Chapter 5
Inference in Two-Step Panel Data Models with Time-Invariant Regressors: Bootstrap Versus Analytic Estimators

Scott E. Atkinson and Christopher Cornwell

5.1 Introduction

Panel data are useful because of the opportunity they afford the researcher to control for unobserved heterogeneity or effects that do not vary over time. Typically, exploiting this opportunity means employing the fixed-effects (FE) estimator, because it produces consistent estimates of the coefficients of time-varying variables under weak assumptions about their relationship with the effects. As is well-understood, the FE estimator achieves this through a data transformation that eliminates the effects. The downside to FE estimation, however, is that this data transformation also eliminates any time-invariant variables. Consequently, the FE estimator is sometimes abandoned entirely for a random-effects (RE) approach, whose requirements for consistency frequently are not satisfied, since unobserved heterogeneity is often correlated with the regressors.

While the partial effects of time-invariant variables can be recovered in a second-step regression, this fact is generally omitted in most textbook treatments of panel-data methods (Wooldridge 2010 is an exception). The practical necessity of recovering the partial effects of time-invariant variables shows up in many different empirical contexts. In the familiar exercise of estimating wage regressions, human-capital variables such as experience and tenure are taken to be correlated with the unobserved effect, which is commonly interpreted as "ability". While FE estimation eliminates such time-invariant unobservables, it also eliminates race, gender and education (when schooling is completed before the sample period), the effects of which are of great interest. Similarly, when estimating production relationships, inputs may be correlated with fixed "environmental" factors, which again would

S.E. Atkinson (✉) • C. Cornwell
Department of Economics, Terry College of Business, University of Georgia,
503 Brooks Hall, 310 Herty Drive, Athens, GA 30602, USA
e-mail: atknsn@uga.edu; cowrnl@uga.edu

swept away by the FE estimator. However, policy-relevant firm characteristics, such as public or private ownership in the case of electric utilities, would be swept away as well.

Interest in recovering the partial effects of time-invariant variables may be even more common in panels of state and country-level aggregates, as with empirical growth and comparative political-economy studies, where policy-relevant fixed institutional variables are differenced out of a FE regression. The recovery of these partial effects motivated the work of Plümper and Troeger (2007), who proposed a three-step method for estimating the effects of time-invariant variables in linear panel-data models. Plümper's and Troeger's so-called "fixed-effects vector decomposition" (FEVD) became widely utilized through their Stata program (xtfevd) that computes the estimator and corresponding standard errors. However, as shown in the critiques of Breusch et al. (2011) and Greene (2011), the main substantive claims about the FEVD estimator are false. Importantly, from our perspective, the standard errors produced by the FEVD estimator are incorrect, because their method does not estimate the correct asymptotic covariance matrix, which should incorporate the first-step estimated covariance matrix.

In this paper, we focus on the two-step estimation procedure and compare conventional inference based on the asymptotic formula to bootstrap alternatives. Bootstrapping has a natural appeal, because of the complications associated with estimating the asymptotic covariance matrix and the inherent finite-sample bias of the resulting standard errors. Our paper contributes to the panel data and bootstrapping literature in four ways. First, we derive the correct asymptotic covariance matrix for the second-step coefficient estimators, allowing for heteroskedasticity and autocorrelation of unknown form. Second, we develop the steps required to perform bootstrap estimation of the covariance matrix of the second-step estimated coefficients. Third, we prove that the pairs and wild bootstrap coefficient estimators are unbiased. This stands in contrast to Flachaire (2005) who asserts that the pairs is a biased estimator. Unbiasedness implies that the error in rejection probability (ERP) of corresponding t-tests, measured as the difference between their actual and nominal size, should be small.

Finally, using Monte Carlo methods, we compare the size and power of the naive asymptotic estimator, which ignores the first-step estimation error, the correct asymptotic covariance matrix estimator, which does not, and the bootstrap alternatives. We consider a variety of panel sizes (N) and lengths (T) relevant to the common large-N, small-T setting. We find that the bootstrap methods consistently provide more accurate inference than the asymptotic formulae. The performance gain is largest for N less than 250 and shrinks as N grows. Although a small ERP remains for the correct asymptotic covariance matrix estimator when $N = 1,000$, as N grows to 250 and beyond, both bootstrap estimators generally produce the correct size. Comparing the two bootstrap methods, the pairs tends to moderately over-reject and the wild to moderately under-reject for smaller values of N, with the pairs having a slightly smaller ERP. While we find a slight advantage of the pairs method over the others in terms of power with $N = 250$ and $T = 5$, few differences are observed with larger values of N.

The remainder of this paper is organized as follows. Section 5.2 introduces the model and two-step estimator. Section 5.3 presents the correct asymptotic covariance matrix for the second-step estimator and outlines the bootstrap alternatives for second-step standard-error estimation. In Sect. 5.4, we review previous studies of the wild and pairs procedures and prove that they are both unbiased for the problems we address. Section 5.5 explains our Monte Carlo experiments and reports our findings on the size and power of t-tests produced by conventional and bootstrap procedures. Conclusions follow in Sect. 5.6.

5.2 The Two-Step Model and Parameter Estimation

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

$$y_{it} = \mathbf{x}_{it}\beta + \mathbf{z}_i\gamma + \xi_{it} \quad i = 1, \ldots, N; t = 1, \ldots, T, \tag{5.1}$$

where $\xi_{it} = c_i + e_{it}$, y_{it} is the dependent variable, $\mathbf{x}_{it}$ is a $(1 \times K)$ vector of time-varying regressors, $\mathbf{z}_i$ 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.[1] The e_{it} may be heteroscedastic and serially correlated. The coefficient vectors, β and γ, 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:

$$\mathbf{y}_i = \mathbf{X}_i\beta + (\mathbf{j}_T \otimes \mathbf{z}_i)\gamma + +\mathbf{j}_T c_i + \mathbf{e}_i, \tag{5.2}$$

where $\mathbf{X}_i$ is $(T \times K)$, $\mathbf{j}_T$ is a T-vector of ones, and $\mathbf{y}_i$ and $\mathbf{e}_i$ are $(T \times 1)$ vectors.

Our interest is in estimating γ, allowing for the possibility that some or all of the variables in $\mathbf{X}_i$ are correlated with the unobserved effect. Formally, we adopt the standard FE assumption that

$$E(e_{it} \mid \mathbf{X}_i, \mathbf{z}_i, c_i) = 0, \quad t = 1, \ldots, T. \tag{5.3}$$

Additionally, we assume

$$E(c_i \mid \mathbf{z}_i) = 0, \tag{5.4}$$

which treats the time-invariant variables as uncorrelated with the unobserved effects. We invoke (5.4) to focus attention on the transmission of the first-step estimation

---

[1]Although the model setup assumes a balanced panel, this is not necessary. The asymptotic covariance matrix and bootstrap procedures can readily accommodate settings in which the number of time-series observations varies with the cross-section unit.

error to the second step, without the confounding influence of endogeneity in $\mathbf{z}_i$. While this is a strong assumption, it is reasonable in some of the empirical contexts referenced above. For example, in the production relationships category, Agee et al. (2009, 2012) use directional distance functions to estimate the efficiency of a household's production of health and human capital in children. In both cases, all time-invariant child and household characteristics are treated as exogenous on the grounds that the characteristics are either fixed by nature (e.g., race and gender) or by circumstances of the household or the parents. County, state, and country-level panels provide other examples, many of which come from social science papers outside of economics (as described in Plümper and Troeger (2007)). A common practice has been either to include time-invariant variables in a pooled ordinary least-squares (POLS) or RE regression, or use FE estimation and ignore them. Knack (1993), which is concerned with the relationship between the prospect of jury service and voter registration, does both. Over Knack's two-year panel, certain state characteristics, like whether there is a senate contest, are time-invariant. Because the timing of senate contests are fixed by law, this indicator is exogenous. Knack estimates OLS regressions on each year separately, including the fixed characteristics, and FE regressions using both years of the panel, dropping these characteristics.[2]

In the two-step approach to estimating γ, we begin by applying FE to (5.1), which produces

$$\hat{\beta}_{FE} = \left(\sum_i \mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i \right)^{-1} \sum_i \mathbf{X}_i' \mathbf{Q}_T \mathbf{y}_i, \tag{5.5}$$

where $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{j}_T (\mathbf{j}_T' \mathbf{j}_T)^{-1} \mathbf{j}_T'$ is the idempotent projection that time de-means the data. The FE estimator is unbiased and consistent under (5.3).

Next, we take $\hat{\beta}_{FE}$ and compute individual or group-level residuals,

$$\hat{\delta}_i = \bar{y}_i - \bar{\mathbf{x}}_i \hat{\beta}_{FE}, \tag{5.6}$$

and formulate the second-step regression model

$$\hat{\delta}_i = \mathbf{z}_i \gamma + u_i, \tag{5.7}$$

where

$$u_i = \bar{\xi}_i - \bar{\mathbf{x}}_i (\hat{\beta}_{FE} - \beta), \tag{5.8}$$

---

[2] Atkinson and Cornwell (2013) extend the analysis here to allow some of the elements of $\mathbf{z}_i$ to be correlated with the unobserved effect.

$\bar{\xi}_i = c_i + \bar{e}_i$, the over-bar indicates the sample-period mean for unit i (e.g., $\bar{\mathbf{x}}_i = \frac{1}{T} \sum_t \mathbf{x}_{it}$), and the first element of $\mathbf{z}_i$ is 1. Equation (5.8) is obtained by combining (5.7), (5.6), and the sample-period mean for unit i of (5.2). We then estimate γ by applying OLS to (5.7).[3] The resulting estimator, which we label $\hat{\gamma}_{FE}$ because it is derived from $\hat{\beta}_{FE}$, can be written as

$$\hat{\gamma}_{FE} = \left(\sum_i \mathbf{z}_i' \mathbf{z}_i \right)^{-1} \left(\sum_i \mathbf{z}_i' \hat{\delta}_i \right). \tag{5.9}$$

5.3 Second-Step Standard-Error Estimation

In this section, we first derive the asymptotic covariance matrix of $\hat{\gamma}_{FE}$ and explain how to estimate it consistently. Then we outline the procedures for computing the wild and pairs bootstrap alternatives.

5.3.1 Asymptotic Covariance Matrix

As Wooldridge (2010) points out, the asymptotic covariance matrix for $\hat{\gamma}_{FE}$ can be obtained by applying standard arguments for two-step estimators (see, for example, Murphy and Topel 1985). We begin by writing the sampling error of $\hat{\gamma}_{FE}$ as

$$\hat{\gamma}_{FE} - \gamma = \left(\sum_i \mathbf{z}_i' \mathbf{z}_i \right)^{-1} \left(\sum_i \mathbf{z}_i' u_i \right). \tag{5.10}$$

Then we can show that $\sqrt{N}(\hat{\gamma}_{FE} - \gamma)$ is asymptotically normal with a limiting covariance matrix that can be expressed as

$$(\mathbf{B}_{zz})^{-1} \mathbf{A} \, (\mathbf{B}_{zz})^{-1}, \tag{5.11}$$

where, $\mathbf{B}_{zz} = \operatorname{plim} \frac{1}{N} \sum_i \mathbf{z}_i' \mathbf{z}_i$. As implied by (5.8)

$$\mathbf{A} = \operatorname{plim} \frac{1}{N} \sum_i \bar{\xi}_i^2 \mathbf{z}_i' \mathbf{z}_i + \operatorname{plim} \frac{1}{N} \sum_i \mathbf{z}_i' \bar{\mathbf{x}}_i \, \mathbf{V}_{\hat{\beta}_{FE}} \, \bar{\mathbf{x}}_i' \mathbf{z}_i, \tag{5.12}$$

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

---

[3]As discussed in Atkinson and Cornwell (2013), allowing some of the elements of $\mathbf{z}_i$ to be correlated with the unobserved effect leads to the two-step "simple, consistent" instrumental variables estimator of Hausman and Taylor (1981). From this perspective, you can view our two-step estimator as an instrumental variables estimator using $[\mathbf{Q}_T \mathbf{X}_i, (j_T \otimes \mathbf{z}_i)]$ as instruments.

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

$$\hat{\mathbf{V}}_{\hat{\beta}_{FE}} = \left(\sum_i \mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i \right)^{-1} \sum_i \mathbf{X}_i' \mathbf{Q}_T \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \mathbf{Q}_T \mathbf{X}_i \left(\sum_i \mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i \right)^{-1} \qquad (5.13)$$

(see Arellano 1987), and extracting an estimator of ξ_i from the group-level version of (5.1) evaluated at $(\hat{\beta}_{FE}, \hat{\gamma}_{FE})$.

5.3.2   Bootstrap Methods

There are two important reasons to prefer bootstrap estimators of standard errors to estimators based on asymptotic formulae. First, bootstrapping standard errors is often easier than estimating the asymptotic covariance matrix. Second, bootstrapping often produces better small-sample performance in terms of ERP.

We consider the wild and pairs bootstrap procedures because they produce estimated standard errors that are robust to heteroskedasticity. Davidson and Flachaire (2008) have shown that the wild bootstrap yields a heteroskedasticity-consistent covariance matrix estimator when the residuals are divided by h_i, the diagonal element of the projection matrix corresponding to the right-hand-side variables of the original equation estimated. T. Lancaster (2003, A note on bootstraps and robustness, unpublished manuscript. Department of Economics, Brown University) has proven that the pairs bootstrap yields a similar covariance estimator. Below we outline how each can be adapted to our two-step estimation problem.

Following Cameron and Trivedi (2005), for fixed-T panels, consistent (as $N \to \infty$) standard errors can be obtained by using cross-sectional resampling. Hence, we employ this method for both the pairs and wild bootstrap, assuming no cross-sectional or temporal dependence.[4]

5.3.2.1   Wild Bootstrap Estimator

The wild bootstrap procedure can the applied to the estimation of the standard errors of $\hat{\gamma}_{FE}$ by executing the following steps.

---

[4]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.

1. Compute $\hat{\beta}_{FE}$ in (5.5).
2. Using $\hat{\beta}_{FE}$, compute $\hat{\delta}_i$ in (5.6).
3. Compute $\hat{\gamma}_{FE}$ in (5.9).
4. Since $\xi_{it} = c_i + e_{it}$ in (5.1), compute

$$\hat{\xi}_{it} = y_{it} - \mathbf{x}_{it}\hat{\beta}_{FE} - \mathbf{z}_i\hat{\gamma}_{FE}, \tag{5.14}$$

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

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

where h_{it} is the diagonal element of the projection matrix corresponding to the right-hand-side variables of (5.2). Thus, the transformed residual is homoskedastic by definition so long as the error term, ξ_{it}, is homoskedastic.[5]

5. We follow Davidson and Flachaire (2008) and MacKinnon (2002) and define ϵ_i as the two-point Rademacher distribution:

$$\epsilon_i = \begin{cases} -1 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases}. \tag{5.15}$$

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

$$y_{it}^w = \mathbf{x}_{it}\hat{\beta}_{FE} + \mathbf{z}_i\hat{\gamma}_{FE} + \xi_{it}^w, \tag{5.16}$$

where

$$\xi_{it}^w = f(\hat{\xi}_{it})\epsilon_i. \tag{5.17}$$

Davidson and Flachaire (2008) provide evidence that this version of the wild bootstrap is superior to other wild methods. This is due to the fact that $E(\epsilon_i) = 0$, $E(\epsilon_i^2) = 1$, $E(\epsilon_i^3) = 0$, and $E(\epsilon_i^4) = 1$. Since $\hat{\xi}_{it}$ and ϵ_i are independent, $E(\xi_{it}^w) = E(\hat{\xi}_{it})\epsilon_i\vartheta = 0$, its variance is that of $\hat{\xi}_{it}\vartheta$, its third moment is zero (which implies zero skewness in $\hat{\xi}_{it}$), but its fourth moment is again that of $\hat{\xi}_{it}\vartheta$. Thus, the first, second, and fourth moments of $\hat{\xi}_{it}\vartheta$ are reproduced exactly in the wild bootstrap data using (5.15).

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

$$\hat{\beta}_{FE}^w = \left(\sum_i \mathbf{X}_i'\mathbf{Q}_T\mathbf{X}_i \right)^{-1} \sum_i \mathbf{X}_i'\mathbf{Q}_T\mathbf{y}_i^w. \tag{5.18}$$

---

[5]Further, this transformation is needed to obtain a heteroskedastic-consistent covariance matrix as explained above.

7. Compute the group-mean residuals as

$$\delta_i^w = \bar{y}_i^w - \bar{\mathbf{x}}_i \hat{\beta}_{FE}^w. \tag{5.19}$$

Note that in step (5) it was necessary to generate y_{it}^w rather than time-demeaned y_{it}, because in the current step one must compute group means to generate the residuals. Group means cannot be recovered from the time-demeaned data.

8. Formulate the true bootstrap model

$$\delta_i^w = \mathbf{z}_i \hat{\gamma}_{FE} + u_i^w, \tag{5.20}$$

where u_i^w is a bootstrap error, and compute the second-step estimator of $\hat{\gamma}_{FE}$ using the bootstrap data:

$$\hat{\gamma}_{FE}^w = \left(\sum_i \mathbf{z}_i' \mathbf{z}_i \right)^{-1} \sum_i \mathbf{z}_i' \delta_i^w. \tag{5.21}$$

9. Iterate steps 5–8 and compute the sample standard deviation of $\hat{\gamma}_{FE}^w$, $s_{\hat{\gamma},w}$, as an estimator of the standard error of $\hat{\gamma}_{FE}$, where w denotes the wild procedure.

5.3.2.2 Pairs Bootstrap Estimator

The pairs bootstrap procedure discussed in T. Lancaster (2003, A note on bootstraps and robustness, unpublished manuscript. Department of Economics, Brown University) can be extended to our problem as follows:

1. Compute $\hat{\beta}_{FE}$ in (5.5).
2. Using $\hat{\beta}_{FE}$, compute $\hat{\delta}_i$ in (5.6).
3. Compute $\hat{\gamma}_{FE}$ in (5.9).
4. Draw randomly with replacement among $i = 1, \ldots, N$ blocks, using all T observations in the chosen block, with probability $1/T$ from $\{y_{it}, \mathbf{x}_{it}, \mathbf{z}_{it}\}$ to obtain $\{y_{it}^p, \mathbf{x}_{it}^p, \mathbf{z}_{it}^p\}$, 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 ξ_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, ξ_i^p, as

$$\mathbf{y}_i^p = \mathbf{X}_i^p \hat{\beta}_{FE} + (\mathbf{j}_T \otimes \mathbf{z}_i^p) \hat{\gamma}_{FE} + \xi_i^p. \tag{5.22}$$

Compute the FE estimator of $\hat{\beta}$ using the pairs bootstrap data $(\mathbf{y}_i^p, \mathbf{X}_i^p)$

$$\hat{\beta}^p_{FE} = \left(\sum_i \mathbf{X}^{p'}_i \mathbf{Q}_T \mathbf{X}^p_i \right)^{-1} \sum_i \mathbf{X}^{p'}_i \mathbf{Q}_T \mathbf{y}^p_i. \qquad (5.23)$$

6. Using $\hat{\beta}^p_{FE}$ compute the residuals

$$\delta^p_i = \bar{y}^p_i - \bar{\mathbf{x}}^p_i \hat{\beta}^p_{FE}. \qquad (5.24)$$

7. Formulate the second-step pairs bootstrap model as

$$\delta^p_i = \mathbf{z}^p_i \hat{\gamma}_{FE} + u^p_i, \qquad (5.25)$$

where u^p_i is the pairs second-step error, then compute the second-step estimator of $\hat{\gamma}_{FE}$ using the bootstrap data:

$$\hat{\gamma}^p_{FE} = \left(\sum_i \mathbf{z}^{p'}_i \mathbf{z}^p_i \right)^{-1} \sum_i \mathbf{z}^{p'}_i \delta^p_i. \qquad (5.26)$$

8. Iterate steps 4–7 and compute the sample standard deviation of $\hat{\gamma}^p_{FE}$, $s_{\hat{\gamma},p}$, as an estimator of the standard error of $\hat{\gamma}_{FE}$.

5.4 The Size and Power of Bootstrap Estimators

5.4.1 Previous Studies of the Size and Power of Bootstrap Estimators

We are unaware of any Monte Carlo study that examines the ERP and size of estimator t-values for two-step panel-data models of the type we consider. However, there is a substantial literature on bootstrap performance in cross-section regressions. Horowitz (2001) compares the actual size of the pairs and wild bootstrap to the size associated with the asymptotic formula for White's information matrix test, the t-test in a heteroskedastic regression model, and the t-test in a Box-Cox regression model. For relatively small sample sizes, he finds that the wild and pairs dramatically reduce the ERP of the asymptotic formulas, and in many cases the wild essentially eliminates this error. The wild method outperforms the pairs and both outperform the jackknife method. Davidson and Flachaire (2008) obtain similar results when they compare the wild and pairs estimators to those obtained using the asymptotic formula. Using an Edgeworth expansion, they trace the wild's advantage to the fact that the ERPs of the pairs depend on more higher-order raw moments of the original errors and the bootstrap residuals, which are greater under heteroskedasticity. With homoskedastic errors, there is little difference between the

wild and pairs estimators and their ERP is very small. The results we present below are consistent with this finding. Inference based on the asymptotic formula also improves, but exhibits a substantially larger ERP.

In summary, for single-equation models with heteroskedastic errors, Monte Carlo results generally show that the wild bootstrap outperforms the pairs and that both improve on inference based on the estimator of the asymptotic formula. However, we are not aware of bootstrap performance comparisons that address the empirical context of a two-step panel data model, where estimation error from the first step is the primary complicating factor. Next, we analytically examine the wild and pairs procedures for the two-step estimation problem, identifying the conditions required for both to be unbiased.

5.4.2   The Unbiasedness of Our Two-Step Bootstrap Estimators

The unbiasedness of the first and second-step wild estimators follows directly from the fact that the ϵ_i are zero-mean random variables generated independently of ξ_{it}. See Theorems 1 and 2 in the Appendix. Flachaire (2005) compares the conditional expectation of the bootstrap error given the explanatory variables for the wild versus pairs methods in a simple linear model. In terms of our setup, he asserts that the wild bootstrap satisfies $E(u_i^w|\mathbf{z}_i) = 0$, and hence is unbiased.

He also asserts that the pairs does not satisfy $E(u_i^p|\mathbf{z}_i^p) \neq 0$, and therefore is biased (because u_i^p depends on $\mathbf{z}_i^p$). Thus, he argues that the wild should produce a smaller ERP than the pairs estimator. In the Appendix, however, we prove that both estimators are unbiased given the exogeneity conditions (5.3) and (5.4).

To show that the pairs estimator is unbiased, we need to reformulate it in terms of the residual of the original model. Since by definition $\mathbf{y}_i$ equals the fitted model plus the residual for observation i,

$$\sqrt{v_i}\mathbf{y}_i = \sqrt{v_i}\mathbf{X}_i\hat{\beta}_{FE} + \sqrt{v_i}(\mathbf{j}_T \otimes \mathbf{z}_i)\hat{\gamma}_{FE} + \sqrt{v_i}\hat{\boldsymbol{\xi}}_i, \qquad (5.27)$$

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

$$\hat{\beta}_{FE}^p = \left(\sum_i v_i\mathbf{X}_i'\mathbf{Q}_T\mathbf{X}_i\right)^{-1}\sum_i v_i\mathbf{X}_i'\mathbf{Q}_T\mathbf{y}_i. \qquad (5.28)$$

Hence, (5.28) becomes a weighted regression version of (5.23), where $\mathbf{y}_i^p$ is replaced by $\sqrt{v_i}\mathbf{y}_i$, $\mathbf{X}_i^p$ is replaced by $\sqrt{v_i}\mathbf{X}_i$, and $(\mathbf{j}_T \otimes \mathbf{z}_i^p)$ is replaced by $\sqrt{v_i}(\mathbf{j}_T \otimes \mathbf{z}_i)$. Again by definition, reformulate (5.25) as

$$\sqrt{v_i}\delta_i = \sqrt{v_i}\mathbf{z}_i\hat{\gamma}_{FE} + \sqrt{v_i}\hat{u}_i. \qquad (5.29)$$

where $\hat{u}_i$ is the residual computed from (5.7) and obtain a more useful formulation of the second-step pairs estimator as

$$\hat{\gamma}^p_{FE} = \left(\sum_i v_i \mathbf{z}'_i \mathbf{z}_i \right)^{-1} \sum_i v_i \mathbf{z}'_i \delta_i. \qquad (5.30)$$

Theorem 3 says that the pairs estimator is unbiased in the first step if $E(\mathbf{Q}_T \hat{\boldsymbol{\xi}}_i | v_i, \mathbf{X}_i) = 0$. This result also applies to any linear, single-equation model estimated by the pairs estimator. Finally, using (5.30), and assuming that (5.3), (5.4), and Theorem 3 hold, Theorem 4 shows that in the second step, the pairs estimator is unbiased for $\hat{\gamma}_{FE}$.

Davidson and MacKinnon (1999) demonstrate that the ERP depends on estimator bias. Thus, a biased bootstrap estimator should have a larger ERP than an unbiased estimator assuming that the errors, ξ_i, are i.i.d.[6] The size of t-values for both the second-step wild and pairs estimators should be highly accurate since their biases are zero, given that the assumptions in (5.3) and (5.4) hold.

5.5 Monte Carlo Estimation

5.5.1 Data Generation

We create the data for the Monte Carlo experiments in the following steps.

1. Generate the x_{itk} and z_{itg} ($k = 1, \ldots, 10$; $g = 1, \ldots, 3$) as multivariate normal with zero means and unit variances. We set $cov(z_g, z_{g'}) = 0.2$, $g \neq g'$ (implying simple correlations of 0.2), $cov(x_k, x_{k'}) = 0.3$, $k \neq k'$. We also set $cov(x_k, z_g) = 0.3$ and draw x_{itk} and z_{itg}. For each g, we then create the group mean of z_{itg} and use this for z_{ig} (which is time invariant) so that the group means of x_k and z_g have correlation of 0.3.
2. Generate c_i and e_{it} as i.i.d. normal random variables with mean zero and variance of 10 and 100, respectively. The large variance for e_{it} guarantees a relatively low R^2 for the first-step regression. This in turn implies a greater difference between the estimated "naive" and correct asymptotic covariance matrices for the second-step coefficients.
3. Using the data from steps 1–2, generate y_{it} in Eq. (5.1).

The bootstrap estimators do not require the i.i.d. assumption. As described in Sect. 5.3.2, they have heteroskedastic-consistent covariance matrices. Using cross-sectional resampling as defined above, these bootstrap methods will deal with dependent data by generating correlated errors that exhibit approximately the

---

[6]As indicated above, Davidson and Flachaire (2008) find that many other factors in addition to bias, especially heteroskedasticity, can increase the ERP of bootstrap and asymptotic estimators.

same pattern of autocorrelation as ξ_{it}. However, in this paper we focus on the i.i.d. case (as defined in step two), because our primary interest here is how the bootstrap estimators handle the transmission of the first-step estimation error to the second step relative to the analytical alternative in the simplest of contexts. In Atkinson and Cornwell (2013), we explore the effects, in some cases substantial, of heteroskedasticity and serial correlation in the ξ_{it} on the second-step inference problem.

5.5.2 Monte Carlo Results

We perform a number of Monte Carlo experiments to compare the actual size and power of the t-statistics derived from four estimators of the covariance matrix of $\hat{\gamma}$– a "naive" method which is the asymptotic formula without adjusting for the first-step parameter estimators, the correct asymptotic formula which makes this adjustment, the pairs bootstrap, and the wild bootstrap. For our size calculations, we assume that $\beta = \gamma = 1$ during data generation and test the null that $\gamma = 1$ using a two-sided equal-tailed 95 % confidence interval, so that the total type-I error is $\alpha = 0.05$.

We set the number of bootstrap draws, B, to 399 following MacKinnon (2002) who 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 on ERP up to three significant digits beyond the decimal point. Within each of $M(m = 1, \ldots, M)$ Monte Carlo trials, for each bootstrap method, we estimate the unrestricted model and obtain $\hat{\gamma}^*_{m,b,g}$, the bootstrap estimator of $\hat{\gamma}_{m,g}$, $b = 1, \ldots, B$. For each m, we calculate the actual size of the test-statistic,

$$t^*_{m,g} = (\hat{\gamma}_{m,g} - \gamma_g)/s^*_{m,g}, \quad g = 1, \ldots, G, \tag{5.31}$$

where $s^*_{m,g}$ is the bootstrap estimator of the standard error of $\hat{\gamma}_{m,g}$, computed as the standard deviation of $\hat{\gamma}^*_{m,b,g}$ over all bootstrap replications. Note that the $t^*_{m,g}$ statistic is not asymptotically pivotal and no asymptotic refinements obtain; however, we employ it since applied researchers may have difficulty computing the asymptotic formula. See MacKinnon (2002) for details.

For each Monte Carlo trial, m, we calculate the size for each bootstrap estimator as the percentage of $t^*_{m,g}$ values greater than the nominal level of $t^*_{\alpha/2} = 1.96$ or less than the nominal level of $t^*_{1-\alpha/2} = -1.96$, with $\alpha = 0.05$. We choose $M = 1,999$, so that $\frac{1}{2}\alpha(M + 1)$ is an integer.

For each Monte Carlo trial, we compute the size for the naive and asymptotic formula methods using

$$t_{m,g} = (\hat{\gamma}_{m,g} - \gamma_g)/s_{\hat{\gamma}_{m,g}}, \tag{5.32}$$

Table 5.1 Monte Carlo actual size calculations

	$T = 5$			$T = 10$			$T = 20$		
	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3
$N = 50$									
Naive	0.080	0.077	0.088	0.073	0.085	0.085	0.070	0.085	0.084
Asy.	0.069	0.069	0.080	0.070	0.081	0.083	0.070	0.084	0.083
Pairs	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.051
Wild	0.046	0.047	0.046	0.046	0.046	0.046	0.047	0.046	0.046
Avg.	0.062	0.061	0.066	0.060	0.066	0.066	0.059	0.066	0.066
$N = 100$									
Naive	0.067	0.073	0.058	0.057	0.070	0.067	0.067	0.073	0.070
Asy.	0.060	0.069	0.055	0.055	0.069	0.065	0.066	0.072	0.069
Pairs	0.050	0.051	0.050	0.050	0.051	0.051	0.050	0.050	0.050
Wild	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048	0.048
Avg.	0.056	0.060	0.053	0.052	0.059	0.058	0.058	0.061	0.059
$N = 250$									
Naive	0.055	0.064	0.055	0.051	0.061	0.064	0.061	0.053	0.055
Asy.	0.053	0.062	0.051	0.051	0.060	0.062	0.061	0.052	0.054
Pairs	0.050	0.049	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wild	0.050	0.049	0.049	0.049	0.049	0.049	0.049	0.049	0.049
Avg.	0.052	0.056	0.051	0.050	0.055	0.056	0.055	0.051	0.052
$N = 500$									
Naive	0.057	0.050	0.065	0.057	0.060	0.051	0.043	0.056	0.055
Asy.	0.052	0.048	0.062	0.057	0.059	0.050	0.042	0.055	0.053
Pairs	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wild	0.050	0.050	0.049	0.050	0.050	0.049	0.049	0.050	0.050
Avg.	0.052	0.049	0.056	0.053	0.054	0.050	0.046	0.053	0.052
$N = 1,000$									
Naive	0.051	0.054	0.053	0.048	0.045	0.048	0.053	0.047	0.052
Asy.	0.049	0.050	0.050	0.048	0.044	0.048	0.053	0.047	0.051
Pairs	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Wild	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
Avg.	0.050	0.051	0.050	0.049	0.047	0.049	0.051	0.048	0.051

where $s_{\hat{\gamma}_{m,g}}$ for the naive method is the standard error estimator ignoring the existence of first-step random variables and for the asymptotic method is the square root of the gth diagonal element of (5.13). Then, we estimate size over all M observations as the percentage of times that $t_{m,g}$ exceeds the nominal level of $t^*_{\alpha/2} = 1.96$ or is less than the nominal level of $t^*_{1-\alpha/2} = -1.96$ for $\alpha = 0.05$.

Because we are interested in performance under large-N asymptotics, we consider the following cases: $N = 50, 100, 250, 500$ and $1,000$ crossed with $T = 5, 10$ and 20. We compute the actual size, also termed type-I error or rejection probability (RP), and the absolute value of the ERP. Table 5.1 reports actual RPs, while Table 5.2 reports the sum over all parameters of the absolute ERPs.

Table 5.2 Total Monte Carlo ERP

	$T = 5$	$T = 10$	$T = 20$
$N = 50$			
Naive	0.095	0.093	0.088
Asy.	0.067	0.083	0.086
Pairs	0.004	0.004	0.003
Wild	0.011	0.012	0.011
Avg.	0.044	0.048	0.047
$N = 100$			
Naive	0.047	0.043	0.059
Asy.	0.034	0.039	0.056
Pairs	0.001	0.001	0.001
Wild	0.005	0.005	0.005
Avg.	0.022	0.022	0.030
$N = 250$			
Naive	0.023	0.025	0.018
Asy.	0.015	0.022	0.017
Pairs	0.001	0.000	0.000
Wild	0.002	0.002	0.003
Avg.	0.010	0.012	0.009
$N = 500$			
Naive	0.022	0.017	0.017
Asy.	0.016	0.016	0.016
Pairs	0.000	0.000	0.000
Wild	0.001	0.001	0.001
Avg.	0.010	0.009	0.009
$N = 1,000$			
Naive	0.007	0.009	0.008
Asy.	0.002	0.010	0.007
Pairs	0.001	0.000	0.000
Wild	0.001	0.001	0.001
Avg.	0.002	0.005	0.004

Table 5.1 shows that the naive and asymptotic methods seriously over-reject for small values of N, the pairs bootstrap slightly over-rejects, and the wild bootstrap under-rejects. Both bootstrap methods are considerably more accurate than the non-bootstrap methods, with the advantage going to the pairs. For $N = 50$, actual sizes are 0.051 for the pairs and between 0.046 and 0.047 for the wild, but range from 0.069 to 0.088 for the non-bootstrap methods. Thus, the upward bias of the pairs is extremely small (2 %); the downward bias of the wild is larger (6–8 %), but still far smaller than the bias of the conventional methods (at least 40 %).

Increasing N generally improves the accuracy of all methods, but the performance rankings do not change. By N of 250, the bootstrap methods produces the correct, or very close to the correct, size in every case, while the conventional methods still overstate the significance of t-values by at least 10 % in more than

half of the cases. Increasing N to 500 and then 1,000 brings actual and nominal size into alignment in every bootstrap case. However, even when $N = 1,000$, the conventional approaches get the size right in only two cases and are off by as much as 12 % in the others.

The sums of absolute ERPs in Table 5.2 concisely summarize the advantage of the bootstrap methods. For $N > 50$, the pairs bootstrap absolute ERP sum never exceeds 0.001. The wild bootstrap is not as impressive for smaller values of N, but competes with the pairs when N is at least 250. In contrast, the absolute ERP sums of conventional methods are often an order of magnitude larger, even for N as large as 500.

To quantify the roles of N and T in reducing size distortions we regress the ln(actual size) for each estimated coefficient on the logs of N and T by method, where the observations are the 15 combinations of N and T considered in Table 5.3. Asymptotic t-values are reported in parentheses. The results show that increasing N by a given percentage affects size to a considerably greater degree than increasing T by the same percentage. The cross-section dimension effect is also always highly significant, whereas the panel length is never significant at even the 10-% level. This makes sense, because T affects second-step estimation only through its effect on $\hat{\beta}_{FE}$ and its estimated covariance matrix.

Finally we examine the power of the conventional and bootstrap tests, computed using level-adjusted sizes. Since both bootstrap methods always reject less frequently than the conventional methods, the former will appear to have less power. Therefore, we compute power based on level-adjusted $t_{m,g}^*$ values, so that critical values are used for which the actual RP is exactly equal to the nominal RP.[7] These levels, $t_{\alpha/2,g}^*$ and $t_{1-\alpha/2,g}^*$, are the $\alpha/2$ and $1 - \alpha/2$ quantiles of the sorted $t_{m,g}^*$. For each Monte Carlo replication, they are found by first sorting the $t_{m,g}^*$ values from large to small and then taking the $(\alpha/2)(B + 1)$ and $(1 - \alpha/2)(B + 1)$ values for each g. We compute the power curves for γ_3 for each method as the alternative value of γ_3 (denoted as ALT in the figures) is increased from -1 to 1 in increments of .1 by calculating the percentage of the B bootstrap estimates that fall outside the critical region. For the conventional methods, we use the same range of alternative parameter values to compute power as the percentage of M Monte Carlo estimates that falls outside the interval defined by their level-adjusted t-values, computed using the same sorting method just described.

Figures 5.1–5.3 present the power curves for $N = 250, 500$, and 1000, where $T = 5$ throughout. With $N = 250$, the pairs method holds a slight advantage in terms of power. However, with larger values of N, all methods are highly similar. With $N = 250$ the power of all methods is quite low relative to $N = 1000$, where power has risen to approximately .7 for alternatives of -1 and 1.

---

[7]See Davidson and MacKinnon (2006a,b) for further discussion.

Table 5.3 Regressions
explaining ln (actual size)
(number of observations
= 15)

	Intercept	$\ln(N)$	$\ln(T)$	R^2
Naive				
γ_1	2.5426	−0.1277	−0.0369	0.7180
	(15.4134)**	(−5.4644)**	(−0.8293)	
γ_2	2.8238	−0.1766	−0.0173	0.8894
	(22.2146)**	(−9.8096)**	(−0.5066)	
γ_3	2.6794	−0.1550	−0.0073	0.7569
	(14.9573)**	(−6.1099)**	(−0.1514)	
Avg.	2.6819	−0.1531	−0.0205	0.7881
Asy.				
γ_1	2.3006	−0.1128	0.0169	0.6857
	(14.7321)**	(−5.1009)**	(0.4023)	
γ_2	2.6490	−0.1663	0.0211	0.8702
	(20.1923)**	(−8.9502)**	(0.5962)	
γ_3	2.5232	−0.1474	0.0282	0.7411
	(14.1429)**	(−5.8312)**	(0.5873)	
Avg.	2.4910	−0.1422	0.0221	0.7657
Pairs				
γ_1	1.6651	−0.0088	−0.0008	0.8064
	(188.2853)**	(−7.0626)**	(−0.3330)	
γ_2	1.6575	−0.0083	0.0013	0.6901
	(146.0756)**	(−5.1508)**	(0.4405)	
γ_3	1.6622	−0.0088	0.0000	0.8025
	(187.1058)**	(−6.9817)**	(0.0179)	
Avg.	1.6616	−0.0086	0.0002	0.7663
Wild				
γ_1	1.4655	0.0220	−0.0014	0.7810
	(61.5810)**	(6.5375)**	(−0.2115)	
γ_2	1.4651	0.0229	−0.0035	0.8314
	(69.5469)**	(7.6683)**	(−0.6200)	
γ_3	1.4590	0.0228	−0.0009	0.8420
	(72.5098)**	(7.9953)**	(−0.1599)	
Avg.	1.4632	0.0226	−0.0019	0.8181

** indicates significance of the t-values at the .05 level using a
two-tailed asymptotic-t test.

5.6 Conclusions

The primary advantage of panel data is the ability they provide to control for
unobserved heterogeneity or effects that are time-invariant. The fixed-effects (FE)
estimator is by far the most popular technique for exploiting this advantage,
because it makes no assumption about the relationship between the explanatory
variables in the model and the effects. However, a well-known problem with the
FE estimator is that any time-invariant regressor in the model is swept away by the
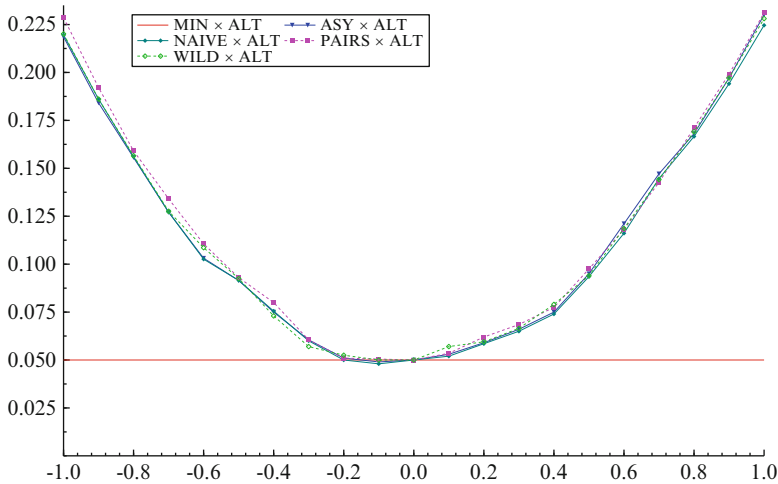data transformation that eliminates the effects. The partial effects of time-invariant

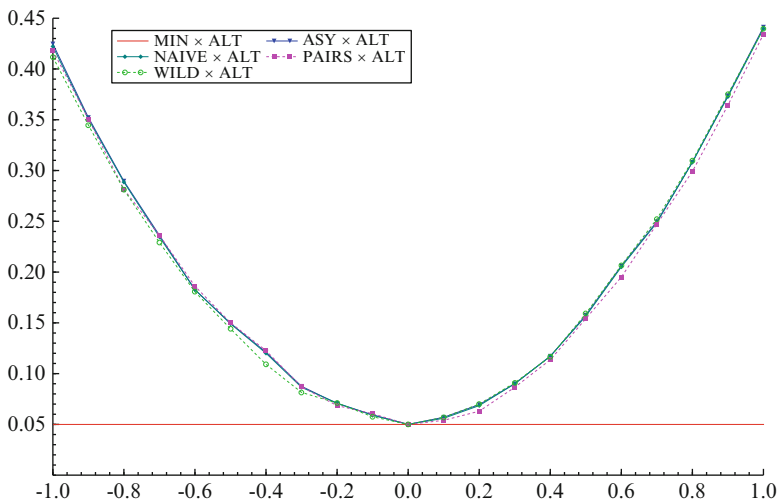**Fig. 5.1** Power vs. H_A for $N = 250$, $t = 5$ for γ_3

Fig. 5.2 Power vs. H_A for $N = 500$, $t = 5$ for γ_3

variables can be estimated in a second-step regression, but this fact is generally overlooked in textbook discussions of panel-data methods.

In this paper, we have shown how to conduct inference on the coefficients of time-invariant variables in linear panel-data models, estimated in a two-step framework. Our estimation framework is rooted in Hausman and Taylor's (1981) "consistent, but inefficient" estimator, albeit under weaker FE assumptions. We derive the asymptotic covariance matrix of the two-step estimator and compare inference based on the asymptotic standard errors with bootstrap alternatives. Bootstrapping
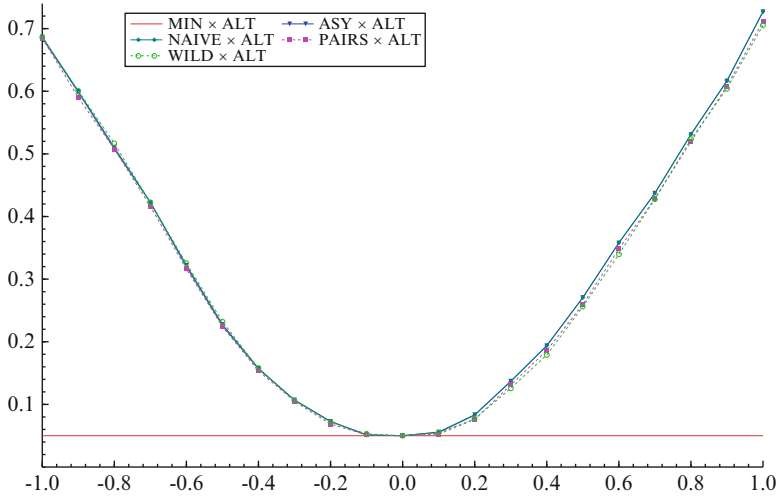
Fig. 5.3 Power vs. H_A for $N = 1000$, $t = 5$ for γ_3

has a natural appeal, because of the complications associated with estimating the asymptotic covariance matrix and the inherent finite-sample bias of the resulting standard errors. We adapt the pairs and wild bootstrap to this two-step problem. Then we prove that both bootstrap coefficient estimators are unbiased, a result that is important since bootstrap ERPs are a function of bias. Using Monte Carlo methods, we compare the size and power of the naive asymptotic estimator, which ignores the first-step estimation error, the correct asymptotic covariance matrix estimator, which does not, and the bootstrap alternatives.

In terms of size, bootstrap methods are the clear winners. For values of N less than 250, the pairs somewhat over-rejects and the wild somewhat under-rejects, with the pairs having a small advantage. The positive ERP of the pairs is 2 %, while the negative ERP of the wild is 6–8 %. In contrast, both are considerably more accurate than the methods based on asymptotic formulae, which are typically biased by more than 40 %. For values of N equal to 250 and larger, the bootstrap methods converge to the correct size and the advantage of the pairs becomes negligible. The correct asymptotic covariance matrix estimator remains somewhat biased even for $N = 1,000$. The pairs bootstrap method slightly out-performs the other methods in terms of power with $N = 250$, although power curves are highly similar with larger values of N.

The Monte Carlo findings are consistent with the results of our analytical examination of the pairs and wild bootstrap procedures. The implication of these results is that both bootstrap ERPs should be very close to zero. The bottom line of both our Monte Carlo exercise and analytical results is that researchers interested in estimating the effects of time-invariant variables in a two-step framework should

rely on bootstrapped standard errors. This conclusion holds particularly strongly in small-N panels like those encountered in cross-state and cross-country studies.

What remains is to consider the advantages to bootstrapping when some of the second-step regression variables may be correlated with the unobserved effects and when the true model errors are heteroskedastic and autocorrelated. Atkinson and Cornwell (2013) extend the work of this paper to that case.

Appendix

Unbiasedness of the Wild First-Step Estimator, $\hat{\boldsymbol{\beta}}_{FE}^{w}$

Lemma 1: *Since ϵ_i is drawn independently and $E(\epsilon_i = 0)$, $E(\boldsymbol{\xi}_i^w|\mathbf{X}_i) = 0$.*

Proof of Lemma 1: From (5.17), $\boldsymbol{\xi}_i^w = \hat{\boldsymbol{\xi}}_i \epsilon_i \vartheta$. Thus, $E(\boldsymbol{\xi}_i^w|\mathbf{X}_i) = E(\hat{\boldsymbol{\xi}}_i \epsilon_i \vartheta|\mathbf{X}_i) = E(\hat{\boldsymbol{\xi}}_i|\mathbf{X}_i)\vartheta E(\epsilon_i|\mathbf{X}_i) = E(\hat{\boldsymbol{\xi}}_i|\mathbf{X}_i)\vartheta E(\epsilon_i) = 0$, since ϵ_i is independent of $\hat{\boldsymbol{\xi}}_i$ and $\mathbf{X}_i$ and in addition $E(\epsilon_i) = 0$ by definition in (5.15). ∎

Theorem 1: *Given the FE conditional-mean assumption in (5.3) and 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 (5.16) as $\mathbf{y}_i^w = \mathbf{X}_i\hat{\beta}_{FE} + \mathbf{z}_i\hat{\gamma}_{FE} + \boldsymbol{\xi}_i^w$ and substituting into (5.18), the first-step wild estimator can be written as

$$\hat{\beta}_{FE}^{w} = \hat{\beta}_{FE} + \left(\sum_i \mathbf{X}_i'\mathbf{Q}_T\mathbf{X}_i\right)^{-1}\sum_i \mathbf{X}_i'\mathbf{Q}_T\boldsymbol{\xi}_i^w, \qquad (5.33)$$

where $\boldsymbol{\xi}_i^w$ is a $(T \times 1)$ vector. Then

$$E(\hat{\beta}_{FE}^{w}|\mathbf{X}_i) = \hat{\beta}_{FE} + \left(\sum_i \mathbf{X}_i'\mathbf{Q}_T\mathbf{X}_i\right)^{-1}\sum_i \mathbf{X}_i'\mathbf{Q}_T E(\boldsymbol{\xi}_i^w|\mathbf{X}_i) = \hat{\beta}_{FE}, \qquad (5.34)$$

using Lemma 1. Further, $E[E(\hat{\beta}_{FE}^{w}|\mathbf{X}_i)] = E(\hat{\beta}_{FE}^{w}) = \hat{\beta}_{FE}$. ∎

Unbiasedness of the Wild Second-Step Estimator, $\hat{\boldsymbol{\gamma}}_{FE}^{w}$

To show that the second-step wild estimator is unbiased, we substitute (5.19) into (5.20) to obtain

$$u_i^w = \bar{y}_i^w - \bar{\mathbf{x}}_i\hat{\beta}_{FE}^{w} - \mathbf{z}_i\hat{\gamma}_{FE}. \qquad (5.35)$$

Now average (5.16) over t to obtain

$$\bar{y}_i^w = \bar{\mathbf{x}}_i \hat{\beta}_{FE} + \mathbf{z}_i \hat{\gamma}_{FE} + \bar{\xi}_i^w \qquad (5.36)$$

and substitute (5.36) into (5.35) to yield

$$u_i^w = \bar{\mathbf{x}}_i (\hat{\beta}_{FE} - \hat{\beta}_{FE}^w) + \bar{\xi}_i^w. \qquad (5.37)$$

Lemma 2: *Since ϵ_i is drawn independently and $E(\epsilon_i = 0)$, $E(\bar{\xi}_{it}^w | \mathbf{z}_i, \bar{\mathbf{x}}_i) = 0$.*

Proof of Lemma 2: Use the definition of ξ_{it}^w in (5.17) and condition on $\mathbf{z}_i, \bar{\mathbf{x}}_i$. Then use the independence of ϵ_i from $\mathbf{z}_i, \bar{\mathbf{x}}_i$. ∎

Theorem 2: *Given Theorem 1 and Lemma 2, the wild second-step estimator, $\hat{\gamma}_{FE}^w$, is unbiased for $\hat{\gamma}_{FE}$.*

Proof of Theorem 2:

$$E(\hat{\gamma}_{FE}^w | \mathbf{z}_i, \bar{\mathbf{x}}_i) = \hat{\gamma}_{FE} + \left(\sum_i \mathbf{z}_i' \mathbf{z}_i \right)^{-1} \left(\sum_i \mathbf{z}_i' E(u_i^w | \mathbf{z}_i, \bar{\mathbf{x}}_i) \right)$$

$$= \hat{\gamma}_{FE} + \left(\sum_i \mathbf{z}_i' \mathbf{z}_i \right)^{-1} \left(\sum_i \mathbf{z}_i' E\{ [\bar{\mathbf{x}}_i (\hat{\beta}_{FE} - \hat{\beta}_{FE}^w) + \bar{\xi}_i^w] | \mathbf{z}_i, \bar{\mathbf{x}}_i \} \right)$$

$$= \hat{\gamma}_{FE}, \qquad (5.38)$$

after substituting from (5.37) for u_i^w and then applying Theorem 1 and Lemma 2. Finally, $E[E(\hat{\gamma}_{FE}^w | \mathbf{z}_i, \bar{\mathbf{x}}_i)] = E(\hat{\gamma}_{FE}^w) = \hat{\gamma}_{FE}$. ∎

Unbiasedness of the Pairs First-Step Estimator, $\hat{\beta}_{FE}^p$

To show the unbiasedness of the pairs first-step estimator, we need (5.3).

Lemma 3: *Given (5.3), $E(\mathbf{Q}_T \hat{\xi}_i | v_i, \mathbf{X}_i) = 0$.*

Proof of Lemma 3: First,

$$E(\mathbf{Q}_T \hat{\xi}_i | v_i, \mathbf{X}_i) = E(\mathbf{M}_i \mathbf{Q}_T \xi_i | v_i, \mathbf{X}_i)$$

$$= E(\mathbf{Q}_T \xi_i | v_i, \mathbf{X}_i) - \mathbf{Q}_T \mathbf{X}_i \left(\sum_i \mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{Q}_T E(\mathbf{Q}_T \xi_i | v_i, \mathbf{X}_i)$$

$$= 7 E\{ \mathbf{Q}_T [(\mathbf{j}_T \otimes c_i) + \mathbf{e}_i] | v_i, \mathbf{X}_i \}$$

$$\quad - \mathbf{Q}_T \mathbf{X}_i \left(\sum_i \mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i \right)^{-1} \mathbf{X}_i' \mathbf{Q}_T E\{ \mathbf{Q}_T [(\mathbf{j}_T \otimes c_i) + \mathbf{e}_i]$$

$$\quad \times | v_i, \mathbf{X}_i \}, \qquad (5.39)$$

using $\mathbf{M}_i = \mathbf{I}_T - \mathbf{Q}_T\mathbf{X}_i\left(\sum_i \mathbf{X}_i'\mathbf{Q}_T\mathbf{X}_i\right)^{-1}\mathbf{X}_i'\mathbf{Q}_T$ and $\boldsymbol{\xi}_i = (\mathbf{j}_T \otimes c_i) + \mathbf{e}_i$. Then, using (5.3) and the fact that $\mathbf{Q}_T$ eliminates c_i completes the proof. ∎

Theorem 3: *Given Lemma 3, the bootstrap pairs first-step estimator, $\hat{\beta}_{FE}^p$, is unbiased for $\hat{\beta}_{FE}$.*

Proof of Theorem 3: Substitute $\mathbf{Q}_T\mathbf{y}_i$ in (5.28) and take expectations. Then

$$E(\hat{\beta}_{FE}^p|v_i,\mathbf{X}_i) = \hat{\beta}_{FE} + \left(\sum_i v_i\mathbf{X}_i'\mathbf{Q}_T\mathbf{X}_i\right)^{-1}\sum_i v_i\mathbf{X}_i'\mathbf{Q}_T\,E(\mathbf{Q}_T\hat{\boldsymbol{\xi}}_i|v_i,\mathbf{X}_i)$$

$$= \hat{\beta}_{FE}, \tag{5.40}$$

using Lemma 3. ∎

Unbiasedness of the Pairs Second-Step Estimator, $\hat{\gamma}_{FE}^p$

Theorem 4: *Given (5.3), (5.4), and Theorem 3, the pairs second-step estimator, $\hat{\gamma}_{FE}^p$, is unbiased for $\hat{\gamma}_{FE}$.*

Proof of Theorem 4: Substituting (5.29) into (5.30) we obtain

$$\hat{\gamma}_{FE}^p = \hat{\gamma}_{FE} + \left(\sum_i v_i\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\left(\sum_i v_i\mathbf{z}_i'\hat{u}_i\right). \tag{5.41}$$

We can relate $\hat{u}_i$ to u_i as follows:

$$\hat{u}_i = u_i - \mathbf{z}_i\left(\sum_i \mathbf{z}_i'\mathbf{z}_i\right)^{-1}\sum_i \mathbf{z}_i'u_i. \tag{5.42}$$

Then substitute (5.42) into (5.41) to obtain

$$\hat{\gamma}_{FE}^p = \hat{\gamma}_{FE} + \left(\sum_i v_i\mathbf{z}_i'\mathbf{z}_i\right)^{-1}\sum_i v_i\mathbf{z}_i'u_i - \left(\sum_i \mathbf{z}_i'\mathbf{z}_i\right)^{-1}\sum_i \mathbf{z}_i'u_i \tag{5.43}$$

Conditioning on $(\mathbf{z}_i, \bar{\mathbf{x}}_i, v_i)$, we use (5.8) and take the expectation of both sides to obtain

$$E(u_i|\mathbf{z}_i, \bar{\mathbf{x}}_i, v_i) = E(\bar{\xi}_i|\mathbf{z}_i, v_i) + \bar{\mathbf{x}}_i\,E[(\hat{\beta}_{FE} - \beta)|\mathbf{z}_i, \bar{\mathbf{x}}_i, v_i]. \tag{5.44}$$

The first term on the right-hand-side of (5.44) is zero due to (5.3) and (5.4), while $\hat{\beta}_{FE}$ is unbiased for β from Theorem 3. ∎

References

Agee MD, Atkinson SE, Crocker TD (2009) Multi-input and multi-output estimation of the efficiency of child health production: a two-step panel data approach. South Econ J 75:909–927

Agee MD, Atkinson SE, Crocker TD (2012) Child maturation, time-invariant, and time-varying inputs: their interaction in the production of human capital. J Product Anal 35:29–44

Arellano M (1987) Computing robust standard errors for within-groups estimators, Oxf Bull Econ Stat 49:431–34

Atkinson SE, Cornwell C (2013) Inference in two-step panel data models with weak instruments and time-invariant regressors: bootstrap versus analytic estimators. Working paper, University of Georgia

Breusch T, Ward MB, Nguyen HTM, Kompas T (2011) On the fixed-effects vector decomposition. Polit Anal 19:123–134

Cameron AC, Trivedi, PK (2005) Microeconometrics: methods and applications. Cambridge University Press, New York

Davidson R, Flachaire, E (2008) The wild bootstrap, tamed at last. J Econom 146:162–169

Davidson R, MacKinnon, JG (1999) Size distortion of bootstrap tests. Econom Theory 15:361–376

Davidson R, MacKinnon, JG (2006) The power of bootstrap and asymptotic tests. J Econom 133:421–441

Davidson R, MacKinnon, JG (2006) Bootstrap methods in econometrics. Chapter 23 In: Mills TC, Patterson KD (ed) Palgrave handbooks of econometrics: Volume I Econometric Theory. Palgrave Macmillan, Basingstoke, pp 812–838

Flachaire E (2005) Bootstrapping Heteroskedastic regression models: wild bootsrap vs. pairs bootstrap. Comput Stat Data Anal 2005:361–376

Greene WH (2011) Fixed effects vector decomposition: A magical solution to the problem of time-invariant variables in fixed effects models? Polit Anal 19:135–146

Hausman JA, Taylor W (1981) Panel data and unobservable individual effects. Econometrica 49:1377–1399

Horowitz J (2001) The bootstrap. In: Heckman JJ, Leamer E (ed) Handbook of econometrics, vol 5, ch. 52. North-Holland, New York

Kapetanios G (2008) A bootstrap procedure for panel data sets with many cross-sectional units. Econom J 11:377–395

Knack S (1993) The voter participation effects of selecting jurors from registration. J Law Econ 36:99–114

MacKinnon JG (2002) Bootstrap inference in econometrics. Can J Econ 35:615–645

Murphy KM, Topel RH, (1985) Estimation and inference in two-step econometric models. J Bus Econ Stat 3:88–97

Plümper T, Troeger VE (2007) Efficient estimation of time-invariant and rarely changing variables in finite sample panel analysis with unit fixed effects. Polit Anal 15:124–139

Wooldridge JM (2010) Econometric analysis of cross section and panel data, 2nd edn. MIT: Cambridge