Fixed-Effects IV Panel Data Models with Time-Invariant Regressors: Recovering their Partial Effects and Performing Inference

Scott E. Atkinson
Department of Economics
University of Georgia
Athens, GA 30602

Christopher Cornwell
Department of Economics
University of Georgia
Athens, GA 30602

September 18, 2015

DRAFT: Do not cite or quote without the permission of the authors.
Abstract

With panel data, the typically-used fixed-effects estimator controls for unobserved heterogeneity while eliminating time-invariant regressors. Although often treated as lost, we recover their partial effects in a second-step IV regression, where some time-invariant variables are correlated with the unobserved effect. We derive the adjusted asymptotic covariance matrix for the second-step estimator and employ this matrix to develop a pairs and two wild bootstrap alternatives that produce asymptotic refinements. Using Monte Carlo experiments, the wild estimators produce much smaller error in rejection probability and somewhat better power, especially with weak instruments and strong endogeneity, than pairs and asymptotic methods.

JEL CODES: C150, C330, C360

KEY WORDS: Multiple Equation Panel Data, Bootstrap Methods, Endogeneity, Simultaneous Equations, Instrumental Variables, Weak Instruments.
1. Introduction

In panel regression models, frequently interest centers on the partial effects of time-invariant variables which are correlated with the unobserved effects. A classic example is found in panel wage regressions, where the estimation of the returns to education is typically carried out on individuals who have completed their schooling, which is generally correlated with unobserved ability.\textsuperscript{1} Another example is the estimation of production functions for electric utilities, where fixed factors, such as public or private ownership or the presence of rate-of-return regulation, are typically correlated with the unobserved characteristics like the utility’s political goodwill or management ability.\textsuperscript{2} The issue may be even more common in panels of state and country-level aggregates, where policy-relevant fixed institutional variables are likely to be correlated with unmeasured effects, such as voter attitudes toward government, the strength of religious beliefs, or cultural norms.\textsuperscript{3}

The primary advantage of panel data is that the econometrician can easily control for fixed, unobserved effects that may be correlated with explanatory variables using a fixed-effects (FE) or a random effects (RE) method. One can consistently and directly estimate the effects of the time-invariant variables by adopting a RE approach. However, in doing so one loses the robustness of the FE estimator for the coefficients of the time-varying variables. Thus, the FE estimator is by far the most popular technique for controlling for fixed, unobserved effects, because it does not require the RE assumption of the uncorrelatedness of the time-varying explanatory variables and the effects. However,

\textsuperscript{1} See for example Cornwell and Rupert (1988), Heineck and Anger (2010), and García-Mainar and Montuenga-Gómez (2005).

\textsuperscript{2} See Atkinson and Primont (2005) for examples of numerous studies which could be easily formulated in this manner.

\textsuperscript{3} See the highly cited paper by the political scientists Plümper and Troeger (2007), which unfortunately ignores endogeneity and provides an incorrect solution to estimation of this type of model even without endogeneity.
the time-demeaning data transformation that defines the FE estimator sweeps away any
time-invariant variables from the model. A review of the empirical panel-data literature
finds that many studies have justified a RE approach on the belief that the partial effects
of the time-invariant variables were permanently lost in a FE context, due to this data
transformation.\textsuperscript{4}

This popular perception is incorrect. We employ an instrumental variable (IV) FE
estimator in a second step to recover the partial effects of the time-invariant variables
and their adjusted standard errors, based on our derivation of the adjusted asymptotic
covariance matrix. This IV procedure encompasses the “simple, consistent” estimator of
Hausman and Taylor (hereafter HT) (1981). While they allowed some of the time-invariant
variables to be correlated with the effects, they did not derive the asymptotic covariance
matrix estimator for their second-step IV estimator.\textsuperscript{5} We also examine the extent to
which bootstrap estimators which use the asymptotic formula to perform inference in the
second step can generate asymptotic refinements, relative to a non-bootstrap approach.
These refinements are available because asymptotic formulas can be highly biased in small
samples, especially with strong endogeneity and weak instruments.

In a widely-cited paper, Murphy and Topel (1985) derive an adjusted asymptotic
covariance in a general two-step framework, but do not consider the IV or panel-data
setting in which our HT-type procedure is relevant. Wooldridge (2010, p. 359) covers HT-
type estimators, but only refers to the “standard arguments” for adjusting second-step

\textsuperscript{4} This is consistent with statements in most advanced econometrics texts. One important exception is
Wooldridge (2010).

\textsuperscript{5} Atkinson and Cornwell (2014) consider the simple case in which the time-invariant variables are un-
correlated with the unobserved effects. They also perform Monte Carlo experiments comparing the size
and power of estimated $t$-values for bootstrap and asymptotic estimators with those that naively ignored
the first-step estimation error.
estimator standard errors. Thus, the reader is given only general guidance about how to recover the second-step estimated coefficients and perform classical inference about them.\footnote{Plümper and Troeger (2007) derived a three-step FE method for recovering estimated coefficients and standard errors for time-invariant variables, assuming the time-invariant variables and effects are uncorrelated, complete with Stata software to implement it. However, as shown in the critiques of Breusch et al. (2011) and Greene (2011), their method is incorrect. What remains significant about their work is the interest it reflects among empirical researchers for guidance on the recovery of the partial effects of the time-invariant variables and their significance in the two-step estimation problem, based on the numerous citations and applications of their erroneous method.}

Although many Monte Carlo studies have examined the asymptotic refinements of bootstrap estimators, none have considered the two-step, panel-data context we treat here. A number have focused exclusively on the performance of the pairs estimator in cross-section settings with heteroskedastic errors and exogeneous regressors. Studies by Horowitz (2001), MacKinnon (2002), Flachaire (2005), Davidson and MacKinnon (2006), and Davidson and Flachaire (2008) typically find that the pairs dominates the non-wild residual counterparts.

Other studies have examined the importance of instrument strength for non-wild procedures in cross-section models with endogenous regressors. Based on Moreira, Porter, and Suarez (2004), who build on Rothenberg (1984), we can expect that the error rejection probability (ERP) (defined as the difference between the nominal and actual probability of rejection under the hull hypothesis) will increase as instruments become weaker. They point out that due to the asymptotic equivalence up to higher-order terms of the empirical Edgeworth expansion and the bootstrap, the latter provides asymptotic refinements with strong instruments. The Monte Carlo experiments of Moreira, Porter, and Suarez (2004) find that the pairs bootstrap estimator performs poorly with very weak instruments, but somewhat better with moderately weak or strong instruments. Davidson and MacKinnon (2008) examine weak/strong endogeneity crossed with weak/strong instruments for a cross-section model with one endogenous regressor and compare the pairs method and a number
of residual (non-wild) bootstrap methods in terms of size and power of estimated $t$-values. An estimator is termed “restricted” if it imposes the null hypothesis when generating the bootstrap data and an estimator is termed “efficient” if it employs an efficient estimator of the reduced-form equation. They examine all combinations of restricted/unrestricted and efficient/inefficient residual estimators as well as a pairs estimator. The restricted estimators typically exhibit smaller ERP with weak instruments and strong endogeneity.

More comprehensive Monte Carlo experiments are carried out for cross-section models by Davidson and MacKinnon (2010), who compare the ERP of estimated $t$-values based on the wild restricted-efficient (WRE) estimator, a number of residual methods (the restricted-inefficient residual, the unrestricted-inefficient residual, and the restricted-efficient residual), the pairs, and the asymptotic formula itself. They find that the ERP for the non-WRE methods are often substantially larger with weak instruments and strong endogeneity than those for the WRE. However, they do not consider the simpler wild unrestricted-inefficient (WU) estimator, which is the more commonly-estimated wild method, and do not consider two-step IV panel data models.

Therefore, a number of important bootstrap issues remain to be addressed regarding two-step estimation procedures of the sort we consider here. First, there is no Monte Carlo evidence about the extent to which the bootstrap estimators using the second-step asymptotic formula offer asymptotic refinements relative to using the asymptotic formula to construct $t$-values in the classical manner. Hence, we carry out extensive Monte Carlo comparisons of ERP and power with weak/strong instruments and weak/strong endogeneity using the asymptotic formula and three bootstrap methods which yield asymptotic refinements: and pairs, WRE, and WU. The question of their relative performance is especially important with weak instruments and strong endogeneity, since IV estimators
perform poorly in their presence. Second, although the consistency of the bootstrap estimator is well known, little work has been done regarding bootstrap bias. We prove the unbiasedness of the wild estimator and the bias of the pairs estimator in the second step of our model. These results offer insights into the relative performance of the bootstrap methods with a small number of cross-sectional units, since bias affects ERP, based on the work of Davidson and MacKinnon (1999).

The remainder of this paper is organized as follows. We derive the two-step estimator in section 2 and the corrected asymptotic covariance formula for the second-step estimated standard errors in section 3, along with the pairs and wild bootstrap alternatives which use this covariance formula. Section 4 examines the bias of bootstrap estimators (with theorems on their bias in the Appendix, section 7). In section 5, we describe the Monte Carlo simulations and summarize results on ERP and power of t-tests. Conclusions follow in section 6.

2. Linear Panel Models and Two-Step Estimation

We consider the estimation of linear panel-data models of the form

$$y_{it} = x_{it}\beta + z_i\gamma + \xi_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T,$$

(2.1)

where $\xi_{it} = c_i + e_{it}$, $y_{it}$ is the dependent variable, $x_{it}$ is a $(1 \times K)$ vector of time-varying regressors, $z_i$ is a $(1 \times G)$ vector of time-invariant regressors, $c_i$ is an unobserved effect that is fixed for the cross-section unit, and $e_{it}$ is an error term. The coefficient vectors, $\beta$ and $\gamma$, are $(K \times 1)$ and $(G \times 1)$, respectively. For most of the discussion that follows we
work with the form of the model that combines all $T$ observations for each cross-section unit:

$$y_i = X_i\beta + (j_T \otimes z_i)\gamma + j_T c_i + e_i, \quad (2.2)$$

where $y_i$, $e_i$ and the columns of $X_i$ are $T \times 1$ vectors, and $j_T$ is a $T$-vector of ones. Two-step estimation is motivated by an interest in the effects of time-invariant variables, which we compute, allowing for correlation between some of the variables in $z_i$ and the unobserved effect. Consistent with the an HT-type model, we make the following assumptions:

$$E(e_{it} \mid X_i, z_i, c_i) = 0 \quad \text{and} \quad (2.3)$$

$$E(c_i \mid z_{i1}) = 0, \quad (2.4)$$

$$E(c_i \mid z_{i2}) \neq 0, \quad (2.5)$$

where $z_i = (z_{i1}, z_{i2})$, whose components have column dimensions $G_1$ and $G_2$, respectively, and the first element of $z_{i1}$ is 1. Thus, we assume that $z_{i1}$ is strictly exogenous with respect $c_i$ and that the only source of endogeneity is the correlation between the time-invariant regressors in $z_{i2}$ and $c_i$. We make no assumption about the relationship between the variables in $X_i$ and the unobserved effects.

Our estimation of $\gamma$ begins with the first-step FE estimator of $\beta$:

$$\hat{\beta}_{FE} = \left(\sum_i X_i'Q_iX_i\right)^{-1}\sum_i X_i'Q_iy_i, \quad (2.6)$$

where $Q_i = I_T - j_T(j_T'j_T)^{-1}j_T'$ is the projection that time de-means the data. As is well-known, $\hat{\beta}_{FE}$ is consistent even if $X_i$ is correlated with $c_i$, because time-demeaning removes the unobserved effects. We exploit this robustness of the FE estimator in the first step and invoke (2.4)-(2.5) in the second step.
The second step uses $\hat{\beta}_{FE}$ to compute individual or group-level residuals,

$$\hat{\delta}_i = \bar{y}_i - \bar{x}_i \hat{\beta}_{FE}, \quad (2.7)$$

and formulate the regression model

$$\hat{\delta}_i = z_i \gamma + u_i, \quad (2.8)$$

where

$$u_i = \bar{\xi}_i - \bar{x}_i (\hat{\beta}_{FE} - \beta), \quad (2.9)$$

$\bar{\xi}_i = c_i + \bar{e}_i$ and the over-bar indicates the sample-period mean for unit $i$ (e.g., $\bar{x}_i = \frac{1}{T} \sum_t x_{it}$). Equation (2.9) follows from combining (2.7), (2.8), and the sample-period mean for unit $i$ of (2.2).

Given (2.4) and (2.5), we can rewrite (2.8) as

$$\hat{\delta}_i = z_{i1} \gamma_1 + z_{i2} \gamma_2 + u_i, \quad (2.10)$$

and specify a set of reduced-form equations corresponding to the variables in $z_i$ that are correlated with $c_i$:

$$z_{i2,g} = w_i \pi + v_{i2,g}, \quad g = 1, \ldots, G_2, \quad (2.11)$$

where $z_{i2,g}$ is element $g$ of $z_{i2}$, $\pi$ is a ($J \times 1$) coefficient vector, $w_i$ is a $1 \times J$ vector of instruments (whose first element is 1) satisfying

$$E(w_i | u_i) = 0, \quad (2.12)$$

and $v_{i2,g}$ is an error that is uncorrelated with $w_i$, $\forall g$ and $u_i$ The panel-data literature offers several possibilities for instrument sets. Following Amemiya and MaCurdy (1986)
and Breusch, Mizon, and Schmidt (1989), each time realization of the time-varying variables or their demeaned counterparts (depending on the precise nature of (2.12)) could qualify as a separate instrument. The potential weakness of such instrumental variables is an obvious concern and warrants investigation using Monte Carlo analysis.

We then estimate $\gamma$ by applying IV to (2.10) utilizing the instruments implied by (2.12). The estimator, which we label $\hat{\gamma}_{FE}$ because it is derived from the FE estimator of $\beta$, is

$$
\hat{\gamma}_{FE} = \left[ \left( \sum_i z'_i w_i \right) \left( \sum_i w'_i w_i \right)^{-1} \left( \sum_i w'_i z_i \right) \right]^{-1} \left( \sum_i w'_i w_i \right)^{-1} \sum_i w'_i \hat{\delta}_i. \tag{2.13}
$$

Equation (2.13) encompasses HT’s “consistent but inefficient” estimator of the coefficients of time-invariant variables in linear panel models (HT, p. 1383-4).


3.1. Asymptotic Covariance Matrix

While HT propose an estimator like (2.13), they do not derive its asymptotic covariance matrix. However, as noted in Wooldridge (2010), the asymptotic covariance matrix for $\hat{\gamma}_{FE}$ can be obtained by applying standard arguments for two-step estimators. We begin by writing the sampling error of $\hat{\gamma}_{FE}$ as

$$
\hat{\gamma}_{FE} - \gamma = \left[ \left( \sum_i z'_i w_i \right) \left( \sum_i w'_i w_i \right)^{-1} \left( \sum_i w'_i z_i \right) \right]^{-1} \left( \sum_i w'_i w_i \right)^{-1} \sum_i w'_i u_i. \tag{3.1}
$$
Then we can show that \( \sqrt{N}(\hat{\gamma}_{FE} - \gamma) \) is asymptotically normal with a limiting covariance matrix that can be expressed as

\[
(\mathbf{B}_{zw} \mathbf{B}_{ww}^{-1} \mathbf{B}_{wz})^{-1} \mathbf{B}_{zw} \mathbf{B}_{ww}^{-1} \mathbf{A} \mathbf{B}_{ww}^{-1} \mathbf{B}_{wz} (\mathbf{B}_{zw} \mathbf{B}_{ww}^{-1} \mathbf{B}_{wz})^{-1},
\]

(3.2)

where, e.g., \( \mathbf{B}_{zw} = \text{plim} \frac{1}{N} \sum_{i} z'_i \mathbf{w}_i \). As implied by (2.9)

\[
\mathbf{A} = \text{plim} \frac{1}{N} \sum_{i} \xi_i^2 \mathbf{w}'_i \mathbf{w}_i + \text{plim} \frac{1}{N} \sum_{i} \mathbf{w}'_i \mathbf{x}_i \mathbf{V}_{\hat{\beta}_{FE}} \mathbf{x}'_i \mathbf{w}_i,
\]

(3.3)

where \( \mathbf{V}_{\hat{\beta}_{FE}} \) is the limiting covariance matrix of \( \sqrt{N}(\hat{\beta}_{FE} - \beta) \).

A consistent estimator of the asymptotic covariance matrix of \( \hat{\gamma}_{FE} \) hinges on the consistent estimation of \( \mathbf{A} \). The latter is accomplished by utilizing the robust covariance matrix estimator of \( \mathbf{V}_{\hat{\beta}_{FE}} \),

\[
\hat{\mathbf{V}}_{\hat{\beta}_{FE}} = \left( \sum_{i} \mathbf{X}'_i \mathbf{Q}_i \mathbf{X}_i \right)^{-1} \sum_{i} \mathbf{X}'_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}'_i \mathbf{Q}_i \mathbf{X}_i \left( \sum_{i} \mathbf{X}'_i \mathbf{Q}_i \mathbf{X}_i \right)^{-1}
\]

(3.4)

(see Arellano (1987)), and extracting an estimator of \( \xi_i \) from the group-level version of (2.1) evaluated at \( (\hat{\beta}_{FE}, \hat{\gamma}_{FE}) \). The estimator of the asymptotic covariance matrix of \( \hat{\gamma}_{FE} \), which is also used to obtain asymptotic refinements in the bootstrap calculations of \( t \)-values as explained below, is obtained by computing (3.2) after replacing the probability limits with their sample analogs.

### 3.2. Panel Data Bootstrap Resampling Methods

With our panel-data bootstrap methods, three resampling schemes are available. These are cross-sectional (also called panel bootstrap) resampling, temporal (also called block bootstrap) resampling, and cross-sectional/temporal resampling. With panel-bootstrap resampling, one randomly selects cross-sectional units and uses all \( T \) observations for each. If cross-sectional dependence exists, one can select the relevant blocks
of cross-sectional units. With temporal resampling, one randomly selects temporal units and uses all $N$ observations for each. If temporal dependence exists, one can select the relevant blocks of temporal units. With cross-sectional/temporal resampling, both methods are utilized. Following Cameron and Trivedi (2005), consistent (as $N \to \infty$) standard errors can be obtained using the cross-sectional bootstrap method. Hence, we employ this method for both the pairs and wild methods, where we assume no cross-sectional or temporal dependence.\footnote{Also, see Kapetanios (2008) who shows that if the data do not exhibit cross-sectional dependence but exhibit temporal dependence, then cross-sectional resampling is superior to block bootstrap resampling. Further, he shows that cross-sectional resampling provides asymptotic refinements. Monte Carlo results using these assumptions indicate the superiority of the cross-sectional method.}

3.2.1 Pairs Bootstrap Estimator

A desirable property of the pairs bootstrap is that it yields a heteroskedasticity-consistent covariance matrix estimator (HCCME), as proven by Lancaster (2003). Davidson and MacKinnon stress the importance of estimating restricted bootstrap models, so that the null hypothesis is imposed on both the structural equation and the bootstrap disturbances. With the pairs bootstrap, however, only the unrestricted model can be estimated, since one cannot impose parametric restrictions.\footnote{Flachaire (1999) developed a hybrid pairs bootstrap estimator that does allow one to impose parametric restrictions. However, it has not been widely adopted and is not treated in our paper.}

The pairs bootstrap procedure can be extended to our problem as follows:

1. Compute $\hat{\beta}_{FE}$ in (2.6).
2. Using $\hat{\beta}_{FE}$, compute $\hat{\delta}_i$ in (2.7).
3. Compute $\hat{\gamma}_{FE}$ in (2.13).
4. Draw randomly with replacement among $i = 1, \ldots, N$ blocks, using all $T$ observations in the chosen block, with probability $1/N$ from $\{y_{it}, x_{it}, z_{it}, w_{it}\}$ to obtain
\{y_{it}^p, x_{it}^p, z_{it}^p, w_{it}^p\}, where the superscript denotes the pairs estimator. Resampling all variables in this manner preserves the correlation of the corresponding time-invariant and group-mean variables in the second-step regression with the first-step variables.

(5) For the pairs bootstrap, define \(\xi_i^p\) as a \((T \times 1)\) vector made up of \(\{\xi_{i1}^p, \ldots, \xi_{iT}^p\}\) for observation \(i\). Write the first-step regression model with unknown error term, \(\xi_i^p\), as

\[
y_i^p = X_i^p \hat{\beta}_{FE} + (j_T \otimes z_i^p) \hat{\gamma}_{FE} + \xi_i^p.
\]

(3.5)

Compute the FE estimator of \(\hat{\beta}\) using the pairs bootstrap data \((y_i^p, X_i^p)\)

\[
\hat{\beta}_{FE}^p = \left( \sum_i X_i^{p'} Q_i X_i^p \right)^{-1} \sum_i X_i^{p'} Q_i y_i^p.
\]

(3.6)

(6) Using \(\hat{\beta}_{FE}^p\) compute the residuals

\[
\delta_i^p = \bar{y}_i^p - \bar{x}_i^p \hat{\beta}_{FE}^p.
\]

(3.7)

(7) Formulate the second-step pairs bootstrap model as

\[
\delta_i^p = z_i^p \hat{\gamma}_{FE} + u_i^p,
\]

(3.8)

where \(u_i^p\) is the pairs second-step error,

Then compute the second-step IV estimator of \(\hat{\gamma}_{FE}\) using the bootstrap data:

\[
\hat{\gamma}_{FE}^p = \left[ \left( \sum_i z_i^{p'} w_i^p \right) \left( \sum_i w_i^{p'} w_i^p \right)^{-1} \left( \sum_i w_i^{p'} z_i^p \right) \right]^{-1}
\]

\[
\left( \sum_i z_i^{p'} w_i^p \right) \left( \sum_i w_i^{p'} w_i^p \right)^{-1} \sum_i w_i^{p'} \delta_i^p.
\]

(3.9)

(8) Iterate steps 4-7 for a given number of bootstrap trials.
3.2.2 Wild Bootstrap Estimators

Davidson and Flachaire (2008) have shown that the wild bootstrap also yields a HC-CME. We examine the WU and WRE bootstrap estimators. Davidson and MacKinnon (2010) argue that the WRE estimator should be computed when a single hypothesis is being tested, so that the null hypothesis is imposed on both the structural equation and the bootstrap disturbances. This means that the DGP is estimated more efficiently when true restrictions are imposed. Further, the WRE produces more efficient estimators of the reduced-form equations. Together these efficiency gains should produce lower ERP than the other bootstrap estimators. We compare this estimator with the more commonly-used and easier to compute, WU estimator by initially executing the following steps (1)-(7) for both:

1. Compute $\hat{\beta}_{FE}$ in (2.6).
2. Using $\hat{\beta}_{FE}$, compute $\hat{\delta}_i$ in (2.7).
3. Compute $\hat{\gamma}_{FE}$ in (2.13).
4. From (2.1) compute
   \[ \hat{\xi}_{it} = y_{it} - x_{it}\hat{\beta}_{FE} - z_i\hat{\gamma}_{FE}, \]
   where $h_{it}$ is the diagonal element of the projection matrix corresponding to the right-hand-side variables of (2.2).\(^9\)

Then define $f(\hat{\xi}_{it})$ as:

\[ f(\hat{\xi}_{it}) = \frac{\hat{\xi}_{it}}{(1 - h_{it})^{1/2}}, \]

Davidson and Flachaire (2008) show this projection-matrix adjustment to the wild bootstrap produces homoskedastic resampled residuals so long as the error terms are homoskedastic.

\(^9\) Davidson and Flachaire (2008) show this projection-matrix adjustment to the wild bootstrap produces homoskedastic resampled residuals so long as the error terms are homoskedastic.
(5) We follow Davidson and Flachaire (2008) and MacKinnon (2002) and define \( \epsilon_i \) as the two-point Rademacher distribution:

\[
\epsilon_i = \begin{cases} 
-1 & \text{with probability } \frac{1}{2} \\
1 & \text{with probability } \frac{1}{2}
\end{cases}
\]  

(3.11)

This assigns the same value to all \( T \) observations for each \( i \). Then, we generate

\[
y_{it}^w = x_{it}\hat{\beta}_{FE} + z_i\hat{\gamma}_{FE} + \xi_{it}^w,
\]

(3.12)

where

\[
\xi_{it}^w = f(\hat{\xi}_{it})\epsilon_i.
\]

(3.13)

Davidson and Flachaire (2008) show that this version of the wild bootstrap is superior to other wild methods because \( E(\epsilon_i) = 0, E(\epsilon_i^2) = 1, E(\epsilon_i^3) = 0 \), and \( E(\epsilon_i^4) = 1 \). Since \( \hat{\xi}_{it} \) and \( \epsilon_i \) are independent, \( E(\xi_{it}^w) = E(f(\hat{\xi}_{it})\epsilon_i) = 0 \), the variance of \( \xi_{it}^w \) is that of \( f(\hat{\xi}_{it}) \), its third moment is zero (which implies zero skewness in \( f(\hat{\xi}_{it}) \)), but its fourth moment is again that of \( f(\hat{\xi}_{it}) \). Thus, the first, second, and fourth moments of \( f(\hat{\xi}_{it}) \) are reproduced exactly in the wild bootstrap data using (3.11).

(6) Compute the FE estimator of \( \hat{\beta} \) using the wild bootstrap data:

\[
\hat{\beta}_{FE}^w = \left( \sum_i X_i'Q_iX_i \right)^{-1} \sum_i X_i'Q_iy_i^w.
\]

(3.14)

(7) Formulate the true bootstrap second-step model

\[
\delta_{it}^w = z_i\hat{\gamma}_{FE} + u_{it}^w,
\]

(3.15)

where we obtain \( u_{it}^w \) using the bootstrap analog to (2.9):

\[
u_{it}^w = \bar{\xi}_{it} - x_i(\beta_{FE}^w - \hat{\beta}_{FE}).
\]

(3.16)
To formulate the WRE bootstrap DGP for the second step, we consider a system of two equations: one structural equation with one endogenous explanatory variable, \( z_{i2} \), and one reduced-form equation. For the structural equation, we write the bootstrap analog to (2.10):

\[
\delta_{i}^{w} = z_{i1} \hat{\gamma}_{1} + z_{i2}^{w} \hat{\gamma}_{2} + u_{i}^{w},
\]

where the reduced-form equation is

\[
z_{i2}^{w} = w_{i} \hat{\pi} + \nu_{i2}^{w}.
\]

Assuming that we wish to test the null hypothesis that \( \gamma_{2} = 0 \), we compute the WRE following Davidson and MacKinnon (2010). As they point out, computing the OLS estimator of \( \hat{\pi} \) does not produce an efficient estimator, and if the instruments are weak, this estimator may be very inefficient. We obtain their more efficient estimator by generating the residuals, \( \hat{u}_{i} \), from the estimation under the null of equation (2.10) by restricted OLS (since no other endogenous variable appears on the right-hand-side of the restricted equation (2.10)) and then regressing \( z_{i2} \) on \( w_{i} \) and \( \hat{u}_{i} \). Then we use the OLS estimates of \( \hat{\pi} \) and, modifying (2.11), we obtain \( \tilde{v}_{i2} = z_{i2} - w_{i} \hat{\pi} \), which yields \( v_{i2}^{w} = \tilde{v}_{i2} \epsilon_{i} \). Finally we obtain \( z_{i2}^{w} \) using (3.18). We then generate \( \delta_{i}^{w} \) from (3.17), where under the null we impose the restriction that \( \hat{\gamma}_{2} = 0 \). Finally, we regress \( \delta_{i}^{w} \) on \( z_{i1} \) and \( z_{i2}^{w} \) to obtain \( \gamma_{2}^{wre} \), which is the WRE bootstrap estimator of \( \hat{\gamma}_{2} \). To be clear, one must not draw a second set of \( \epsilon_{i} \) values for the second-step estimation, but instead reuse the same set of \( \epsilon_{i} \) drawn in step (5).

To obtain the WU estimator for (3.17), we parallel the WRE approach without imposing the null. First, we use the OLS estimates of \( \hat{\pi} \) and \( \tilde{v}_{i2} = z_{i2} - w_{i} \hat{\pi} \)
to obtain \( v'_{i2} = \hat{v}_{i2} \epsilon_i \), (3.18) to obtain \( z'_{i2} \), and (3.17) to generate \( \delta''_i \). Then we regress \( \delta''_i \) on \( z_{i1} \) and \( z'_{i2} \) to obtain \( \gamma''_{2} \), which is the WU bootstrap estimator of \( \hat{\gamma}_2 \).

(10) Iterate steps 5-9 for a given number of bootstrap trials.

Although not used in our Monte Carlo simulations, we can generalize estimation of the WRE estimator to the case of two endogenous explanatory variables as follows. Assume that we are testing the null that \( \gamma_{2,1} = 0 \) and consider the structural equation,

\[
\delta''_i = \hat{\gamma}_0 + z''_{i1} \hat{\gamma}_1 + z''_{i2} \hat{\gamma}_2 + u''_i
\]  
(3.19)

and two reduced-form equations,

\[
z''_{i2,1} = w_i \hat{\pi}_1 + u''_{i2,1},
\]  
(3.20)

\[
z''_{i2,2} = w_i \hat{\pi}_2 + u''_{i2,2}.
\]  
(3.21)

To obtain the WRE estimator, first impose the restriction that \( \gamma_{2,1} = 0 \) and compute the restricted 2SLS estimator of (3.19) and its residual \( \hat{u}_i \). We then separately regress \( z_{i2,1} \) and \( z_{i2,2} \) on \( w_i \) and \( \hat{u}_i \), obtaining \( \hat{\pi}_1 \) and \( \hat{\pi}_2 \). From these we compute \( \hat{v}_{i2,1} = z_{i2,1} - w_i \hat{\pi}_1 \) and \( \hat{v}_{i2,2} = z_{i2,2} - w_i \hat{\pi}_2 \) and the bootstrap residuals \( u''_{i2,1} = \hat{v}_{i2,1} \epsilon_i \) and \( u''_{i2,2} = \hat{v}_{i2,2} \epsilon_i \) in (3.20) and (3.21), respectively. We then compute the left-hand sides of these equations and insert them into (3.19) to obtain \( \delta''_i \), where the null is imposed. Finally we regress \( \delta''_i \) on \( z''_{i1} \) and \( z''_{i2} \) to obtain \( \gamma''_{2,2} \). The WU estimator is straightforward for the case of one or more endogenous variables using the following matrix form:

\[
\hat{\gamma}''_{F,E} = \left[ \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \left( \sum_i w_i' z_i \right) \right]^{-1} \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \sum_i w_i' \delta''_i.
\]  
(3.22)
4. The Bias of Bootstrap Estimators

To determine the bias of the pairs estimator, we need to reformulate it so that we stochastically resample the error term. Since by definition \( y_i \) equals the fitted model plus the residual for observation \( i \), we can write the pairs resampling process as

\[
\sqrt{\nu_i} y_i = \sqrt{\nu_i} X_i \hat{\beta}_{FE} + \sqrt{\nu_i} (j_T \otimes z_i) \hat{\gamma}_{FE} + \sqrt{\nu_i} \hat{\xi}_i, \tag{4.1}
\]

where the \( \nu_i \) is the number of times (from 0 to \( N \)) that each \((y_i, X_i)\) pair for observation \( i \) is reused in the pairs bootstrap sample. Using \((\sqrt{\nu_i} y_i, \sqrt{\nu_i} X_i)\) we obtain an alternative formulation of the pairs first-step estimator as

\[
\hat{\beta}_{FE}^p = \left( \sum_i \nu_i X'_i Q_i X_i \right)^{-1} \sum_i \nu_i X'_i Q_i y_i. \tag{4.2}
\]

Hence, (4.2) is a weighted regression version of (3.6), where \( y_i^p \) is replaced by \( \sqrt{\nu_i} y_i \), \( X_i^p \) is replaced by \( \sqrt{\nu_i} X_i \), and \((j_T \otimes z_i^p)\) is replaced by \( \sqrt{\nu_i} (j_T \otimes z_i) \). Then, by definition, (3.8) becomes

\[
\sqrt{\nu_i} \delta_i = \sqrt{\nu_i} z_i \hat{\gamma}_{FE} + \sqrt{\nu_i} \hat{u}_i, \tag{4.3}
\]

where \( \hat{u}_i \) is the residual computed from (2.8), which leads to this more useful expression of the second-step pairs estimator:

\[
\hat{\gamma}_{FE}^p = \left[ \left( \sum_i \nu_i z'_i w_i \right) \left( \sum_i \nu_i w'_i w_i \right)^{-1} \left( \sum_i \nu_i w'_i z_i \right) \right]^{-1} \left( \sum_i \nu_i z'_i w_i \right) \left( \sum_i \nu_i w'_i w_i \right)^{-1} \sum_i \nu_i w'_i \delta_i. \tag{4.4}
\]

Given this formulation and our previous assumptions about the error term of the original model in (2.2) and the residuals in the bootstrap models, we now prove theorems establishing the bias of the wild and pairs estimators. Theorems 1 and 2 in the Appendix
relate to the bias of the wild estimator. Theorem 1 establishes that the first-step WU and WRE estimators are unbiased under (2.5), where some of the time-invariant regressors are correlated with the effects. Theorem 2 shows that both second-step wild estimators are unbiased under (2.5).

Theorems 3 and 4 relate to the bias of the pairs estimator. Theorem 3 shows that the pairs estimator is unbiased under (2.5)(endogeneity) in the first step if $E(Q_i \hat{\xi}_i | \nu_i, X_i) = 0$, while Theorem 4 says that the pairs estimator is biased under (2.5) in the second step, because there is no analog to the condition $E(\epsilon_i) = 0$, which exists for the wild estimators. This theorem also shows how weak instruments and strong endogeneity in the original data affect the bias of the pairs estimator. To prove Theorems 3 and 4, we use (4.4) and follow the methodology of Atkinson and Wilson (1992) by relating the pairs bootstrap residuals to the true model residuals, whose properties are known. Assuming that all data are standardized (mean zero and variance one), without loss of generality, increasing $\sum_i z'_iw_i$ implies an increase in the correlation of the instruments and the explanatory variables, i.e., instrument strength, while decreasing $\sum_i w'_iu_i$ implies a decrease in the correlation of the instruments with the error term. These effects will decrease each of the terms comprising the bias in Theorem 4. Because of Theorems 2 and 4, the exact strength and validity of the instruments used to estimate $\hat{\gamma}_{FE}$ should be less important for the WRE and WU bootstrap methods than for the pairs.

Davidson and MacKinnon (1999) show that the ERP of bootstrap estimators is a function of the bias in the bootstrap parameter estimators. If some of the variables in $z_i$ are endogenous, we have shown that the wild second-step estimators are unbiased but that the pairs second-step estimator is biased. Thus, the ERP of this estimator should be
larger than that of the wild estimator, which should in theory be zero. The Monte Carlo results presented below are consistent with these theoretical results.

Building on Theorems 1-4, we can easily show that our bootstrap estimators are consistent. However, we omit these proofs since Horowitz (2001) has already demonstrated these properties.

Because many of the instruments generated from within HT-type panel-data models may be weak, we examine the implications of weak and strong instruments crossed with weak and strong endogeneity in our Monte Carlo simulations. As we have shown, weak instruments coupled with strong degrees of endogeneity and small $N$ will increase the bias of the second-step pairs bootstrap in estimating $\hat{\gamma}_{FE}$. Small-sample $t$-values for the pairs second-step estimator should be less accurate in terms of size than those for the wild, with inaccuracies disappearing as $N$ becomes larger, as the degree of endogeneity becomes weaker, and as instruments become stronger.

5. Monte Carlo Estimation

The data-generation process must define (a) which variables to treat as endogenous, (b) the degree of correlation between $u_i$ and the reduced-form errors, $v_{it2.g}$, and (c) the number of instruments and their correlation with the endogenous variables.

5.1 Data Generation

We create the data for each of our Monte Carlo experiments, which cross weak/strong instrument strength with high/low endogeneity, using the following steps:

(1) Set $\beta = \gamma = 1$.

(2) Assume that $x_{it}$ is an $(1 \times 10)$ vector and that $z_i$ is an $(1 \times 3)$ vector, whose first element, $z_{i,0} = 1$ in (2.1).
(3) Set $g = 1$ and designate $z_{i2,1} = z_{i2}$ as the single endogenous variable in equations (2.10) and (2.11); we generate $z_{i2}$ in step (10). We also specify one exogenous element of $z_{i1,1}$.

(4) Generate the $x_{it,k}$ ($k = 1, \ldots, 10$) and $z_{i1,1}$ with zero means, unit variances, and $\text{cov}(x_{it,k}, z_{i1,1}) = .3$. We set $\text{cov}(x_k, x_{k'}) = .3, \ k \neq k'$. Then we compute the time means of $z_{i1,1}$, which we use for $z_{i1}$.

(5) Generate an $(NT \times 1)$ vector $e_{it}$ and an $(N \times 1)$ vector for $c_i$ as i.i.d. normal random variables with mean zero and variance of 100 and 10, respectively. The large relative variance for $e_{it}$ guarantees a relatively low $R^2$ for the first-step regression, so that considerable variation remains to be explained in the second step.

(6) Generate group means for $e_{it}$ and $x_{it}$.

(7) Generate $u_i$ using (2.9) and (2.6).

(8) Following Hahn and Hausman (2002), solve for the implied reduced-form coefficients as

$$
\pi_j = \left\{R^2_f/[(1 - R^2_f)J]\right\}^{1/2},
$$

where $j = 1, \ldots, J$ and $J$ is the number of instruments, conditional on a value for $R^2_f$, which is the theoretical first-stage $R^2$ for the structural equation after the reduced-form equation has been substituted into it. The instrument set $w_i$ consists of the first time realization of each time-demeaned $x_{itk}$ (in the spirit of Breusch, Mizon, and Schmidt (1989)), the exogenous $z_{i1,1}$, and five additional randomly drawn instruments (excluded from (2.1)), which are normally distributed with mean zero and variance 1. Thus, $J = 16$. We alternately set $R^2_f$ to .2 and .9 to generate data sets with weak and strong instruments.
(9) Generate \( v_i \) by first specifying the correlation between \( u_i \) and \( v_i \), \( \rho_{uv} \), as .2 and .9, which we refer to as weak and strong endogeneity. Then we generate \( v_i \) which has the specified \( \rho_{uv} \) by inserting \( u_i \), its standard deviation, \( \sigma_u \), and \( \rho_{uv} \) into
\[
v_i = \rho_{uv} u_i / \sigma_u + (1 - \rho_{uv}^2)^{\frac{1}{2}} \tilde{u}_i,
\]
where \( \tilde{u}_i \sim N(0,1) \). One can easily verify that
\[
\text{corr}(v_i, u_i) = \rho_{uv}.
\]
(10) Use equation (2.11) to generate \( z_{i2} \), where we replace \( v_{i2,1} \) with \( v_i \).
(11) Expand the variable \( z_{i2} \) into an \((NT \times 1)\) variable by replicating each \( i \) observation \( T \) times, and generate \( y_{it} \) using equation (2.1).

5.2 Monte Carlo Experiments and Results

We perform 10,000 Monte Carlo trials for each of the four combinations of instrument strength and degree of endogeneity to compare the actual size and power of the \( t \)-statistics, which we construct for our four estimators of the asymptotic variance of \( \widehat{\gamma}_2 \): the asymptotic formula, the pairs bootstrap, and the two wild bootstrap methods. The number of bootstrap draws, \( B \), is set to 399 so that \( \frac{1}{2} \alpha (B + 1) \) is an integer.\(^1\) After employing the bootstrap DGPs explained above, one obtains \( \gamma_{2,b}^* \), the bootstrap estimator of \( \gamma_{2,b} \). Using these estimates, we construct
\[
t_{2,b}^* = (\gamma_{2,b}^* - \hat{\gamma}_2) / \text{se}(\gamma_{2,b}^*),
\]
where \( \text{se}(\gamma_{2,b}^*) \) is the estimator of the standard error of \( \gamma_{2,b}^* \) for each \( b \) obtained by substituting the bootstrap data into the estimator of the asymptotic covariance of \( \hat{\gamma}_{FE} \), which

---

\(^1\) MacKinnon (2002) states that while this number may be smaller than should be used in practice, any randomness due to \( B \) of this size averages out across the replications. We find this to be true for our experiments where larger values of \( B \) did not change our results up to three significant digits beyond the decimal point.
is the finite sample covariance matrix corresponding to (3.2). The WRE estimator uses \( \hat{\gamma}_2 = 0 \), since this null was imposed during estimation.

For each Monte Carlo trial, we compare the nominal critical value of each bootstrap method’s \( t \) statistic, \( \hat{t}^* \), with the empirical distribution of the test statistics \( t_{j,2}^* \) for \( j = 1, \ldots, B \). Then, we compute the equal-tailed bootstrap \( P \) values as

\[
\hat{p}^*(\hat{t}) = 2 \min \left[ \frac{1}{B} \sum_{j=1}^{B} I(t_{j,2}^* \leq \hat{t}_l^*), \frac{1}{B} \sum_{j=1}^{B} I(t_{j,2}^* \geq \hat{t}_u^*) \right].
\]

This amounts to two tests, one with regard to the values in the lower (\( l \)) tail of the distribution of \( t_{j,2}^* \) and one with regard to the upper (\( u \)) tail. We reject the null if either yields a bootstrap \( P \) value less than \( \alpha/2 \). Over all \( M \) trials, we compute the rejection probability (RP) as the percent of trials which lead to rejection. We set the upper-tail nominal level of \( \hat{t}_u^* = t_{\alpha/2} = 1.96 \) and the lower-tail nominal level of \( \hat{t}_l^* = t_{1-\alpha/2} = -1.96 \) for \( \alpha = .05 \).

For each Monte Carlo trial, we also compute the size of the estimated \( t \)-values using the estimator of the asymptotic standard errors of \( \hat{\gamma}_2 \). For the \( m^{th} \) trial, \( m = 1, \ldots, M \), we obtain

\[
t_{2,m} = (\hat{\gamma}_{2,m} - \gamma_2)/s_{\hat{\gamma}_{2,m}},
\]

where \( s_{\hat{\gamma}_{2,m}} \) is the standard error estimator taken from the finite sample covariance matrix corresponding to (3.2). For each trial, we compute the equal-tailed \( P \) value as

\[
\hat{p}(\hat{t}) = 2 \min \left[ (t_{2,m} \leq \hat{t}_l), (t_{2,m} \geq \hat{t}_u) \right].
\]

We set the upper nominal level of \( \hat{t}_u = t_{\alpha/2} = 1.96 \) and the lower nominal level of \( \hat{t}_l = t_{1-\alpha/2} = -1.96 \) for \( \alpha = .05 \). This again amounts to two tests, one with regard to the
values in the lower tail of the distribution of $\hat{t}$ and one in the upper tail. We reject the null hypothesis if either yields a bootstrap $P$ value less than $\alpha/2$. Over $M$ trials, we compute the RP.

Because we are interested in performance under large-$N$ asymptotics, we consider the following cases: $N = 50, 100, 200, 400, 800$, and with weak instruments, $N = 1,600$. Extensive preliminary investigation found that increasing $T$ to 10 and 20 did not significantly affect ERP, so we only report results with $T = 5$.$^{11}$ Figures 1–4 report the actual RP with $z_{i2}$ endogenous. We consider all combinations of low/high values of $\rho_{uv}$ and low/high $R^2_f$ in the four figures. In Figures 1 and 2, instruments are very strong, with $R^2_f = .9$. Despite this, the non-wild methods over-reject, exhibiting an ERP that is substantially larger than those of the wild methods with small $N$. In Figure 1, where endogeneity is high, with $\rho_{uv} = .9$, the asymptotic formula substantially over-rejects for $N < 200$, the pairs does little better, while the two wild bootstrap methods are clearly superior. They produce very accurate RP which are very similar even for $N = 50$. In Figure 2, where endogeneity is low, with $\rho_{uv} = .2$, the asymptotic formula still substantially over-rejects with small $N$, although slightly less so than with high endogeneity. The pairs clearly outperforms the asymptotic method until $N = 800$. The two wild bootstrap methods do slightly better than the pairs, producing very small and similar ERP.

In Figures 3 and 4 instruments are very weak, with $R^2_f = .2$, causing the performance of the pairs and asymptotic methods to deteriorate substantially. In Figure 3, where endogeneity is high, with $\rho_{uv} = .9$, the ERP of the pairs and asymptotic methods are unacceptably large. While the asymptotic formula grossly over-rejects and the pairs does a little better for $N < 400$, the two wild methods again produce highly similar ERP that

$^{11}$ For the cases of $T = 5, 10,$ and 20 and $N = 50, 100, 200, 400, 800$, and 1,600, interacted with two levels of endogeneity and instrument strength, we regressed actual size on an intercept, $T$, $N$, $R^2_f$, and $\rho_{uv}$. The estimated coefficient on $T$ was highly insignificant while the coefficient estimates for the other variables were highly significant with the expected signs at the .05 level using a two-tailed asymptotic $t$-test.
are very small at all \(N\). In Figure 4, where endogeneity is low, with \(\rho_{uv} = .2\), the ERP of the pairs and asymptotic methods are considerably smaller, but still almost 100% too large with \(N = 50\). The pairs is slightly worse than the asymptotic formula. The two wild methods under-reject somewhat with small \(N\) but their ERP approach zero quickly as \(N\) increases. Again, the wild methods clearly dominate the pairs and asymptotic methods at all values of \(N\).

Our results for the second-step WRE, WU, and pairs estimators are consistent with Theorems 1-4, which indicate that the WU and WRE are unbiased and that the bias of the second-step pairs estimator will decrease as \(R^2_f\) increases and will increase as \(\rho_{uv}\) rises. Further, all biases decrease as \(N\) increases, although ERP are still positive for the pairs and asymptotic methods with \(N = 1,600\) under weak instruments.

Finally we examine the power of the conventional and bootstrap tests. Since the bootstrap methods nearly always reject less frequently than the conventional methods, the former will nearly always appear to have less power. Therefore, we compute power based on level-adjusted \(t_{2,b}^*\) values, so that critical values are used for which the actual RP is exactly equal to the nominal RP for each method. These levels, \(t_{\alpha/2}^*\) and \(t_{1-\alpha/2}^*\), are the \(\alpha/2\) and \(1 - \alpha/2\) quantiles of the \(t_{2,b}^*\). For each Monte Carlo replication, they are found by first sorting the \(t_{2,b}^*\) values from large to small and then taking the \((\alpha/2)(B + 1)\) and \((1 - \alpha/2)(B + 1)\) values. We compute the power curves for the alternative hypothesis for \(\gamma_2\), as it is increased from -1 to 1 in increments of .1, as the percentage of the \(B\) bootstrap estimates that falls outside the level-adjusted critical region. These percentages are then averaged over all Monte Carlo trials. For the asymptotic method, we use the same range of alternative parameter values to compute power, defined as the percentage of \(M\) estimated \(t_{2,m}\) values that falls outside the interval defined by their level-adjusted critical
region. In Figures 5-7 we present power curves for the case of strong endogeneity and weak instruments for $N = 100, 200, \text{ and } 400$. We see little systematic difference among the methods and therefore conclude that the wild methods are equal or superior to the other two methods in terms of power and size. However, overall we find few differences between the WRE and WU methods.

6. Conclusions

Panel data are widely employed because they allow one to deal with the correlation of regression variables with unobserved heterogeneity by simply applying OLS to time-demeaned data. However, the time-demeaning also removes time-invariant variables from the regression, and such variables are frequently of policy interest and themselves correlated with the unobserved effects. A common assertion in the panel-data literature is that recovering the partial effects of these variables requires the specification and estimation of a restrictive RE model. That is not the case, since the effects of time-invariant variables can also be recovered with a two-step procedure in the spirit of Hausman and Taylor’s (1981) “consistent, but inefficient” IV estimator. We resurrect the HT estimator, modify it to a FE context, and derive its asymptotic covariance matrix, allowing for heteroscedasticity and serial dependence in the errors.

Because of the finite-sample bias of the asymptotic formula, especially with weak instruments, we consider bootstrapping alternatives that use the asymptotic formula to achieve asymptotic refinements rather than computing $t$-values in the classical manner. We develop a pairs and two wild bootstrap estimators for the second-step panel data IV estimation problem. For this model, Theorems 1 and 2 show that the second-step WU and WRE parameter estimators are unbiased, while Theorems 3 and 4 show that the second-step pairs parameter estimator is biased. Theorem 4 also shows that the bias of the pairs
bootstrap will increase with weaker instruments, stronger endogeneity, and smaller $N$. Greater bias should translate into greater ERP for the pairs estimator, using the results of Davidson and MacKinnon (1999).

Monte Carlo experiments with weak and strong instruments in combination with weak and strong endogeneity, allow us to compare the size and power of $t$-statistics for our alternative estimators. In terms of ERP our results are highly consistent with those expected from our theorems on bias. We report results for increasing $N$ with $T = 5$, since consistent with the notion of large–$N$ asymptotics, increasing $T$ has no significant effect. With all four scenarios, the WU and WRE methods generate highly accurate ERP and are clearly preferable, although there is no clear distinction between them. When $R_j^2$ and $\rho_{uw}$ are high with $N = 50$, the ERP of the asymptotic method is almost twice those of the wild estimators, which are very small. The ERP for the wild methods fall to zero when $N = 800$. The performance of the pairs is about midway between that of the asymptotic and wild methods.

When $R_j^2$ is high and $\rho_{uw}$ is low, the asymptotic method performs better but is still substantially worse than the bootstrap methods, which exhibit very low ERP. Differences between the wild and pairs methods are small.

When $R_j^2$ is low and $\rho_{uw}$ is high, the asymptotic and pairs methods exhibit their worst performance with unacceptably high ERP. They are about three times too large with $N = 400$. The WRE and WU bootstrap methods exhibit very small size distortions under this scenario.

When $R_j^2$ and $\rho_{uw}$ are low, compared to the case of high $R_j^2$ and low $\rho_{uw}$, both non-wild methods perform somewhat worse. The wild bootstrap methods exhibit vary small ERP.
The ERP shrinks for all methods as \( N \) increases. However, ERP fail to reach zero even with \( N = 1, 600 \) for the asymptotic and pairs methods with low \( R^2_f \), regardless of \( \rho_{uv} \), which emphasizes the importance of strong instruments the for non-wild estimators.

The wild bootstrap methods are slightly better than the other methods in terms of power. Judged by size and power, the bootstrap methods are clearly preferable. Moreover, the performance of the WRE is very similar to that of the WU.

7. Appendix

We assume (2.4) and (2.5) so that some \( z_{i2,g} \) variables in the second-step regression are endogenous but that the first-step estimation process is subject to weak exogeneity. Then we obtain the following results for the first-step WU estimator and for the first-step WRE, so long as the null is true.

1. Unbiasedness of the WU and WRE first-step estimator, \( \hat{\beta}_{FE}^w \)

**Lemma 1:** Since \( \epsilon_i \) is drawn independently and \( E(\epsilon_i) = 0 \), \( E(\xi_i^w | X_i) = 0 \).

**Proof of Lemma 1:** From (3.13), \( \xi_i^w = \hat{\xi}_i \epsilon_i \vartheta \), where \( \vartheta = 1/(1 - h_{it})^{1/2} \). Thus, \( E(\xi_i^w | X_i) = E(\hat{\xi}_i \epsilon_i \vartheta | X_i) = E(\hat{\xi}_i | X_i)E(\epsilon_i | X_i) \vartheta = E(\hat{\xi}_i | X_i)E(\epsilon_i) \vartheta = 0 \), since \( \epsilon_i \) is independent of \( \hat{\xi}_i \) and \( X_i \) and in addition \( E(\epsilon_i) = 0 \) by definition in (3.11).

**Theorem 1:** Assuming the exogeneity conditions in (2.3) for the first step and using Lemma 1, the wild bootstrap first-step estimator \( \hat{\beta}_{FE}^w \) is unbiased for \( \hat{\beta}_{FE} \).

**Proof of Theorem 1:** Writing the vector form of (3.12) as \( y_i^w = X_i \hat{\beta}_{FE} + z_i \gamma_{FE} + \xi_i^w \) and substituting into (3.14), the first-step wild estimator can be written as

\[
\hat{\beta}_{FE}^w = \hat{\beta}_{FE} + \left( \sum_i X_i'Q_iX_i \right)^{-1} \sum_i X_i'Q_i \xi_i^w. \tag{7.1}
\]
Then
\[ E(\hat{\beta}_{FE}^w|X_i) = \hat{\beta}_{FE} + \left(\sum_i X'_iQ_iX_i\right)^{-1}\sum_i X'_iQ_iE(\xi_i^w|X_i) = \hat{\beta}_{FE}, \] (7.2)
using Lemma 1. Further, \( E[E(\hat{\beta}_{FE}^w|X_i)] = E(\hat{\beta}_{FE}^w) = \hat{\beta}_{FE}. \)

2. The bias of the WU and WRE second-step estimators, \( \hat{\gamma}_{FE}^w. \)

First, consider the WU estimator.

**Lemma 2:** Since \( \epsilon_i \) is drawn independently and \( E(\epsilon_i = 0) \), \( E(\tilde{\xi}_i^w, z_i, w_i, \bar{x}_i) = 0. \)

**Proof of Lemma 2:**

Use the definition of \( \xi_i^w \) in (3.13) and condition on \( z_i, w_i, \bar{x}_i. \) Then use the independence of \( \epsilon_i \) from \( \tilde{\xi}_i^w, z_i, w_i, \bar{x}_i. \)

**Theorem 2:** Given the endogeneity conditions in (2.4) and (2.5), the WU second-step estimator \( \hat{\gamma}_{FE}^w \) in (3.22) is unbiased for \( \gamma_{FE}. \)

**Proof of Theorem 2:**

First substitute (3.15) into (3.22) and take conditional expectations to obtain
\[ E(\hat{\gamma}_{FE}^w|z_i, w_i, \bar{x}_i) = \gamma_{FE} \]
\[ + \left[ \left( \sum_i z'_iw_i \right) \left( \sum_i w'_iw_i \right)^{-1} \left( \sum_i w'_iz_i \right) \right]^{-1} \]
\[ \left( \sum_i z'_iw_i \right) \left( \sum_i w'_iw_i \right)^{-1} \left( \sum_i w'_iu'^w_i \right). \] (7.3)

Further, substitute (3.16) into (7.3) to obtain
\[ E(\hat{\gamma}_{FE}^w|z_i, w_i, \bar{x}_i) = \gamma_{FE} \]
\[ + \left[ \left( \sum_i z'_iw_i \right) \left( \sum_i w'_iw_i \right)^{-1} \left( \sum_i w'_iz_i \right) \right]^{-1} \]
\[
\left( \sum_i z_i w_i \right) \left( \sum_i w_i w_i \right)^{-1} \left( \sum_i w_i E[(\bar{x}_i(\hat{\beta}_{FE} - \hat{\beta}^w_{FE}) + \hat{\xi}_i)|z_i, w_i, \bar{x}_i] \right) = \gamma_{FE},
\]

(7.4)

where we employ Theorem 1 and Lemma 2, which make the two terms of the conditional expectation equal to zero.

For the WRE, repeat the above steps imposing the null hypothesis in the bootstrap models.

3. Unbiasedness of the pairs first-step estimator, $\hat{\beta}^p_{FE}$.

To show the unbiasedness of the pairs first-step estimator we first prove Lemma 3.

**Lemma 3:** Assuming (2.3), $E(Q_i \hat{\xi}_i|\nu_i, X_i) = 0$.

**Proof of Lemma 3:** First,

\[
E(Q_i \hat{\xi}_i|\nu_i, X_i) = E(M_i Q_i \xi_i|\nu_i, X_i)
\]

\[
= E(Q_i \xi_i|\nu_i, X_i) - Q_i X_i \left( \sum_i X'_i Q_i X_i \right)^{-1} X'_i Q_i E(Q_i \xi_i|\nu_i, X_i)
\]

\[
= E\{Q_i[(j_T \otimes c_i) + e_i]|\nu_i, X_i\}
\]

\[
- Q_i X_i \left( \sum_i X'_i Q_i X_i \right)^{-1} X'_i Q_i E\{Q_i [(j_T \otimes c_i) + e_i]|\nu_i, X_i\},
\]

(7.5)

using $M_i = I_T - Q_i X_i \left( \sum_i X'_i Q_i X_i \right)^{-1} X'_i Q_i$ and $\xi_i = (j_T \otimes c_i) + e_i$. Then substitute using (2.3) and the fact that $Q_i$ eliminates $c_i$ completes the proof.

**Theorem 3:** Given Lemma 3, the bootstrap pairs first-step estimator, $\hat{\beta}^p_{FE}$, is unbiased for $\hat{\beta}_{FE}$.
Proof of Theorem 3: Substitute $Q_iy_i$ into (4.2) and take expectations. Then

$$E(\hat{\beta}_{FE}^p|\nu_i, X_i) = \hat{\beta}_{FE} + \left( \sum_i \nu_i X_i' Q_i X_i \right)^{-1} \sum_i \nu_i X_i' Q_i E(Q_i \hat{\xi}_i|\nu_i, X_i)$$

$$= \hat{\beta}_{FE}, \quad (7.6)$$

using Lemma 3.

4. Bias of the pairs second-step estimator, $\hat{\gamma}_{FE}^p$

Theorem 4: Assuming an unbiased first-step pairs estimator, (2.4) causes the pairs second-step estimator, $\hat{\gamma}_{FE}^p$, to be biased for $\hat{\gamma}_{FE}$.

Proof of Theorem 4: Substituting (4.3) into (4.4) we obtain

$$\hat{\gamma}_{FE}^p = \hat{\gamma}_{FE} + \left[ \sum_i \nu_i z_i' w_i \right] \left( \sum_i \nu_i w_i' w_i \right)^{-1} \left( \sum_i \nu_i w_i' z_i \right)^{-1} \left( \sum_i \nu_i w_i' \hat{u}_i \right). \quad (7.7)$$

Defining $m_i$ as

$$m_i = \left\{ 1 - z_i \left[ \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \left( \sum_i w_i' z_i \right) \right]^{-1} \right.$$ \n
$$\left. \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \sum_i w_i' \right\} \quad (7.8)$$

we can relate $\hat{u}_i$ to $u_i$ using $\hat{u}_i = m_i u_i$:

$$\hat{u}_i = \left\{ 1 - z_i \left[ \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \left( \sum_i w_i' z_i \right) \right]^{-1} \right.$$ \n
$$\left. \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \sum_i w_i' \right\} u_i$$

$$= u_i - z_i \left[ \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \left( \sum_i w_i' z_i \right) \right]^{-1} \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \left( \sum_i w_i' u_i \right) \quad \quad (7.9)$$
Then substitute (7.9) into (7.7) to obtain

\[
\hat{\gamma}_{FE}^p = \hat{\gamma}_{FE} + \left[ \sum_i \nu_i z_i' w_i \left( \sum_i \nu_i w_i' w_i \right)^{-1} \left( \sum_i \nu_i w_i' z_i \right) \right]^{-1}
\]

\[
\left( \sum_i \nu_i z_i' w_i \right) \left( \sum_i \nu_i w_i' w_i \right)^{-1} \left( \sum_i \nu_i w_i' u_i \right)
\]

\[
- \left[ \left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \left( \sum_i w_i' z_i \right) \right]^{-1}
\]

\[
\left( \sum_i z_i' w_i \right) \left( \sum_i w_i' w_i \right)^{-1} \left( \sum_i w_i' u_i \right)
\]

(7.10)

Conditioning on \((z_i, w_i, \bar{x}_i, \nu_i)\), we take the expectation of both sides and then focus on

\[
E(u_i|z_i, w_i, \bar{x}_i, \nu_i) = E(\hat{\xi}_i|z_i, w_i, \nu_i) + x_i E[(\hat{\beta}_{FE} - \beta)|z_i, w_i, \bar{x}_i, \nu_i],
\]

where we have employed (2.9). Since \(\hat{\beta}_{FE}\) is unbiased for \(\beta\) in the first-step estimation via Theorem 3, the second term in (7.11) equals zero. However, using (2.4), the first term in (7.11) does not equal zero:

\[
E(\hat{\xi}_i|z_i, w_i, \nu_i) = E(c_i|z_i, w_i, \nu_i) + E(\tilde{e}_i|z_i, w_i, \nu_i) \neq 0.
\]

(7.12)

9. References


30


**Fig. 1**

Rejection Probabilities for Various N: $\rho_{u,v} = .9, R^2 = .9$
Fig. 2
Rejection Probabilities for Various N; $\rho_{u,v}=.2$, $R^2=.9$

Fig. 3
Rejection Probabilities for Various N; $\rho_{u,v}=.9$, $R^2=.2$
Fig. 4

Rejection Probabilities for Various N; $\rho_{u,v}=.2$, $R^2=.2$

Fig. 5

Power Curves for $N=100$; $\rho_{u,v}=.9$, $R^2=.2$
Fig. 6

Power Curves for N=200; \( \rho_{u,v} = .9, R^2 = .2 \)

Fig. 7

Power Curves for N=400; \( \rho_{u,v} = .9, R^2 = .2 \)