

ARTICLE TEMPLATE

Model Averaging Estimation of Panel Data Models with Many Instruments and Boosting

Hao Hao^a Bai Huang^b and Tae-hwy Lee^c

^aFord Motor Company, Dearborn, Michigan, USA; ^bCentral University of Finance and Economics, Beijing, China; ^c University of California, Riverside, Riverside, California, USA

ARTICLE HISTORY

Compiled April 7, 2022

ABSTRACT

Applied researchers often confront two issues when using the fixed effect-two-stage least squares (FE-2SLS) estimator for panel data models. One is that it may lose its consistency due to too many instruments. The other is that the gain of using FE-2SLS may not exceed its loss when the endogeneity is weak. In this paper, an L_2 Boosting regularization procedure for panel data models is proposed to tackle the many instruments issue. We then construct a Stein-like model-averaging estimator to take advantage of FE and FE-2SLS-Boosting estimators. Finite sample properties are examined in Monte Carlo and an empirical application is presented.

KEYWORDS

FE-2SLS; weak endogeneity; combined estimator; many instruments; L_2 Boosting; FE-2SLS-Boosting

1. Introduction

In the social and behavioral sciences, research on the causal effect of one variable on another is far from settled. Regression analysis may fail to give a reliable estimate of the causal effect due to many reasons. In a panel data or longitudinal data model, the fixed-effects estimator may suffer from the endogeneity issues that arise due to the correlated unobserved effects or the 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 parameters. The resulted bias can not be removed via the differencing estimation. The commonly used technique to overcome this problem is to use instruments for those endogenous explanatory variables. The most basic approach to doing this is the two-stage least squares (2SLS) estimation. In most applied [settings](#), many analysts have favored the 2SLS approach as shown in [2, 10, 19]. These papers use instrumental variable (IV) procedures to estimate the parameters of the panel data model with endogenous regressors. The IV estimation is first introduced by [28]. A major breakthrough comes out later with the proposition of 2SLS by [6, 27]. [4] studies the widely used fixed effects-two-stage least squares (FE-2SLS) estimator which is very popular mainly because of its efficiency among the class of IV estimators under conditional homoscedasticity.

However, there are two issues that often arise in practice. The first issue is when the endogeneity is weak: the consistency gain of using FE-2SLS may not exceed its efficiency loss. The other issue is when there are too many instruments: FE-2SLS may converge to FE, losing its consistency even when all the instruments are relevant and valid. In this paper, we propose a two-step procedure to handle these two issues.

First, if the endogeneity is strong, FE is inconsistent and the gains from FE-2SLS may be big. However, if the endogeneity is weak, the consistency gain from using FE-2SLS can be small. As FE-2SLS is less efficient than FE, there is a trade-off between the inefficient FE-2SLS estimation and the inconsistent FE estimation. Under this scenario, FE-2SLS and FE estimations can be combined to take advantage of both estimators to obtain a combined estimator with a lower total mean squared error risk.

Second, 2SLS needs enough number of IVs. Otherwise, the finite order integer moments of the 2SLS estimator may not exist [21]. [22] use the Web of Science Database across all papers published between January 2000 and October 2018 on the topic of instruments in the top five economic journals, and find that more than half of these empirical studies report results from a specification based on multiple IVs in the use of 2SLS. Their results confirm that empirical researchers often use more number of IVs than that of endogenous variables in the 2SLS estimation. However, applied researchers often confront problems when having a large number of instruments in panel data models, e.g., in the case when the IVs are formed from lagged endogenous variables. As shown in [7], too many instruments will cause inconsistency in the 2SLS estimator. Since the FE-2SLS estimator is sensitive to a large number of instruments, a regularization method is necessary to reduce the dimension of instruments for consistent estimation. For the cross-sectional models, there is a rich literature on the regularization approaches. For example, in [8, 9], Lasso is used for instrument selection. [12, 13, 15] extend the Lasso-type regularization to the generalized method of moments (GMM). Other than these regularized methods, [14] use information criteria for moment selection, while [23] use L_2 Boosting for instrumental variable selection. For the panel data models, little research is available on data-driven regularization. In this paper, we extend L_2 Boosting by [11] for regularization of the FE-2SLS estimator for a large n and large T panel data model. We call this estimator the “FE-2SLS-Boosting”. We perform a series of Monte Carlo simulations to examine the issue that FE-2SLS becomes inconsistent if too many instruments are used and show that the proposed FE-2SLS-Boosting restores the consistency when there are many instruments.

We finally apply the proposed two-step approach to U.S. real house prices data and examine the extent to which house price fluctuations are driven by fundamental factors. This application is policy relevant. The empirical results indicate that the house prices that deviate from the equilibrium will eventually revert [20] (hereafter HPY). In this paper, the original HPY panel data of 49 States over the 29 year period of 1975–2003 is used to discuss the responses of the U.S. housing market to three fluctuations in incomes, population, and interest rates.

In sum, our procedure handles these two problems in the following two steps: First, L_2 Boosting (modified for a panel data model) is applied to regularize FE-2SLS for selecting instruments. Then FE and FE-2SLS-Boosting is combined to deal with the weak endogeneity. Using too many instruments, FE-2SLS can be inconsistent and can be worse than FE. Regularization by L_2 Boosting maintains the consistency for the FE-2SLS-Boosting estimator. The combination of FE and FE-2SLS-Boosting is shown to further improve over the FE-2SLS-Boosting estimator. Our Monte Carlo experiment demonstrates the relative performance of FE, FE-2SLS, FE-2SLS-Boosting, and the regularized combined (model-averaging) estimators.

The rest of the paper is organized as follows: In Section 2, FE and FE-2SLS estimators are discussed. In Section 3 we present a combined estimator that combines the FE and FE-2SLS estimators. In Section 4, we discuss the needs and benefits of using regularization method when there are many instruments and introduce FE-2SLS-Boosting. Section 5 provides a Monte Carlo simulation. Section 6 presents an empirical example of real house prices in the US. Section 7 concludes with a brief discussion. All proofs are in appendices A and B.

2. FE and FE-2SLS

Consider the following large n and large T panel data model with fixed effects:

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

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

$$y = X\beta + D\alpha + u, \quad (2)$$

where $y = (y_{11}, \dots, y_{1T}, \dots, y_{n1}, \dots, y_{nT})'$ is $nT \times 1$, $X = (x_{11}, \dots, x_{1T}, \dots, x_{n1}, \dots, x_{nT})'$ is $nT \times q$, u is $nT \times 1$, $D \equiv I_n \otimes \iota_T$ 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 Equation (2) by $Q \equiv I_{nT} - D(D'D)^{-1}D'$:

$$Qy = QX\beta + QD\alpha + Qu, \quad (3)$$

where $QD = 0$. Noting that Q is idempotent, $\hat{\beta}_{\text{FE}}$ can be obtained as

$$\hat{\beta}_{\text{FE}} = (X'QX)^{-1} X'Qy. \quad (4)$$

According to Corollary 1 in [3], as $n, T \rightarrow \infty$, under i.i.d. assumption of u_{it} , the asymptotic distribution of $\hat{\beta}_{\text{FE}}$ is

$$\sqrt{nT} (\hat{\beta}_{\text{FE}} - \beta) \xrightarrow{d} N(0, V_1), \quad (5)$$

Let plim denote the probability limit operator as $n, T \rightarrow \infty$, and $V_1 = \sigma_u^2 \left(\text{plim} \frac{X'QX}{nT} \right)^{-1}$. The feasible estimator of $\hat{\sigma}_u^2$ of σ_u^2 can be obtained by first running the ordinary least squares (OLS) regression y on X to get $\hat{\epsilon}_{it} = y_{it} - x_{it}\hat{\beta}_{\text{OLS}}$ as the OLS residual and $\hat{\beta}_{\text{OLS}} = (X'X)^{-1} X'y$. This gives

$$\hat{\sigma}_u^2 = \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T (\hat{\epsilon}_{it} - \bar{\hat{\epsilon}}_i)^2. \quad (6)$$

The endogeneity occurs due to the (i) correlation of α_i with x_{it} or (ii) correlation of u_{it} with x_{it} . We consider the latter case here for which the FE estimator becomes

inconsistent. With u_{it} and x_{it} correlated, the vector x_{it} is endogenous. Performing 2SLS on (3) with QZ as the set of instruments

$$Z'QQy = Z'QQX\beta + Z'QQu, \quad (7)$$

one gets the FE-2SLS estimator

$$\hat{\beta}_{\text{FE-2SLS}} = (X'H_ZX)^{-1} X'H_Zy, \quad (8)$$

where $H_Z = QZ(Z'QZ)^{-1}Z'Q$. Under both large n and large T , the asymptotic distribution of $\hat{\beta}_{\text{FE-2SLS}}$ follows

$$\sqrt{nT} \left(\hat{\beta}_{\text{FE-2SLS}} - \beta \right) \xrightarrow{d} N(0, V_2), \quad (9)$$

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

Remark 1: If a subset of regressors is treated as endogenous, consider the following structural equation of a panel data model:

$$y = X\beta + D\alpha + u, \quad (10)$$

where $X = (X_1 \ Z_1)$ and $\beta = (\beta_1 \ \beta_2)$. Let X_1 be q_1 endogenous variables, Z_1 be ℓ_1 included exogenous variables, and $q = q_1 + \ell_1$. Let $Z = (Z_1 \ Z_2)$ be the set of ℓ ($= \ell_1 + \ell_2$) exogenous variables (instrumental variables). This equation is identified if $\ell \geq q$, and therefore $\ell_2 \geq q_1$. In this case, QZ can be used as the set of instruments to get the FE-2SLS estimator as

$$\hat{\beta}_{\text{FE-2SLS}} = (X'H_ZX)^{-1} X'H_Zy,$$

with $H_Z = QZ(Z'QZ)^{-1}Z'Q$.

3. Combined Estimator of FE and FE-2SLS

The FE-2SLS estimator is preferred to the FE estimator as it is consistent under endogeneity (which can be ensured by the regularization of many instruments), while the FE estimator is inconsistent. However, in small samples, FE-2SLS can have a much larger variance so FE can have smaller mean squared errors (MSE) especially when the extent of endogeneity is not severe. Motivated by this observation, we follow [17] to propose the following combined estimators $\hat{\beta}_c$, which is the weighted average of FE and FE-2SLS estimators with the weights depending on the Hausman statistic [18]. The combination of FE and FE-2SLS is expected to improve the estimation precision. Let

$$\hat{\beta}_c = w\hat{\beta}_{\text{FE}} + (1 - w)\hat{\beta}_{\text{FE-2SLS}}, \quad (11)$$

where $\hat{\beta}_{\text{FE-2SLS}}$ is the FE-2SLS estimator, and

$$w = \begin{cases} \frac{\tau}{H_{nT}} & \text{if } H_{nT} \geq \tau \\ 1 & \text{if } H_{nT} < \tau, \end{cases} \quad (12)$$

$$H_{nT} = nT(\hat{\beta}_{\text{FE-2SLS}} - \hat{\beta}_{\text{FE}})' \left(\hat{V}_2 - \hat{V}_1 \right)^{-1} (\hat{\beta}_{\text{FE-2SLS}} - \hat{\beta}_{\text{FE}}), \quad (13)$$

where H_{nT} is the Hausman statistics, τ is a shrinkage parameter, $\hat{V}_1 = \hat{\sigma}_u^2 \left(\frac{X'QX}{nT} \right)^{-1}$ and $\hat{V}_2 = \hat{\sigma}_u^2 \left(\frac{X'H_ZX}{nT} \right)^{-1}$.

The model has the following reduced-form representation for x_{it} as

$$x_{it} = \Pi' z_{it} + v_{it}, \quad (14)$$

with $E(z_{it}v_{it}) = 0$. Next, write the structural equation error u_{it} as a linear function of v_{it} and ε_{it}

$$u_{it} = v_{it}'\rho + \varepsilon_{it}, \quad (15)$$

with $E(v_{it}\varepsilon_{it}) = 0$. We use the local asymptotic approach. ρ is local to zero

$$\rho = \frac{1}{\sqrt{nT}}\delta, \quad (16)$$

where δ is a $q \times 1$ localizing parameter, which indexes the degree of correlation between u_{it} and v_{it} . δ (and thus ρ) controls the degree of endogeneity.

We make the following assumptions:

Assumption A1. x_i are distributed independently across i . $E\|x_{it}\|^4 < \infty$.

Assumption A2. u_{it} are i.i.d. over t and i . α_i are independent over i . α_i and x_{it} are independent of u_{it} for any i and t . $E(u_{it}^4) < \infty$, $E\|v_{it}\|^4 < \infty$, $E(\varepsilon_{it}^4) < \infty$.

Assumption A3. $\sigma_u^2 (\text{plim}_{n,T \rightarrow \infty} \frac{1}{nT} X'QX)^{-1} = V_1$, $\sigma_u^2 (\text{plim}_{n,T \rightarrow \infty} \frac{1}{nT} X'H_ZX)^{-1} = V_2$. $\hat{\sigma}_u^2 = \sigma_u^2 + o_p(1)$.

Assumption A4. $E\|z_{it}\|^4 < \infty$; $\text{rank}(\Pi) = q$.

In Assumptions A1, x_i are distributed not necessarily identically across different i . More details can be found in [1]. For a vector or matrix A , its norm is defined as $\|A\| = (\text{tr}(A'A))^{1/2}$. Assumption A2 rules out the cases in which the regressors include lagged dependent variables, although a dynamic panel introduces an even larger number of instruments for which the proposed method of using L_2 Boosting to select instruments would be useful. More discussion of the i.i.d. assumption about the idiosyncratic errors in a large T panel is provided in [3]. Assumption A3 is standard in the literature. Assumption A4 is the rank condition on Π to ensure that the coefficient β is identified.

Denote $V = (v'_{11}, \dots, v'_{1T}, \dots, v'_{n1}, \dots, v'_{nT})'$, and $\Sigma = E(VV')$. We then have the following theorem for the joint asymptotic normality of FE and FE-2SLS estimators.

Theorem 1: Under Assumptions A1-A4,

$$\sqrt{nT} \begin{pmatrix} \hat{\beta}_{\text{FE}} - \beta \\ \hat{\beta}_{\text{FE-2SLS}} - \beta \end{pmatrix} \xrightarrow{d} h + \xi, \quad (17)$$

where

$$h = \begin{pmatrix} \sigma_u^{-2} V_1 \text{tr}(Q\Sigma) \delta \\ 0 \end{pmatrix}, \quad (18)$$

$$\xi \sim N(0, V), \quad (19)$$

$$V = \begin{pmatrix} V_1 & V_1 \\ V_1 & V_2 \end{pmatrix}. \quad (20)$$

Furthermore,

$$H_{nT} \xrightarrow{d} (h + \xi)' B (h + \xi), \quad (21)$$

$$\sqrt{nT} (\hat{\beta}_c - \beta) \xrightarrow{d} \Psi = G_2' \xi - \left(\frac{\tau}{(h + \xi)' B (h + \xi)} \right)_1 G' (h + \xi), \quad (22)$$

where $B = G(V_2 - V_1)^{-1} G'$, $G = (-I \ I)'$, and $G_2 = (0 \ I)'$, and $(a)_1 = \min[1, a]$. \square

Proof. See Appendix A.

Theorem 1 extends [17] for the panel data models and gives expressions for the joint asymptotic distribution of $\hat{\beta}_{\text{FE}}$ and $\hat{\beta}_{\text{FE-2SLS}}$ estimators, the Hausman statistic, and the combined estimator under the local-to-exogeneity setup in Equation (16). These alternative estimators (FE, FE-2SLS, and combined estimators) are compared in the asymptotic risk. Following [16, 17], the asymptotic risk of any sequence of estimators $\hat{\beta}$ of β is defined as

$$R(\hat{\beta}, \beta, W) = \lim_{n, T \rightarrow \infty} E \left[nT (\hat{\beta} - \beta)' W (\hat{\beta} - \beta) \right] \equiv R(\hat{\beta}). \quad (23)$$

Denote the largest eigenvalue $\lambda_1 \equiv \lambda_{\max}(W(V_2 - V_1))$ of the matrix $W(V_2 - V_1)$ and the ratio $d \equiv \text{tr}(W(V_2 - V_1)) / \lambda_1$. The following theorem is an extension of the Theorem 2 of [17] for the panel data model with general weighting matrix W .

Theorem 2: Under Assumptions A1-A4, if

$$d > 2 \text{ and } 0 < \tau \leq 2(d - 2), \quad (24)$$

then

$$R\left(\hat{\beta}_{\text{FE-2SLS}}\right) = \text{tr}(WV_2), \quad (25)$$

and

$$R\left(\hat{\beta}_c\right) < R\left(\hat{\beta}_{\text{FE-2SLS}}\right) - \frac{\tau\lambda_1(2(d-2) - \tau)}{\sigma_u^{-4}\delta'\text{tr}(Q\Sigma)V_1(V_2 - V_1)^{-1}V_1\text{tr}(Q\Sigma)\delta + q}. \quad (26)$$

□

Proof. See Appendix A.

Equation (26) shows that the asymptotic risk of the combined estimator is strictly less than that of the FE-2SLS estimator, so long as the shrinkage parameter τ satisfies the condition (24). In the special case $W = (V_2 - V_1)^{-1}$, the condition (24) simplifies to $q > 2$ and $0 < \tau \leq 2(q - 2)$. The assumption $q > 2$ is Stein's (1956) classic condition [26] for shrinkage.

The following two corollaries are obtained with $W = (V_2 - V_1)^{-1}$.

Corollary 1: *Under Assumptions A1-A4 and the local-to-exogeneity setup,*

$$R\left(\hat{\beta}_{\text{FE}}\right) = \text{tr}(WV_1) + \sigma_u^{-4}\delta'\text{tr}(Q\Sigma)V_1WV_1\text{tr}(Q\Sigma)\delta,$$

and

$$\begin{cases} R\left(\hat{\beta}_{\text{FE}}\right) \leq R\left(\hat{\beta}_{\text{FE-2SLS}}\right) & \text{if } \sigma_u^{-4}\delta'\text{tr}(Q\Sigma)V_1WV_1\text{tr}(Q\Sigma)\delta \leq q \\ R\left(\hat{\beta}_{\text{FE}}\right) > R\left(\hat{\beta}_{\text{FE-2SLS}}\right) & \text{otherwise.} \quad \square \end{cases}$$

Corollary 1 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.

Corollary 2: *Under Assumptions A1-A4, If $q < \sigma_u^{-4}\delta'\text{tr}(Q\Sigma)V_1WV_1\text{tr}(Q\Sigma)\delta$, $d > 2$, and $0 < \tau \leq 2(d - 2)$, then $R\left(\hat{\beta}_c\right) < R\left(\hat{\beta}_{\text{FE}}\right)$.* □

Corollary 2 indicates that, when endogeneity is strong, $d > 2$, and $0 < \tau \leq 2(d - 2)$, the combined estimator is better than both the FE and FE-2SLS estimators.

4. FE-2SLS-Boosting

According to [7], the 2SLS estimator is inconsistent when the number of instruments ℓ is large. By replacing X with $X^* = (x_{11}^*, \dots, x_{1T}^*, \dots, x_{n1}^*, \dots, x_{nT}^*)' = QX$ and Z with $Z^* = (z_{11}^*, \dots, z_{1T}^*, \dots, z_{n1}^*, \dots, z_{nT}^*)' = QZ$, we extend the Bekker theorem on the inconsistency result of 2SLS to FE-2SLS in the panel data models, i.e., the FE-2SLS is inconsistent unless $\frac{\ell}{nT} \rightarrow 0$. As shown in Appendix B, while $\hat{\beta}_{\text{FE}}$ is inconsistent due to the endogeneity, $\hat{\beta}_{\text{FE-2SLS}}$ may also be inconsistent due to the large number of instruments.

The reduced form equation for x_{it}^* can be written as

$$x_{it}^* = \Pi' z_{it}^* + v_{it} \quad (27)$$

with $E(z_{it}^* v_{it}) = 0$. Instruments z_{it}^* is $\ell \times 1$. Π is an $\ell \times q$ matrix, and Π_k is the k^{th} column of Π for $k = 1, \dots, q$.

In order to ensure the consistency of the FE-2SLS estimator, we extend the regularization method L_2 Boosting by [11] to Equation (27) for the panel data model. We use this regularization method to select a subset of instruments and compute the FE-2SLS estimator based on the selected instruments. We refer to the FE-2SLS estimator using L_2 Boosting as “FE-2SLS-Boosting”.

Let m denote the m^{th} iteration in the Boosting procedure, and \bar{M} denote the maximum number of iteration. At each iteration m , we have a weak learner $f_{m,it,k}$ that gives a less accurate estimation on X_k^* . But the summation of the weak learners up to step \bar{M} will give a strong estimation on X_k^* . We refer to this summation as a strong learner $F_{\bar{M},it,k} = \sum_{m=0}^{\bar{M}} c_m f_{m,it,k}$, where c_m is the learning rate that controls the step of the learning process in Boosting. To simplify the notation, we drop the subscription k in the procedure. So, $f_{m,it} = f_{m,it,k}$ and $F_{m,it} = F_{m,it,k}$. However, the procedure is repeated for each $k \in \{1, \dots, q\}$. The algorithm for instrument selection using L_2 Boosting for each X_k^* is as follows:

- (1) When $m = 0$, the initial estimate for $x_{it,k}^*$ is

$$f_0 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T x_{it,k}^*, \quad (28)$$

which is the simple mean of $x_{it,k}^*$. Denote $F_{0,it} = f_0$ for all i and t .

- (2) For each iteration $m = 1, \dots, \bar{M}$:

- (a) Compute the “current residual”, $\hat{v}_{m,it} = x_{it,k}^* - F_{m-1,it}$.
- (b) Regress the current residual $\hat{v}_{m,i}$ on each instrument $z_{j,it}^*$, for $j = 1, \dots, \ell$.

The estimators $\hat{\Pi}_{kj}$ are solved as

$$\hat{\Pi}_{m,kj} = \arg \min_{\Pi_{kj}} \sum_{i=1}^n \sum_{t=1}^T (\hat{v}_{m,it} - \Pi_{kj} z_{j,it}^*)^2. \quad (29)$$

We select the instrument that has the minimum sum of squared residuals,

such that

$$j_m = \arg \min_{j \in \{1, \dots, \ell\}} \sum_{i=1}^n \sum_{t=1}^T \left(\hat{v}_{m,it} - \hat{\Pi}_{m,kj} z_{j,it}^* \right)^2. \quad (30)$$

(c) The weak learner is

$$f_{m,it} = \hat{\Pi}_{kj_m} z_{j_m,it}^*, \quad (31)$$

where $z_{j_m,it}^*$ is the instrument that is selected at iteration m .

(d) The strong learner $F_{m,it}$ is updated as

$$F_{m,it} = F_{m-1,it} + c_m f_{m,it}, \quad (32)$$

where $c_m > 0$ is a learning rate.

(3) We repeat Steps 1 and 2 for $k = 1, \dots, q$.

A stopping rule is necessary in order to avoid over-fitting. Extending [11], we introduce a modified AIC for panel data models to choose the optimal number of iterations \hat{M} . Let $\hat{V}_m = (\hat{v}_{m,11} \dots \hat{v}_{m,nT})'$, $f_m = (f_{m,11} \dots f_{m,nT})'$, $F_m = (F_{m,11} \dots F_{m,nT})'$. We define $P_m = Z_{j_m}^* (Z_{j_m}^{\prime*} Z_{j_m}^*)^{-1} Z_{j_m}^{\prime*}$ to be an $nT \times nT$ matrix. From Equation (31),

$$\begin{aligned} Z_{j_m}^* \hat{\Pi}_{kj_m} &= P_m \hat{V}_m \\ f_m &= P_m (X_k^* - F_{m-1}). \end{aligned} \quad (33)$$

When $m = 0$, P_{j_0} is an $nT \times nT$ matrix of $\frac{1}{nT}$. Then the strong learner at each iteration m is

$$\begin{aligned} F_m &= F_{m-1} + c_m P_m (X_k^* - F_{m-1}) \\ &= \left[I_{nT} - \prod_{a=0}^m (I_n - c_{j_a} P_{j_a}) \right] X_k^* \equiv B_m X_k^*. \end{aligned}$$

Since the AIC_c in [11] does not provide enough penalty in the panel data models, we modify the AIC_c and denote it as AIC_c^* ,

$$AIC_c^*(m) = \log(\hat{\sigma}_{k,m}^2) + \frac{1 + \text{trace}(B_m) \log(\frac{nT}{n+T}) / \frac{nT}{n+T}}{1 - (\text{trace}(B_m) + 2) \log(\frac{nT}{n+T}) / \frac{nT}{n+T}}, \quad (34)$$

where $\log(\hat{\sigma}_{k,m}^2) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\hat{v}_{m,i} - c_m f_{m,it})^2$. Then $\hat{M} = \arg \min_{m=1, \dots, \bar{M}} AIC_c^*(m)$ yields the stopping rule for the iterations.

For L_2 Boosting to be consistent for panel data models, we impose the following regularity conditions, under which FE-2SLS-Boosting is consistent for the conditional mean of x_{it}^* in quadratic mean under a panel data model. We make the following assumptions extending [11]:

Assumption B1. The dimension of instruments satisfies $\ell_{nT} = O(\exp(CnT^{1-\eta}))$, $n, T \rightarrow \infty$, for some $0 < \eta < 1$, $0 < C < \infty$.

Assumption B2. $\sum_{j=1}^{\ell_{nT}} |\Pi_j| < \infty$, as $n, T \rightarrow \infty$.

Assumption B3. $\sup_{1 \leq j \leq \ell_{nT}} \|Z_j^*\|_\infty < \infty$, where $\|Z_j^*\|_\infty = \sup_{\omega \in \Omega} |Z_j^*(\omega)|$ and Ω denotes the underlying probability space.

Assumption B4. $E|v_{it}|^s < \infty$ for some $s > 4/\eta$ with η in Assumption B1.

In Assumption B1, the dimension of instruments is allowed to grow exponentially with respect to the number of observations. So instruments can be in a high dimension. Assumption B2 gives an $L1$ -norm sparseness condition that the sum of the coefficient Π_j for all j is bounded. In this case, all instruments may be relevant, but the contribution of many instruments is very small. Hence weakly relevant instruments are allowed in the model. Assumption B3 states that by restricting the growth rate of ℓ_{nT} , the maximum realization of random variable Z_j^* under sample space needs to be bounded. Assumption B4 specifies the existence of some higher moments of the error term v_{it} , and the number of existing moments depends on η from assumption B1. Thus the number of existing moments and the growth rate of ℓ_{nT} are related.

Similarly to Theorem 1 in [11], under Assumptions B1-B4, the L_2 Boosting estimation converges to the conditional mean of x_{it}^* in quadratic mean. Thus, the FE-2SLS-Boosting is able to shrink some elements of the coefficient matrix $\hat{\Pi}$ to zero corresponding to weak instruments, and the subsequent application of FE-2SLS will consider those removed instruments Z^* as irrelevant to the endogenous variables X^* . Once the instruments are selected, we use the selected instruments to compute the FE-2SLS-Boosting estimator, and then the combined estimator of FE and FE-2SLS-Boosting.

Remark 2: Suppose some elements in Π are zeros, then only a subset of z_{it} is relevant to x_{it} . However, L_2 Boosting does not require some elements in Π to be zeros. What is required is $\sum_{j=1}^{\ell_{nT}} |\Pi_j| < \infty$, as $n, T \rightarrow \infty$. Assumption B2 only requires a much weaker form of sparsity, the sparseness condition that the *sum* of the coefficient $|\Pi_j|$ for all j is bounded. Hence, only a finite number of instruments are *strongly* relevant even if all elements of Π_j may not be zero.

5. Monte Carlo

In this section, we carry out a series of simulations to show that FE-2SLS can go bad if too many instruments are used, and then the regularization method FE-2SLS-Boosting restores the consistency of the FE-2SLS estimator and the efficiency of the combined estimator. We also evaluate the finite sample performance of the regularized combined estimator and compare the risk of other estimators. A design similar to [17] is used. We consider the following data generating process (DGP),

$$y_{it} = x_{it}\beta + \alpha_i + u_{it} \quad (35a)$$

$$x_{it} = \theta v_{it-1} + v_{it} \quad (35b)$$

$$\begin{pmatrix} u_{it} \\ v_{it} \end{pmatrix} \sim N \left(0, \begin{pmatrix} 1 & \rho/\sqrt{q} & \cdots & \rho/\sqrt{q} \\ \rho/\sqrt{q} & 1 & 0 & 0 \\ \vdots & 0 & 1 & 0 \\ \rho/\sqrt{q} & 0 & 0 & 1 \end{pmatrix} \right). \quad (35c)$$

Recall that x_{it} is a $q \times 1$ vector. For simplicity, we consider the case when all

elements of the $q \times 1$ moving average (MA) parameter vector θ are the same and set at the value 0.3. ρ is then simplified as a scalar. All elements of v_{it} are i.i.d. $N(0, I_q)$. Each pair of elements of errors u_{it} and v_{it} has covariance $\frac{\rho}{\sqrt{q}}$, but all other correlation zero. α_i are i.i.d. $N(0, 1)$. The parameter ρ can vary in $(-1, 1)$ to control the extent of endogeneity of x_{it} . The results are not quantitatively sensitive to the value of β , thus we set β to be zero without loss of generality.

Our goal is the consistent and efficient estimation of the structural parameter β . In the DGP, the variable x_{it} is endogenous following an invertible vector moving average VMA(1) process in (35b), which we approximate by the VAR(p) model of order $p \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. We consider the lagged variables of x_{it} as instruments, i.e., $z_{it} \equiv (x_{i,t-1} \dots x_{i,t-p})$. The parameter θ controls the strength of the instruments z_{it} as they are taken from the lagged x_{it} . The number of instruments equals to $\ell = q \times p$. To mimic the situation in the empirical application for estimation of the U.S. house prices using three endogenous variables in 49 states over 29 years, we consider $n = 49$, $T = 29$, $q = 3$. We set the range of ρ on a 20-point grid on $[0, 0.975]$. Estimates of three estimators, $\hat{\beta}_{\text{FE-2SLS}}$, $\hat{\beta}_{\text{FE}}$, and $\hat{\beta}_{\text{c}}$, are computed from 2000 Monte Carlo replications. To compare these three estimators, we calculate the median squared error (MSE) of each estimator

$$MSE(\hat{\beta}) = \text{median} \left(\left(\hat{\beta} - \beta \right)' \left(\hat{\beta} - \beta \right) \right). \quad (36)$$

We present the results graphically. In Figure 1: (a) plots $MSE(\hat{\beta})$ for $p = 1$, $\ell = 3$. This context is with just-identified instruments. Fig. 1(a) shows that the combined estimator has lower $MSE(\hat{\beta})$ than FE-2SLS, regardless of the degree of endogeneity. It also shows that the classical 2SLS estimator problem that its moments do not exist for the just-identified case, [cf. 24, 25]. Observe that $MSE(\hat{\beta})$ of FE-2SLS in Fig. 1(a) is quite large compared to that in Fig. 1(b) for all degrees of endogeneity ρ . It shows that the number of instruments in Fig. 1(a) is too small and it would be necessary to increase the number of instruments. If there are more instruments, then the system is over-identified with more moment restrictions than parameters to estimate. It will ensure FE-2SLS has a finite mean with one over-identifying restriction and a finite variance with two over-identifying restrictions [21]. Fig. 1(b) plots the MSE for $p = 2$, $\ell = 6$. Now the model is over-identified and FE-2SLS is well behaved. Fig. 1(b) shows that the combined estimator has similar $MSE(\hat{\beta})$ to FE for the small values of ρ where FE has small $MSE(\hat{\beta})$. The reduction in risk achieved by the combined estimator is large unless ρ is large. Fig. 1(c) plots the $MSE(\hat{\beta})$ for $p = 3$, $\ell = 9$. Fig. 1(d) plots the $MSE(\hat{\beta})$ for $p = 4$, $\ell = 12$. Fig. 1(e) plots the median squared error for $p = 5$, $\ell = 15$. The four figures in (b), (c), (d), (e) look similar. The combined estimator achieves some reduction in $MSE(\hat{\beta})$ relative to FE-2SLS for small values of ρ . However, when the number of instruments becomes larger in Fig. 1(f) with $p = 6$, $\ell = 18$, it starts to show that for large ρ , FE-2SLS becomes biased towards FE. It becomes more apparent as ℓ becomes even larger as shown in the subsequent plots. Continuing to Fig.(g)–Fig.(l), it is easy to see that the FE-2SLS is biased and the bias in FE-2SLS tends to get worse as more instruments are used.

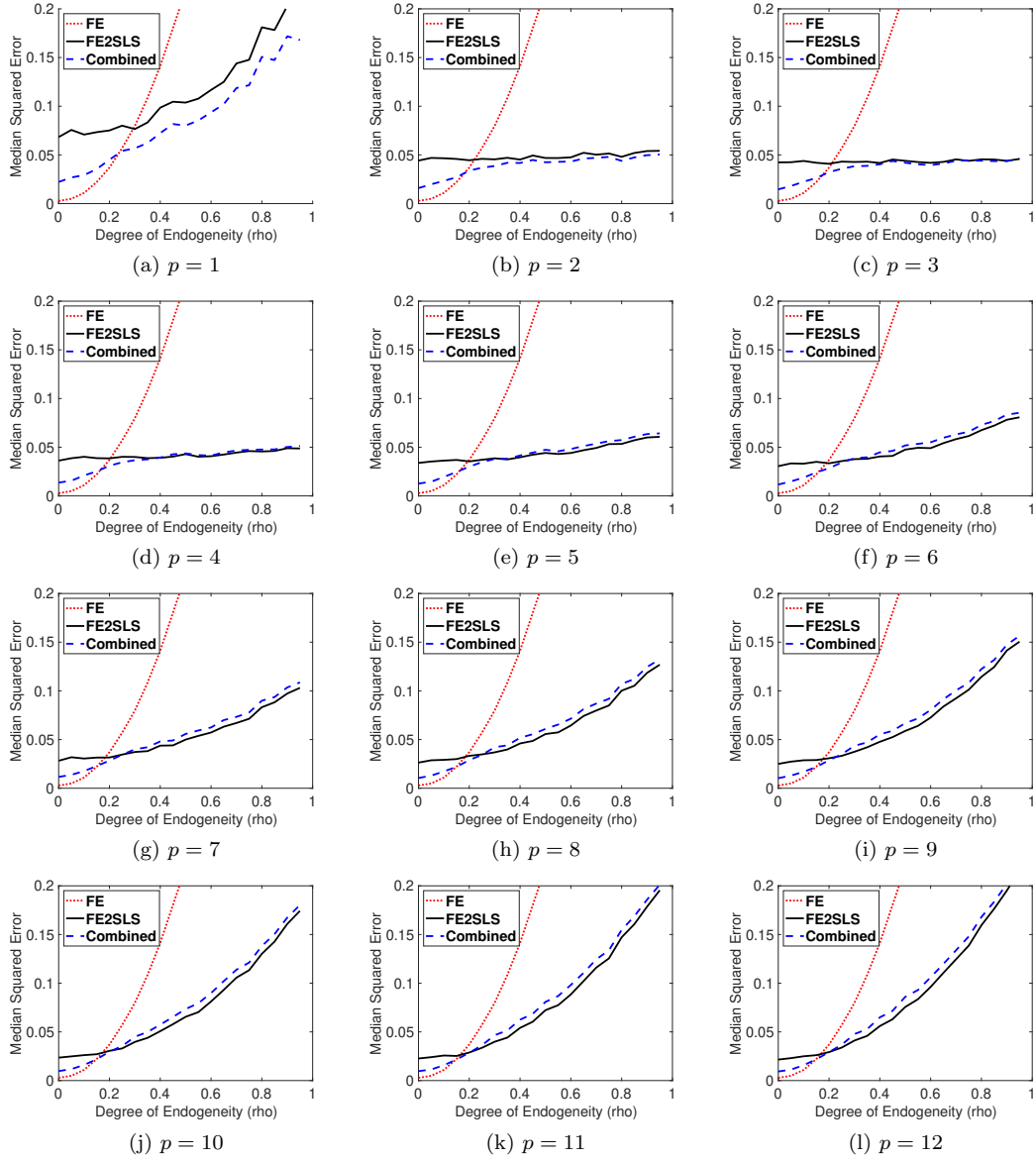


Figure 1. Median Squared Error of FE, FE-2SLS and Combined Estimators, $n = 49$, $T = 29$, $q = 3$.

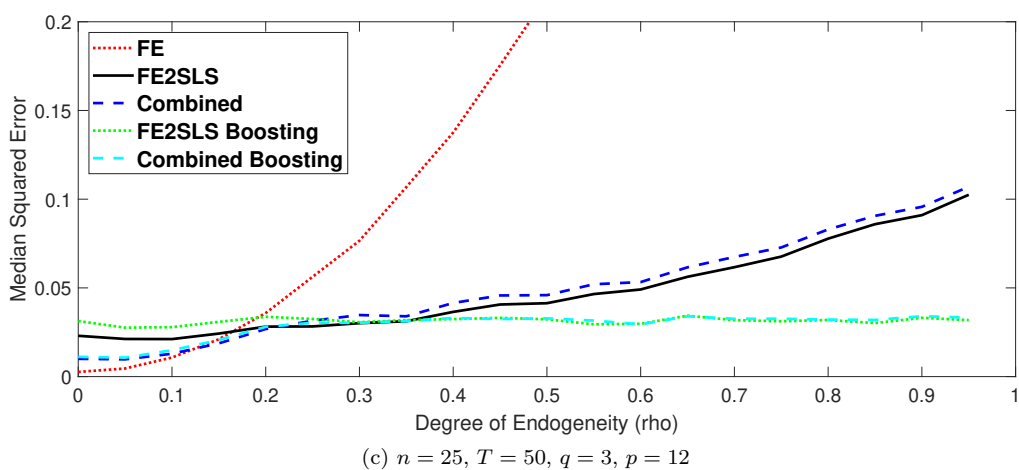
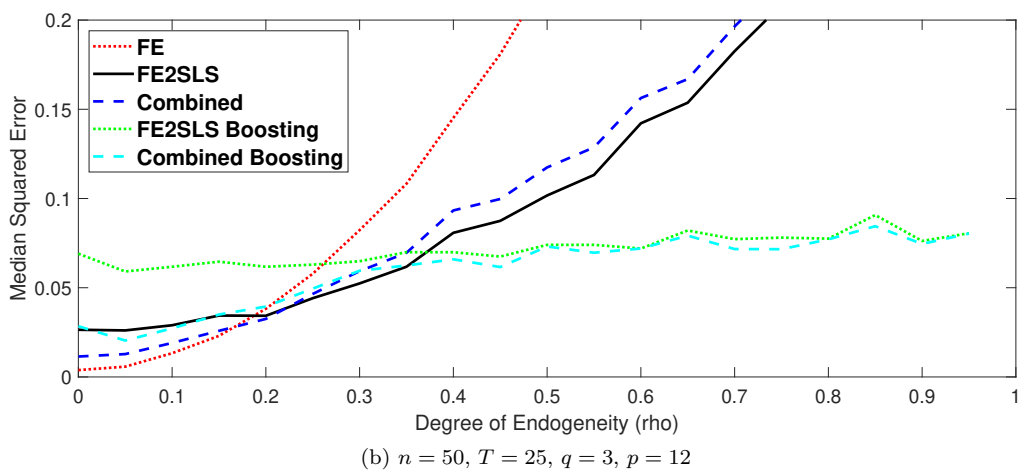
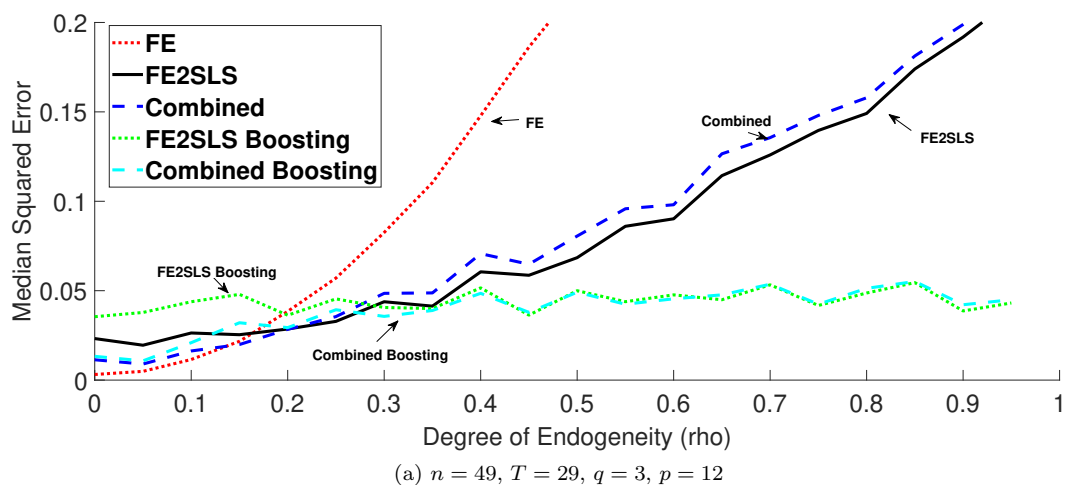


Figure 2. Median Squared Error of FE, FE-2SLS, Combined, F2SLS-Boosting, and Combined Boosting Estimators.

To fix this bias problem in FE-2SLS, we use the extended L_2 Boosting for panel data models to select the instruments, which makes the FE-2SLS estimator and the combined estimator more robust and restore their consistency when there are many potential instruments. To demonstrate this, we consider the setup in Fig. 1(l) with $p = 12$, $\ell = 36$. In Fig. 2(a), we zoom in Fig. 1(l) and add “FE-2SLS-Boosting” and “Combined-Boosting”, where FE-2SLS-Boosting is the Boosting-regularized FE-2SLS estimator and Combined-Boosting is the combined estimator of the FE estimator and the FE-2SLS-Boosting estimator. Notice that the vertical scale of Fig. 2(a) is between 0 and 0.2. By using the L_2 Boosting to regularize a panel data model with many instruments, the FE-2SLS-Boosting estimator restores its consistency. The FE-2SLS-Boosting is the post-selection FE-2SLS estimator. The combined-Boosting estimator is the combined estimator of FE and FE-2SLS-Boosting estimators. The MSEs of the FE-2SLS-Boosting estimator and the Combined-Boosting estimator are significantly reduced when compared to the MSE of the FE-2SLS estimator. When the endogeneity is weak (small values of ρ), the Combined-Boosting estimator (cyan-colored long-dashed) dominates the FE-2SLS-Boosting (green-colored dotted). As the endogeneity gets stronger, the combining weight in the Combined-Boosting goes toward the FE-2SLS-Boosting. The MSE of FE-2SLS-Boosting estimator is significantly reduced when compared to the MSE of the FE-2SLS estimator without instrument selection. The Combined-Boosting estimator dominates the FE-2SLS-Boosting estimator when the endogeneity is weak. However, the weight in the Combined-Boosting increases toward the FE-2SLS-Boosting as the endogeneity gets stronger.

In addition, we examine the effect of the sample size on the risk behavior in finite samples with different n and T . We fix the maximum lag at 12, and thus the effective time length is $T - 12$. When we have $n = 49$ and $T = 29$ (as in Fig. 2(a)), $n(T - 12) = 833$. For Fig. 2(b), $n(T - 12) = 650$. For Fig. 2(c), $n(T - 12) = 950$. As $\delta = \rho\sqrt{nT}$ from Equation (16) increases with \sqrt{nT} , the endogeneity is weaker in Fig. 2(b) than in Fig. 2(c). Therefore the conclusion in Fig. 2(a) is more clearly pronounced in Fig. 2(b) than in Fig. 2(c).

Remark 3: In the presence of many instruments, the combined estimator may behave poorly, and Theorem 2 may not hold for moderate to large values of ρ . Theorem 2 says the combined estimator is always better than the FE-2SLS estimator in the asymptotic risk. Figures 1(g)-(l) and Fig. 2 show, however, that Theorem 2 does not hold when p is large. This is because Theorem 2 holds only when FE-2SLS is consistent. When **both p and ℓ are large**, FE-2SLS is inconsistent and therefore Theorem 2 does not hold when there are many instruments. After selecting instruments through L_2 Boosting, we reduce the dimension of instruments to estimate the FE-2SLS estimator. This restores Theorem 2 and ensures the consistency of the FE-2SLS-Boosting estimator and the efficiency of the Combined estimator.

6. Estimation of House Prices Panel Data Model in United States

The U.S. housing price index, published by the Federal Housing Finance Agency (FHFA), is a measure designed to capture changes in the value of houses in the U.S. based on the data provided by Fannie Mae and Freddie Mac. HPY (2010) suggest a possible spatial pattern in U.S. housing prices using the common correlated effects estimator. [5] replicated the results of HPY by extending the period of research to 2011, and the housing price indexes at the metropolitan area level are used. The HPY results

are shown to be robust. As noted in [5], “The U.S. housing price index rose by nearly 46 percent from 2000 to 2006, followed by a sharp 28 percent drop, unprecedented in American history”, the extended period of 2004–2011 covers quite rare housing statistics of boom, bubble, crash, and recovery circle. What drives house prices? In this paper, the original HPY panel data of 49 States over the 29 year period of 1975–2003 is used to examine the extent to which housing price fluctuations in the U.S. are driven by the fundamental fluctuations.¹ We consider the following panel data model

$$p_{it} = \beta_0 + \beta_y y_{it} + \beta_g g_{it} + \beta_c c_{it} + \alpha_i + u_{it}, \quad (37)$$

where $i = 1, \dots, 49, t = 1, \dots, 29$, p_{it} is the logarithm of the real house price for the i th State of year t . y_{it} is the logarithm of the real disposable income per capita. $c_{it} = r_{it} - \Delta p_{it}$ is the net cost of borrowing. r_{it} is the long-term real interest rate. g_{it} is the population growth rate. α_i is the State-specific factor including the endowment of location, culture and etc. A more detailed description of the data can be found in HPY.

Let $x_{it} \equiv (y_{it} \ g_{it} \ c_{it})'$. We consider the lagged variables of the $q \times 1$ vector x_{it} as instruments, i.e., $z_{it} \equiv (x_{i,t-1} \dots x_{i,t-p})$. Let us assume that $q < \infty$ is fixed while p is allowed to grow with nT . Write a VAR(p) model for x_{it} as

$$\begin{aligned} x_{it} &= \Pi' z_{it} + v_{it} \\ &= \Pi_1 x_{i,t-1} + \dots + \Pi_p x_{i,t-p} + v_{it}, \end{aligned} \quad (38)$$

where $\Pi' = (\Pi_1 \dots \Pi_p)$ and $z_{it} = (x_{i,t-1} \dots x_{i,t-p})'$. Note that the lag order p in VAR(p) may increase with T so that the number of instruments is $\ell_{nT} = q \times p$ may increase with nT . Note that, even with $p \rightarrow \infty$, Assumption B2 is satisfied because $\sum_{k=1}^q \sum_{s=1}^p |\Pi_{sk}| < \infty$ under stationarity. In this example, $x_{it} = (y_{it} \ g_{it} \ c_{it})'$, we consider the lag order $p = 12$, which gives a large number of instruments with $\ell = q \times p = 36$.

Table 1 shows that the income elasticity of real house prices for the combined estimator is significant and positive but widely changing in the range from 0.8038 to 0.2525. We find a positive population growth and a negative net cost of borrowing in the real house prices, which are consistent with our expectation. Both are insignificant in Table 1(c) and 1(d). Similar to the HPY results, we find that the fluctuations in U.S. housing price can be attributed to the fluctuations in incomes, population, and interest rates, although the housing prices appears to have been affected by more fundamental factors. The estimators on the population growth and the net cost of borrowing also exhibit wild variation over different lag orders p and different number of instruments ℓ . We find that the combined estimators are almost the same as the FE-2SLS estimators for all cases. The Hausman statistics are very large, thus most weights are going toward the FE-2SLS estimators. However, when the number of lag order increases, the Hausman statistic decreases from 139.8735 (FE-2SLS when $p = 1$) to 16.4602 (FE-2SLS when $p = 12$). In addition, the estimators of FE-2SLS-Boosting and Combined-Boosting are different from the estimators of FE-2SLS under the same lag order. This is because Boosting selects 19 out of 36 instruments during the selection process. On the other hand, FE-2SLS uses all 36 instruments.

The standard errors of estimators are computed using bootstrap. It is important to note that the combined estimators in each panel in Table 1(c) and 1(d) have smaller

¹Data are available at <http://qed.econ.queensu.ca/jae/2014-v29.3/baltagi-li/>

Table 1. Economics of Real House Prices for 49 U.S. States, 1975–2003.

		p	$\hat{\beta}_y$	$\hat{\beta}_g$	$\hat{\beta}_c$
(a)	FE		0.7303 (0.0395)	1.5180 (0.4681)	-0.2485 (0.0790)
	FE-2SLS	1	0.2525 (0.0633)	3.3145 (0.8154)	-1.8542 (0.1578)
(b)	Combined	1	0.2559 (0.0631)	3.3016 (0.8101)	-1.8427 (0.1581)
	FE-2SLS	12	0.8038 (0.0474)	1.2853 (0.6093)	-0.1185 (0.1082)
(c)	Combined	12	0.7993 (0.0467)	1.2994 (0.5880)	-0.1264 (0.1051)
	FE-2SLS-Boosting	12	0.7643 (0.0493)	1.3192 (0.6410)	-0.2377 (0.1153)
(d)	Combined-Boosting	12	0.7615 (0.0476)	1.3354 (0.6125)	-0.2386 (0.1099)

Notes: Standard errors are in parentheses. “ p ” is the order of VAR(p) for the reduced form equation.

(a) reports FE,

(b) reports FE-2SLS and the combined estimator of FE and FE-2SLS when lag $p = 1$,

(c) reports FE-2SLS and the combined estimator of FE and FE-2SLS when lag $p = 12$,

(d) reports FE-2SLS-Boosting estimator which is the post-selection FE-2SLS estimator using the instruments selected by L_2 Boosting, and Combined-Boosting combines FE and FE-2SLS-Boosting when lag $p = 12$.

standard errors than the corresponding FE-2SLS estimators, which is in accordance with Theorem 2. For example in Table 1(d), the standard errors of the FE-2SLS-Boosting estimators of the three coefficient estimators are 0.0493, 0.6410, and 0.1153 respectively. After combined with FE, the standard errors of the Combined-Boosting estimators reduce to 0.0476, 0.6125, and 0.1099. The combined estimator yields smaller standard errors than FE-2SLS-Boosting.

7. Conclusion

The FE-2SLS estimator for panel data models is a widespread choice in empirical research. However, the FE-2SLS estimator is sensitive to the number of selected instruments and can be inconsistent when many instruments are used even when all the instruments are relevant and valid. In this paper, we propose a two-step procedure – using the L_2 Boosting to select instruments and combining FE with FE-2SLS-Boosting. It is demonstrated that L_2 Boosting for the selection of relevant instruments is important as it ensures the consistency of the FE-2SLS-Boosting estimator and the efficiency of the combined estimator. The proposed procedure provides improvement over both FE and FE-2SLS-Boosting estimators in terms of the asymptotic risk. Our empirical application shows that our results are similar to those in [20] that the economic fundamentals play an important role in affecting the U.S. real house prices at the state level.

Disclosure statement

The authors declare no competing interests. This article does not contain any studies with human participants performed by any of the authors.

References

- [1] S.C. Ahn and H.R. Moon, *Large- N and large- T properties of panel data estimators and the hausman test*, in *Festschrift in Honor of Peter Schmidt*, Springer, 2014, pp. 219–258.
- [2] T. Amemiya and T.E. MaCurdy, *Instrumental-variable estimation of an error-components model*, *Econometrica* 54 (1986), pp. 869–880.
- [3] J. Bai, *Panel data models with interactive fixed effects*, *Econometrica* 77 (2009), pp. 1229–1279.
- [4] B.H. Baltagi, *Econometric Analysis of Panel Data*, John Wiley & Sons, West Sussex, England, 2008.
- [5] B.H. Baltagi and J. Li, *Further evidence on the spatio-temporal model of house prices in the united states*, *Journal of Applied Econometrics* 29 (2014), pp. 515–522.
- [6] R.L. Basman, *A generalized classical method of linear estimation of coefficients in a structural equation*, *Econometrica* 25 (1957), pp. 77–83.
- [7] P.A. Bekker, *Alternative approximations to the distributions of instrumental variable estimators*, *Econometrica* 62 (1994), pp. 657–681.
- [8] A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen, *Sparse models and methods for optimal instruments with an application to eminent domain*, *Econometrica* 80 (2012), pp. 2369–2429.
- [9] A. Belloni and V. Chernozhukov, *Least squares after model selection in high-dimensional sparse models*, *Bernoulli* 19 (2013), pp. 521–547.
- [10] T.S. Breusch, G.E. Mizon, and P. Schmidt, *Efficient estimation using panel data*, *Econometrica* 57 (1989), pp. 695–700.
- [11] P. Bühlmann, *Boosting for high-dimensional linear models*, *The Annals of Statistics* 34 (2006), pp. 559–583.
- [12] M. Caner, *Lasso-type GMM estimator*, *Econometric Theory* 25 (2009), pp. 270–290.
- [13] X. Cheng and Z. Liao, *Select the valid and relevant moments: An information-based lasso for gmm with many moments*, *Journal of Econometrics* 186 (2015), pp. 443–464.
- [14] S.G. Donald, G.W. Imbens, and W.K. Newey, *Choosing instrumental variables in conditional moment restriction models*, *Journal of Econometrics* 152 (2009), pp. 28–36.
- [15] J. Fan and Y. Liao, *Endogeneity in high dimensions*, *Annals of Statistics* 42 (2014), pp. 872–917.
- [16] B.E. Hansen, *Efficient shrinkage in parametric models*, *Journal of Econometrics* 190 (2016), pp. 115–132.
- [17] B.E. Hansen, *Stein-like 2SLS estimator*, *Econometric Reviews* 36 (2017), pp. 840–852.
- [18] J.A. Hausman, *Specification tests in econometrics*, *Econometrica* 46 (1978), pp. 1251–1271.
- [19] J.A. Hausman and W.E. Taylor, *Panel data and unobservable individual effects*, *Econometrica* 49 (1981), pp. 1377–1398.
- [20] S. Holly, M.H. Pesaran, and T. Yamagata, *A spatio-temporal model of house prices in the USA*, *Journal of Econometrics* 158 (2010), pp. 160–173.
- [21] T.W. Kinal, *The existence of moments of k -class estimators*, *Econometrica* 48 (1980), pp. 241–249.
- [22] M. Mogstad, A. Torgovitsky, and C.R. Walters, *The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables*, National Bureau of Economic Research, 2019.
- [23] S. Ng and J. Bai, *Selecting instrumental variables in a data rich environment*, *Journal of*

- Time Series Econometrics 1 (2009), p. Article 4.
- [24] T. Sawa, *The exact sampling distribution of ordinary least squares and two-stage least squares estimators*, Journal of the American Statistical Association 64 (1969), pp. 923–937.
- [25] T. Sawa, *Finite-sample properties of the k-class estimators*, Econometrica 40 (1972), pp. 653–680.
- [26] C. Stein, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, ed., Vol. 1. University of California Press, 1956, pp. 197–206.
- [27] H. Theil, *Economic Forecasts and Policy*, North-Holland Publishing Company, Amsterdam, 1961.
- [28] A. Wald, *The fitting of straight lines if both variables are subject to error*, The Annals of Mathematical Statistics 11 (1940), pp. 284–300.

Appendix A. Proofs of Theorems

Proof of Theorem 1: The joint convergence (17) is a straightforward and standard calculation. Let $h_1 = G_1 h$ and $\xi_1 = G_1 \xi$ with $G_1 = (I \ 0)'$. Here we derive the asymptotic distribution of the FE estimator under the local-to-exogeneity setup. From Equation (4), $\hat{\beta}_{FE} = (X'QX)^{-1} X'Qy$.

$$\hat{\beta}_{FE} - \beta = (X'QX)^{-1} X'QD\alpha + (X'QX)^{-1} X'Qu = (X'QX)^{-1} X'Qu.$$

Given that $u = v\rho + \varepsilon$,

$$\hat{\beta}_{FE} - \beta = (X'QX)^{-1} (X'Qv\rho + X'Q\varepsilon).$$

Since $\rho = \frac{1}{\sqrt{nT}}\delta$, we have

$$\begin{aligned} \hat{\beta}_{FE} - \beta &= (X'QX)^{-1} \left(X'Qv \frac{\delta}{\sqrt{nT}} + X'Q\varepsilon \right) \\ \sqrt{nT} (\hat{\beta}_{FE} - \beta) &= \left(\frac{1}{nT} X'QX \right)^{-1} \left(\frac{1}{nT} X'Qv\delta \right) + \left(\frac{1}{nT} X'QX \right)^{-1} \left(\frac{1}{\sqrt{nT}} X'Q\varepsilon \right) \xrightarrow{d} h_1 + \xi_1, \end{aligned}$$

where

$$h_1 = \left(\text{plim} \frac{1}{nT} X'QX \right)^{-1} \left(\text{plim} \frac{1}{nT} X'Qv\delta \right).$$

Since $X = Z\Pi + v$, we have

$$h_1 = \sigma_u^{-2} V_1 \text{tr}(Q\Sigma) \delta,$$

and

$$\xi_1 \sim \left(\text{plim} \frac{1}{nT} X'QX \right)^{-1} Z,$$

with

$$Z = \frac{1}{\sqrt{nT}} X' Q \varepsilon \sim N \left(0, \sigma_u^2 \left(\text{plim} \frac{1}{nT} X' Q X \right) \right).$$

Hence,

$$\xi_1 \sim N \left(0, \sigma_u^2 \left(\text{plim} \frac{1}{nT} X' Q X \right)^{-1} \right) = N(0, V_1),$$

and

$$\sqrt{nT} (\hat{\beta}_{FE} - \beta) \xrightarrow{d} N(h_1, V_1),$$

where

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

The rest follows by the continuous mapping theorem, as in Theorem 1 of [17].

Proof of Theorem 2: The proof is based on the arguments in Theorem 2 of [17] closely. Noting that $\sqrt{nT} (\hat{\beta}_{FE-2SLS} - \beta) \xrightarrow{d} G'_2 \xi \sim N(0, V_2)$, then

$$R(\hat{\beta}_{FE-2SLS}) = E(\xi' G'_2 W G'_2 \xi) = \text{tr}(W V_2).$$

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

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

Then using the fact that the pointwise quadratic risk of Ψ is strictly smaller than that of Ψ^* , then we have

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

We can calculate that

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

If $Z \sim N(0, V)$ is $q \times 1$, K is $q \times q$, and $\eta(x)$ is absolutely continuous, then by Stein's Lemma

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

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

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

Therefore,

$$\begin{aligned} & E \left(\frac{(h + \xi)' G W G_2' \xi}{(h + \xi)' B (h + \xi)} \right) \\ &= E \text{tr} \left(\frac{G W G_2' V}{(h + \xi)' B (h + \xi)} - \frac{2 G W G_2' V}{((h + \xi)' B (h + \xi))^2} B (h + \xi) (h + \xi)' \right) \\ &= E \left(\frac{\text{tr} (G W G_2' V)}{(h + \xi)' B (h + \xi)} \right) - 2 E \text{tr} \left(\frac{G W G_2' V}{((h + \xi)' B (h + \xi))^2} B (h + \xi) (h + \xi)' \right). \end{aligned}$$

Since

$$\begin{aligned} G W G_2' V &= W G_2' V G = W (V_2 - V_1), \\ G W G_2' V B &= G W G_2' V G (V_2 - V_1)^{-1} G' = G W G', \end{aligned}$$

we have

$$E \text{tr} \left(\frac{G W G_2' V}{((h + \xi)' B (h + \xi))^2} B (h + \xi) (h + \xi)' \right) = E \text{tr} \left(\frac{(h + \xi)' G W G' (h + \xi)}{((h + \xi)' B (h + \xi))^2} \right).$$

Thus

$$\begin{aligned} E (\psi^{*'} W \psi^*) &= R (\hat{\beta}_{\text{FE-2SLS}}) + \tau^2 E \left(\frac{(h + \xi)' G W G' (h + \xi)}{((h + \xi)' B (h + \xi))^2} \right) \\ &\quad + 4\tau E \text{tr} \left(\frac{(h + \xi)' G W G' (h + \xi)}{((h + \xi)' B (h + \xi))^2} \right) - 2\tau E \text{tr} \left(\frac{W (V_2 - V_1)}{(h + \xi)' B (h + \xi)} \right). \end{aligned}$$

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 $G W G_2' V B = G W G' = B_1' A^* B_1$, $B_1' B_1 = B$.

Using the inequality $b' a b \leq (b' b) \lambda_{\max}(a)$ for symmetric a , and let

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

Then

$$\begin{aligned} \text{tr} (B (h + \xi) (h + \xi)' G W G_2' V) &= (h + \xi)' B_1' A^* B_1 (h + \xi) \\ &\leq (h + \xi)' B (h + \xi) \lambda_1. \end{aligned} \tag{A1}$$

Using Equation (A1) and Jensen's inequality, we have

$$\begin{aligned}
E(\psi^{*'}W\psi^*) &\leq R(\hat{\beta}_{\text{FE-2SLS}}) + (\tau^2 + 4\tau) E\left(\frac{\lambda_1}{(h + \xi)'B(h + \xi)}\right) \\
&\quad - 2\tau E\text{tr}\left(\frac{(W(V_2 - V_1))}{(h + \xi)'B(h + \xi)}\right) \\
&= R(\hat{\beta}_{\text{FE-2SLS}}) - E\left(\frac{\tau(2(\text{tr}(W(V_2 - V_1)) - 2\lambda_1) - \tau\lambda_1)}{(h + \xi)'B(h + \xi)}\right) \\
&\leq R(\hat{\beta}_{\text{FE-2SLS}}) - \frac{\tau(2(\text{tr}(W(V_2 - V_1)) - 2\lambda_1) - \tau\lambda_1)}{E((h + \xi)'B(h + \xi))}. \tag{A2}
\end{aligned}$$

Since $\text{tr}(BV) = \text{tr}(G(V_2 - V_1)^{-1}G'V) = q$, we have

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

Substituted into (A2) we have

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

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

Proof of Corollary 1 and Corollary 2: The proofs of Corollary 1 and Corollary 2 are a straightforward and standard calculation.

Appendix B. When there are too many instruments

In this appendix we show the inconsistency of the FE and the FE-2SLS estimators in the panel data models when there are too many instruments.

First, consider the FE estimator under endogeneity. $\hat{\beta}_{\text{FE}}$ is obtained as

$$\hat{\beta}_{\text{FE}} = (X'QX)^{-1}X'Qy = \beta + (X'QX)^{-1}X'Qu,$$

Premultiplying X by Q ,

$$QX = QZ\Pi + Qv. \tag{B1}$$

Let $y^* = Qy$, $X^* = QX$, $Z^* = QZ$, $u^* = Qu$, and $v^* = Qv$, the $\hat{\beta}_{\text{FE}}$ can be rewritten as

$$\hat{\beta}_{\text{FE}} = \beta + (X^{*'}X^*)^{-1}X^{*'}u^*.$$

Define $\xi = [Qu, Qv]$, and the variance-covariance matrix of ξ is

$$\Omega = \begin{bmatrix} \Omega_u & \Omega_{uv} \\ \Omega_{vu} & \Omega_v \end{bmatrix}. \quad (\text{B2})$$

Then

$$\begin{aligned} E\left(\frac{1}{nT}X'Qu\right) &= \frac{1}{nT}E(\Pi'Z^*u^* + v^*u^*) \\ &= \frac{1}{nT}E(v^*u^*) = \Omega_{uv}, \end{aligned}$$

and

$$\begin{aligned} E\left[\frac{1}{nT}X'QX\right] &= \frac{1}{nT}E[(Z\Pi + v)'Q(Z\Pi + v)] \\ &= \frac{1}{nT}E[\Pi'Z^*Z^*\Pi + \Pi'Z^*v^* + v^*Z^*\Pi + v^*v^*] \\ &= \frac{1}{nT}\Pi'E(Z^*Z^*)\Pi + \Omega_v. \end{aligned}$$

If X is endogenous, $\Omega_{uv} \neq 0$. Then the OLS estimator $\hat{\beta}_{\text{FE}}$ is inconsistent because

$$\hat{\beta}_{\text{FE}} - \beta \xrightarrow{p} \left[\frac{1}{nT}\Pi'E(Z^*Z^*)\Pi + \Omega_v\right]^{-1} \Omega_{uv} \neq 0. \quad (\text{B3})$$

Next, we show that the FE-2SLS estimator is inconsistent when there are too many instruments. The $\hat{\beta}_{\text{FE-2SLS}}$ is

$$\hat{\beta}_{\text{FE-2SLS}} = (X'H_ZX)^{-1}X'H_Zy = \beta + (X^*H_ZX^*)^{-1}X^*H_Zu^*, \quad (\text{B4})$$

where $H_Z = QZ(Z'QZ)^{-1}Z'Q = Z^*(Z^*Z^*)^{-1}Z^*$. Let the idempotent matrix H_Z decompose into $H_Z = \Psi\Lambda\Psi$, where $\Lambda = \text{diag}(I_\ell, 0_{nT-\ell})$, and Ψ is a $nT \times nT$ matrix that has orthonormal properties. There exist an $nT \times (q+1)$ matrix $R = \Psi'\xi\Omega^{-1/2}$ that satisfies

$$E(R'R) = E(\Omega^{-1/2}\xi'\Psi\Psi'\xi\Omega^{-1/2}) = E(\Omega^{-1/2}\xi'\xi\Omega^{-1/2}) = I_{k+1}. \quad (\text{B5})$$

We partition $R = [r_1, r_2]'$, where r_1 is $\ell \times q + 1$. Then,

$$\begin{aligned}
E\left(\frac{1}{nT}\xi' H_Z \xi\right) &= \frac{1}{nT} E(\xi' \Psi \Lambda \Psi \xi) \tag{B6} \\
&= \frac{1}{nT} \Omega^{1/2} E\left(\Omega^{-1/2} \xi' \Psi \Lambda \Psi \xi \Omega^{-1/2}\right) \Omega^{1/2} \\
&= \frac{1}{nT} \Omega^{1/2} E(R' \Lambda R) \Omega^{1/2} \\
&= \frac{1}{nT} \Omega^{1/2} E\left([r_1', r_2'] \begin{bmatrix} I_\ell & 0 \\ 0 & 0_{nT-\ell} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}\right) \Omega^{1/2} \\
&= \frac{1}{nT} \Omega^{1/2} E(r_1' r_1) \Omega^{1/2} \\
&= \frac{\ell}{nT} \Omega,
\end{aligned}$$

which implies that $E\left(\frac{1}{nT} v' H_Z v\right) = \frac{\ell}{nT} \Omega_v$ and $E\left(\frac{1}{nT} v' H_Z u\right) = \frac{\ell}{nT} \Omega_{vu}$. So,

$$\begin{aligned}
E\left(\frac{1}{nT} X^{*'} H_Z X^*\right) &= \frac{1}{nT} E\left((Z^* \Pi + v^*)' H_Z (Z^* \Pi + v^*)\right) \tag{B7} \\
&= \frac{1}{nT} E\left(\Pi' Z^{*'} H_Z Z^* \Pi + \Pi' Z^{*'} H_Z v^* + v^{*'} H_Z Z^* \Pi + v^{*'} H_Z v^*\right) \\
&= \frac{1}{nT} E\left(\Pi' Z^{*'} Z^* \Pi\right) + \frac{1}{nT} E\left(v^{*'} H_Z v^*\right) \\
&= \frac{1}{nT} E\left(\Pi' Z^{*'} Z^* \Pi\right) + \frac{\ell}{nT} \Omega_v,
\end{aligned}$$

and

$$\begin{aligned}
E\left(\frac{1}{nT} X^{*'} H_Z u^*\right) &= \frac{1}{nT} E\left((Z^* \Pi + v^*)' H_Z u^*\right) \tag{B8} \\
&= \frac{1}{nT} E\left(v^{*'} H_Z u\right) = \frac{\ell}{nT} \Omega_{vu}.
\end{aligned}$$

Hence,

$$\hat{\beta}_{\text{FE-2SLS}} - \beta \xrightarrow{p} \frac{\ell}{nT} \left[\frac{1}{nT} \Pi' E(Z^{*'} Z^*) \Pi + \frac{\ell}{nT} \Omega_v \right]^{-1} \Omega_{uv} \neq 0, \tag{B9}$$

and $\hat{\beta}_{\text{FE-2SLS}}$ is inconsistent unless $\frac{\ell}{nT} \rightarrow 0$.