^* = .50$



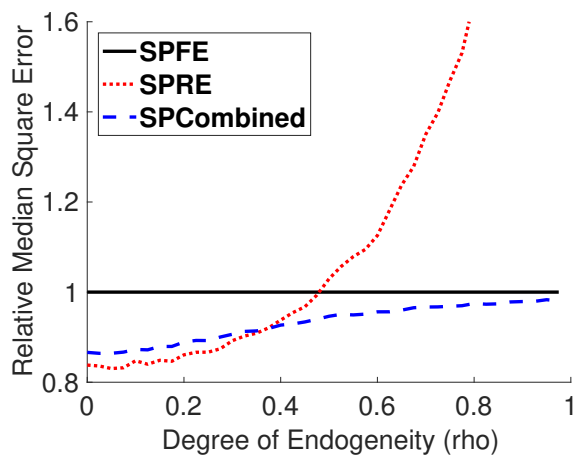
(c) $n = 20, T = 3, q = 8, p = 1, \rho^* = .41$



(d) $n = 100, T = 3, q = 8, p = 1, \rho^* = .41$

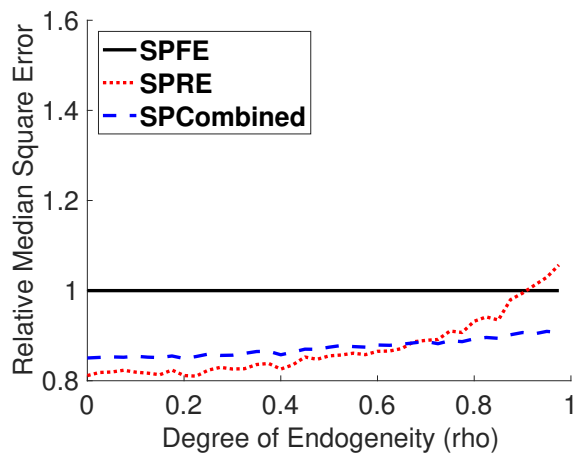


(e) $n = 20, T = 3, q = 8, p = 1, \rho^* = .34$

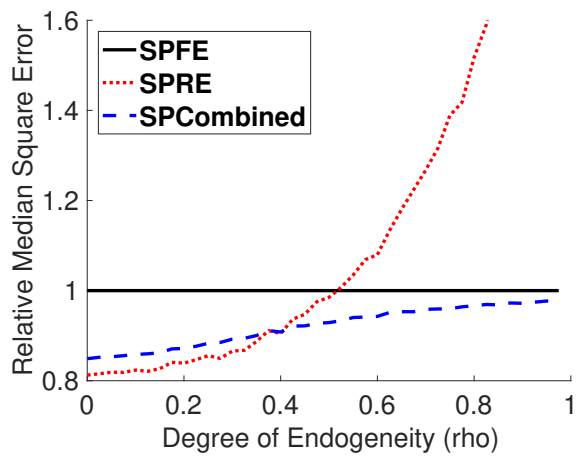


(f) $n = 100, T = 3, q = 8, p = 1, \rho^* = .34$

Figure 2: Median Squared Error of SPFE, SPRE and SPCombined Estimators, $n = \{20, 100\}$, $T = 3, q = 8, p = 1, \rho^* = \{.34, .41, .50\}$.

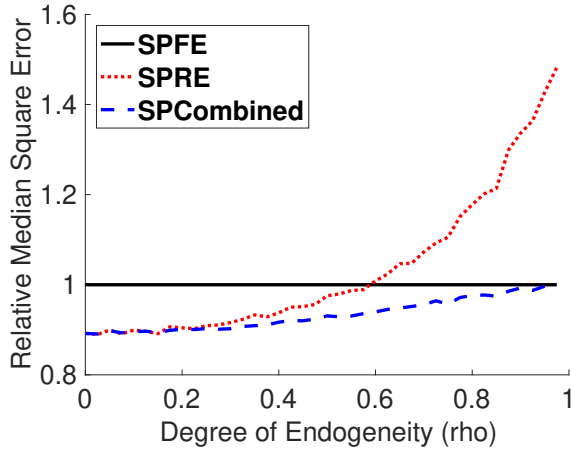


(a) $n = 20, T = 3, q = 8, p = 1, \rho^* = .28$

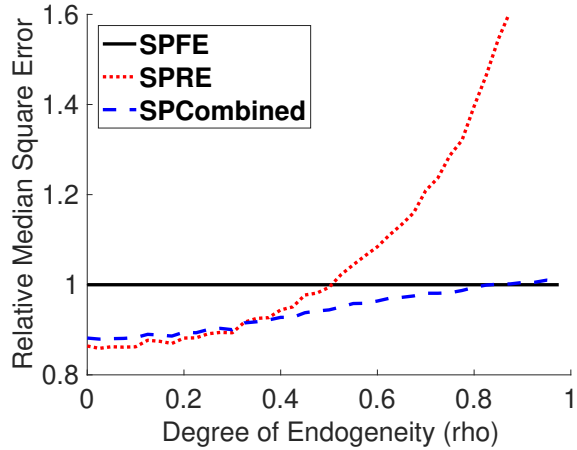


(b) $n = 100, T = 3, q = 8, p = 1, \rho^* = .28$

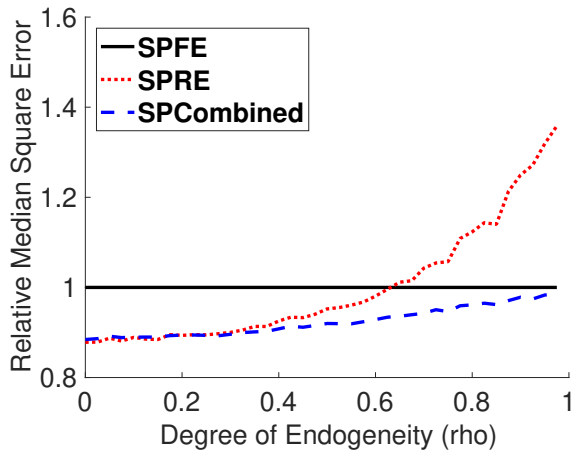
Figure 3: Median Squared Error of SPFE, SPRE and SPCombined Estimators, $n = \{20, 100\}, T = 3, q = 8, p = 1, \rho^* = \{.28\}$.



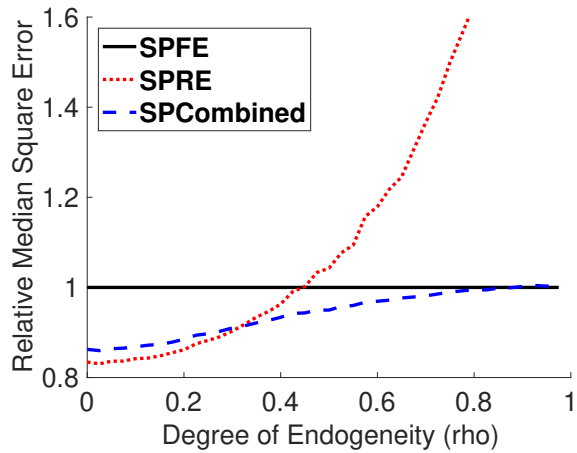
(a) $n = 20, T = 3, q = 8, p = 2, \rho^* = .86$



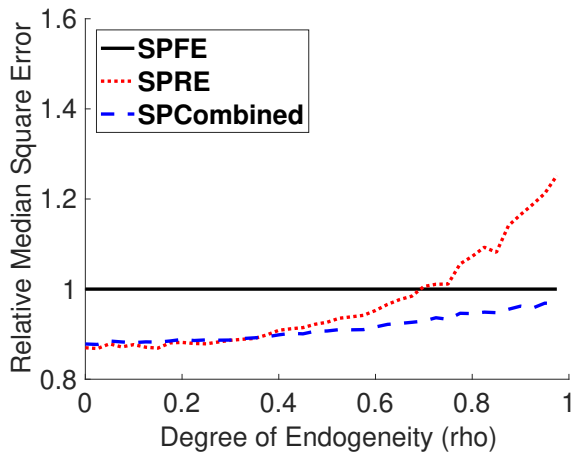
(b) $n = 100, T = 3, q = 8, p = 2, \rho^* = .86$



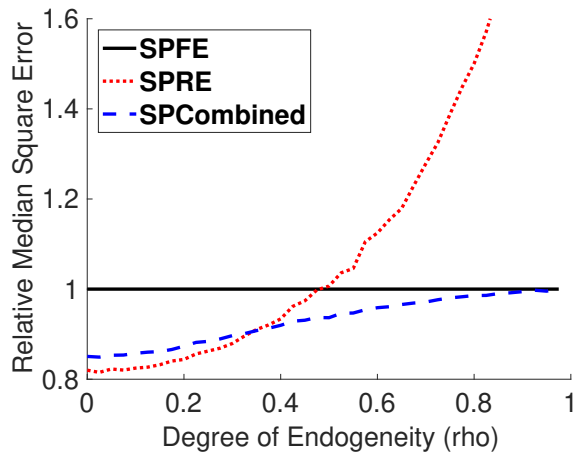
(c) $n = 20, T = 3, q = 8, p = 2, \rho^* = .74$



(d) $n = 100, T = 3, q = 8, p = 2, \rho^* = .74$

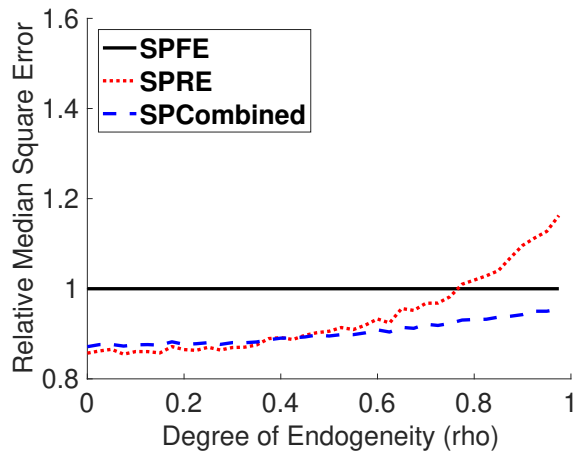


(e) $n = 20, T = 3, q = 8, p = 2, \rho^* = .61$

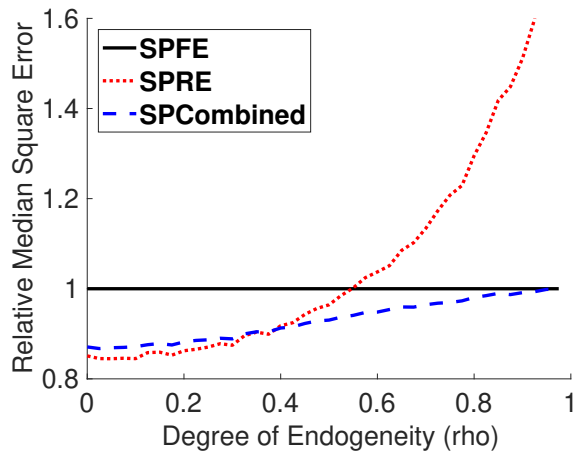


(f) $n = 100, T = 3, q = 8, p = 2, \rho^* = .61$

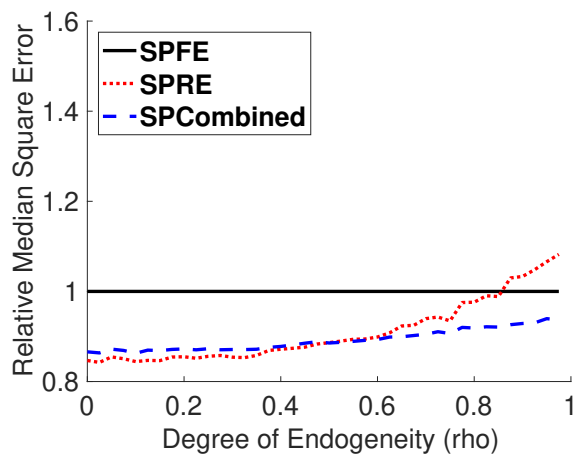
Figure 4: Median Squared Error of SPFE, SPRE and SPCombined Estimators, $n = \{20, 100\}$, $T = 3, q = 8, p = 2, \rho^* = \{.61, .74, .86\}$.



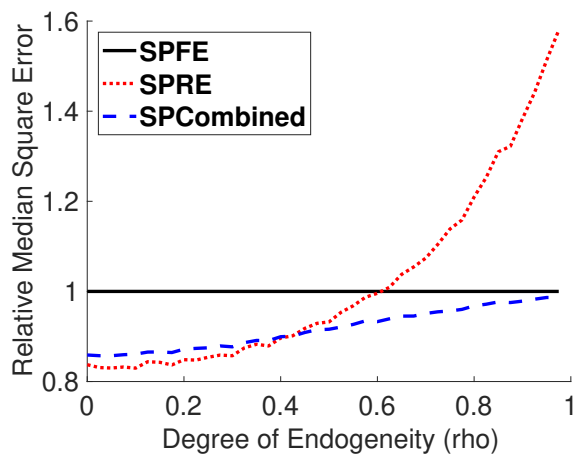
(a) $n = 20, T = 3, q = 8, p = 2, \rho^* = .50$



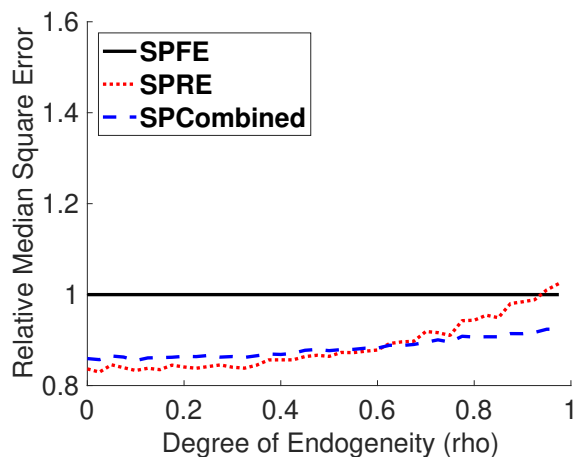
(b) $n = 100, T = 3, q = 8, p = 2, \rho^* = .50$



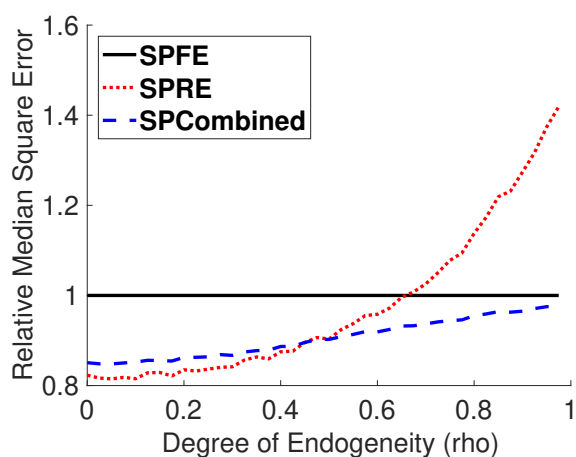
(c) $n = 20, T = 3, q = 8, p = 2, \rho^* = .41$



(d) $n = 100, T = 3, q = 8, p = 2, \rho^* = .41$

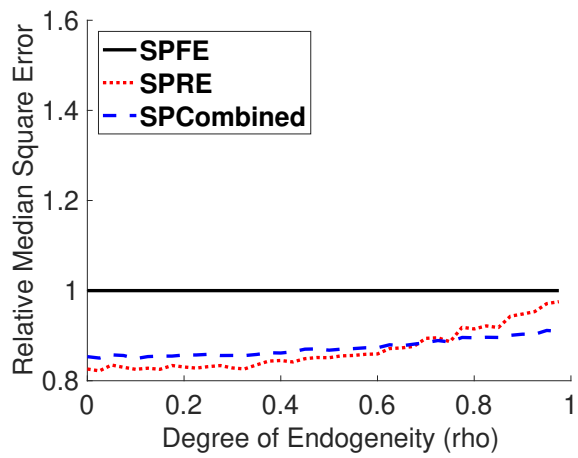


(e) $n = 20, T = 3, q = 8, p = 2, \rho^* = .34$

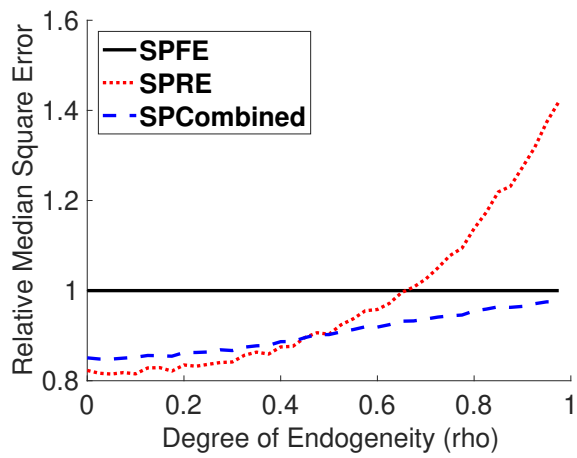


(f) $n = 100, T = 3, q = 8, p = 2, \rho^* = .34$

Figure 5: Median Squared Error of SPFE, SPRE and SPCombined Estimators, $n = \{20, 100\}$, $T = 3, q = 8, p = 2, \rho^* = \{.34, .41, .50\}$.



(a) $n = 20, T = 3, q = 8, p = 2, \rho^* = .28$



(b) $n = 100, T = 3, q = 8, p = 2, \rho^* = .28$

Figure 6: Median Squared Error of SPFE, SPRE and SPCombined Estimators, $n = \{20, 100\}, T = 3, q = 8, p = 2, \rho^* = \{.28\}$.