Panel Bootstrap Standard Errors

The **bootstrap method** provides an alternative way to obtain panel-robust standard errors. The key assumption is that observations are independent over i, so one does a bootstrap pairs procedure that resamples **with replacement over** i and uses all observed time periods for a given individual. For data $\{(\mathbf{y}_i, \mathbf{X}_i), i = 1, \ldots, N\}$ this yields B pseudo-samples and for each **pseudo-sample** one performs OLS regression of $\widetilde{y}_{it}$ on $\widetilde{\mathbf{w}}_{it}$, yielding B estimates $\widehat{\boldsymbol{\theta}}_b$, $b = 1, \ldots, B$.

The **panel bootstrap estimate** of the variance matrix is then

$$\widehat{\mathrm{V}}_{\mathrm{Boot}}[\widehat{\boldsymbol{\theta}}] = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\boldsymbol{\theta}}_b - \overline{\widehat{\boldsymbol{\theta}}}\right)\left(\widehat{\boldsymbol{\theta}}_b - \overline{\widehat{\boldsymbol{\theta}}}\right)', \tag{21.14}$$

where $\overline{\widehat{\boldsymbol{\theta}}} = B^{-1}\sum_b \widehat{\boldsymbol{\theta}}_b$. This bootstrap provides **no asymptotic refinement** (see Section 11.2.2). Given independence over i the estimate is consistent as $N \to \infty$. It is asymptotically equivalent to the estimate (21.13), just as in the cross-section case bootstrap pairs are asymptotically equivalent to White's heteroskedastic consistent estimate. This bootstrap does not offer an asymptotic refinement though bootstrap with asymptotic refinement is possible (see Section 11.6.2).

This bootstrap method can be applied to any panel estimator that relies on independence over i and $N \to \infty$, including the pooled feasible GLS estimators of Section 21.5.2 for short panels. The key is to resample over i only, and not over both i and t.

Discussion

The importance of correcting standard errors for serial correlation in errors at the individual level cannot be overemphasized. Computer packages currently do not automatically do this. Bertrand, Duflo, and Mullainathan (2004) illustrate the resulting downward bias in standard error computation, in the context of difference-in-differences estimation (see Section 22.6). They find that the panel-robust and panel bootstrap methods work well, even though in their application with state-year data N (the number of states) is relatively small whereas the asymptotic theory uses $N \to \infty$.

The following example (see Table 21.2) also shows the importance of correcting standard errors for any error serial correlation and autocorrelation.

21.3. Linear Panel Example: Hours and Wages

An important issue in labor economics is the responsiveness of labor supply to wages. The standard textbook model of labor supply suggests that for people already working the effect of a wage increase on labor supply is ambiguous, with an income effect pushing in the direction of less work offsetting a substitution effect in the direction of more work.

Cross-section analysis for adult males finds a relatively small positive response to hours worked. However, it is possible that this association is spurious, merely reflecting a greater unobserved desire to work being positively associated with higher wages.

Panel data analysis can control for this, under the assumption that the unobserved desire to work is time-invariant. For example, the within estimator does so by measuring the extent to which an individual works above-average (or below-average) hours in periods with above-average (or below-average) wages.

The data on 532 males for each of the 10 years from 1979 to 1988 come from Ziliak (1997). The variable of interest is lnhrs, the natural logarithm of annual hours worked. The single explanatory variable is lnwg, the natural logarithm of hourly wage. We consider the regression model

$$\text{lnhrs}_{it} = \alpha_i + \beta \text{lnwg}_{it} + \varepsilon_{it},$$

where the individual-specific effect α_i is simplified to α in some models and β measures the wage elasticity of labor supply. The error term ε_{it} is assumed to be independent over i, but it may be correlated over t for given i. As noted we expect β, the labor supply elasticity, to be small and positive.

Ziliak (1997) additionally included a quadratic in age, number of children, and an indicator variable for bad health. These regressors and year dummies make relatively small difference to the estimate of β and its standard error, and for simplicity they are omitted here. In Chapter 22 we consider more general models that permit lnwg to be endogenous and permit lags of lnhrs to appear as a regressor.

21.3.1. Data Summary

For the 5,320 observations, the sample means of lnhrs and lnwg are respectively 7.66 and 2.61, implying geometric means of 2,120 hours and $13.60 per hour. The sample standard deviations are respectively 0.29 and 0.43, indicating considerably greater variability in percentage terms in wages rather than hours.

For panel data it is useful to know whether variability is mostly across individuals or across time. The total variation of a series x_{it} around its grand mean $\bar{x}$ can be decomposed as

$$\begin{aligned}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})^2 &= \sum_{i=1}^{N}\sum_{t=1}^{T}[(x_{it}-\bar{x}_i)+(\bar{x}_i-\bar{x})]^2 \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2 + \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{x}_i-\bar{x})^2,\end{aligned}$$

as the cross-product term sums to zero. In words, the total sum of squares equals the **within sum of squares** plus the **between sum of squares**. This leads to **within standard deviation** s_{W} and **between standard deviation** s_{B}, where

$$s_{\mathrm{W}}^2 = \frac{1}{NT-N}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2$$

and

$$s_{\mathrm{B}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})^2.$$

Table 21.2. *Hours and Wages: Standard Linear Panel Model Estimators*[a]

	POLS	Between	Within	First Diff	RE–GLS	RE–MLE
α	7.442	7.483	7.220	.001	7.346	7.346
β	.083	.067	.168	.109	.119	.120
Robust se[b]	(.030)	(.024)	(.085)	(.084)	(.051)	(.052)
Boot se	[.030]	[.019]	[.084]	[.083]	[.056]	[.058]
Default se	{.009}	{.020}	{.019}	{.021}	{.014}	{.014}
R^2	.015	.021	.016	.008	.014	.014
RMSE	.283	.177	.233	.296	.233	.233
RSS	427.225	0.363	259.398	417.944	288.860	288.612
TSS	433.831	17.015	263.677	420.223	293.023	292.773
σ_α	.000		.181		.161	.162
σ_ε	.283		.232		.233	.233
λ	0.000	–	1.000	–	.585	.586
N	5320	532	5320	4788	5320	5320

[a] Shown are pooled OLS (POLS), between, within, first-differences, random effects (RE) GLS and MLE linear panel regression of lnhrs on lnwg. Standard errors for the slope coefficients are panel robust in parentheses, panel bootstrap in square brackets, and default estimates that assume iid errors in curly braces. The R^2, root mean square error (RMSE), residual sum of squares (RSS), total sum of squares (TSS), and sample size come from the appropriate regression given in Section 21.2. The parameter λ is defined after (21.11).

[b] se, standard error.

The within and between sample standard deviations are, respectively, 0.22 and 0.18 for lnhrs and 0.19 and 0.39 for lnwg. The larger total variation in wages compared to hours is therefore due to between individual variation being much higher for wages. Within individuals the variation is actually somewhat smaller for wages than it is for hours.

21.3.2. Comparison of Panel Data Estimators

Table 21.2 summarizes results from application of the standard panel estimators defined in Section 21.2.2 to these data, along with three different estimates of the standard errors. As detailed in the following, statistical inference should use either the panel-robust standard error or the panel bootstrap standard error.

Slope Parameter Estimates

The estimate of the slope parameter β differs across the different estimation methods. The between estimate that uses only cross-section variation is less than the pooled OLS estimate. The within or fixed effects estimate of 0.168 is much higher than the pooled OLS estimate of 0.083 and is borderline statistically significant using a two-tailed test at 5% and standard error estimate of 0.084 or 0.085. The first-differences estimate of 0.109 is also higher than that of pooled OLS but is considerably less than the within estimate, which also uses only time-series variation. The RE estimates of 0.119 or 0.120 lie between the between and within estimates. This is expected, as RE estimates can be shown to be a **weighted average of between and within estimates**. The two RE estimates are very close to each other as here the estimates of the variances σ_α^2 and σ_ε^2 are similar, leading to similar values $\widehat{\lambda} = 0.585$ and $\widehat{\lambda} = 0.586$ in the regression (21.10). The RE estimates are surprisingly less efficient than the pooled OLS estimates, a sign that the RE model fails to model the error correlation well.

Which estimates are preferred? The within and first-difference estimators are consistent under all models (pooled, RE, and FE) whereas the other estimators are inconsistent under the fixed effects model. The most robust estimates are therefore the within or first-differences estimates of 0.168 or 0.109.

There is, however, an efficiency loss in using these more robust estimators, with standard errors of 0.83 to 0.85 that are much larger than those from pooled OLS and RE estimates. A formal Hausman test (see Section 21.4.3 for details and discussion) can be used to test whether or not the individual effects are fixed. Given the relative imprecision of estimation in this example, the Hausman test does not reject the null hypothesis of random effects, despite the large difference between FE and RE estimates. So the more efficient random effects estimates could be used here. Another advantage of random effects estimation is that it permits estimation of the coefficients of time-invariant estimators.

Standard Error Estimation

We now turn to comparison of the standard error estimates. From Section 21.2.3, inference should be based on panel-robust standard errors that permit errors to be correlated over time for a given individual and to have variances and covariances that differ across individuals. Also, as detailed in later sections, the standard errors for estimators based on deviations from means, such as (21.8) and (21.10), need to account for loss of $N + K$ rather than K degrees of freedom.

The first standard error estimate is computed by the panel-robust method given in (21.13), and the second is computed by the panel bootstrap given in (21.14) with 500 replications. For brevity these estimates are called panel robust, though they are additionally robust to heteroskedasticity. The two estimates are very close, aside from the random effects models where the panel-robust standard errors are underestimated because they are computed for the regression (21.10), which ignores estimation error in $\widehat{\lambda}$.

The third standard error estimate is the standard default computer output that is based on the assumption of iid errors. In this example the correctly estimated standard errors are a remarkable three to four times as large as the default standard errors. The one exception is the between estimator, an estimator with standard errors that need only correction for heteroskedasticity since it uses only cross-section variation.

For example, for the pooled OLS estimator of β the default standard error is 0.09, leading to incorrect t-statistic of 9.07. The panel-robust standard error is a much larger 0.30, leading to correct t-statistic of a much smaller 2.83. Default standard errors assume independence of model errors over t for given i when in practice they are likely to be positively correlated. This erroneous assumption overestimates the benefit of additional time periods, leading to downward bias in standard errors (see Section 21.5.4). Additionally, ignoring heteroskedasticity in errors also leads to bias, though this bias could be in either direction. For these data a failure to control for heteroskedasticity also imparts a large downward bias: The standard error of $\widehat{\beta}_{\text{POLS}}$ controlling for heteroskedasticity, but not for correlation over t for given i, is 0.020. For other data, correction for heteroskedasticity is usually much less important than the correction for panel correlation.

For the within and between estimators inclusion of the term α_i should control for some of the correlation in the error across time for a given individual. For these data, however, the differences between panel-robust and nonrobust standard errors remain large, in part because of failure to additionally control for heteroskedasticity.

Clearly panel-robust standard errors should be used.

21.3.3. Graphical Analysis

It is insightful to perform a graphical comparison of overall, between, and fixed effects (within or first-differences) regressions. Such plots are rarely performed in panel data regression, but they are easily applied here as there is only one regressor.

All plots include a nonparametric regression curve using the Lowess smoother (see Section 9.6.2) and a linear regression curve that corresponds to the estimates given in Table 21.2.

Figure 21.1 plots lnhrs against lnwg for all firms in all years (5,320 observations). The plot suggests a positive relationship, roughly linear except at the extreme ends, and from Table 21.2 the line has slope 0.083 with a low $R^2 = 0.015$.

The between estimator (21.7) regresses $\bar{y}_i$ on $\bar{x}_i$. The corresponding plot for the lnhrs–lnwg data is given in Figure 21.2 and again shows a positive relationship.

The within or fixed effects estimator (21.8) regresses $(y_{it} - \bar{y}_i)$ on $(x_{it} - \bar{x}_i)$. Figure 21.3 gives the related plot of $(y_{it} - \bar{y}_i + \bar{y})$ on $(x_{it} - \bar{x}_i + \bar{x})$, where $\bar{y} = N^{-1} \sum_i \bar{y}_i$ and $\bar{x} = N^{-1} \sum_i \bar{x}_i$ are the grand means of y and x. Comparison with Figure 21.1 shows that differencing the individual mean leads to a considerable decrease in the range of variability in lnwg, with less of a decrease in the variability of lnhrs.

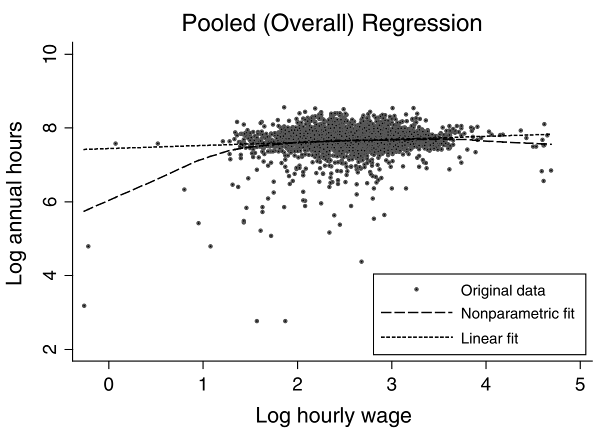

Figure 21.1: Hours and wages: pooled (overall) regression. Natural logarithm of annual hours worked plotted against natural logarithm of hourly wage. Data for 532 U.S. males for each of the ten years 1979–88.

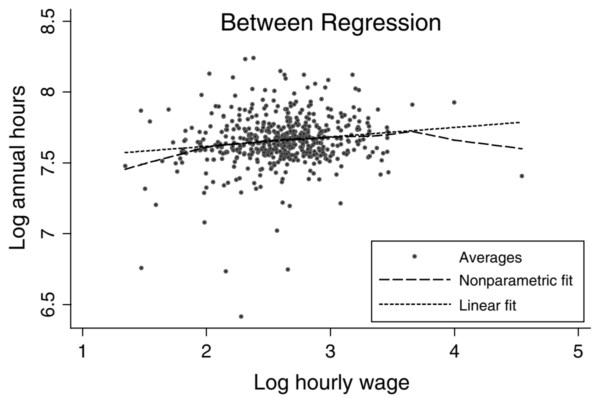

Figure 21.2: Hours and wages: between regression. Ten-year average of log hours plotted against ten-year average of log wage for 532 men. Same sample as Figure 21.1.

The slope does appear steeper than that for pooled OLS, and from Table 21.2 the slope increased from 0.083 to 0.168.

The first-differences estimator (21.9) regresses $(y_{it} - y_{i,t-1})$ on $(x_{it} - x_{i,t-1})$. The corresponding plot for the lnhrs – lnwg data is given in Figure 21.4. The figure is qualitatively similar to Figure 21.3.

The conclusion of the preceding analysis is that there is greater response to wage changes using time-series variation than using cross-section variation.

21.3.4. Residual Analysis

It is instructive to consider the autocorrelation patterns of the data and of residuals. For example, for residuals $\widehat{u}_{it} = y_{it} - \widehat{y}_{it}$ the autocorrelation between period s and period t is calculated as $\widehat{\rho}_{st} = c_{st}/\sqrt{c_{ss}c_{tt}}$, $s, t = 1, \ldots, T$, where the covariance estimate $c_{st} = (N-1)^{-1}\sum_i(\widehat{u}_{it} - \overline{\widehat{u}}_t)(\widehat{u}_{is} - \overline{\widehat{u}}_s)$ and $\overline{\widehat{u}}_t = N^{-1}\sum_i \widehat{u}_{it}$.

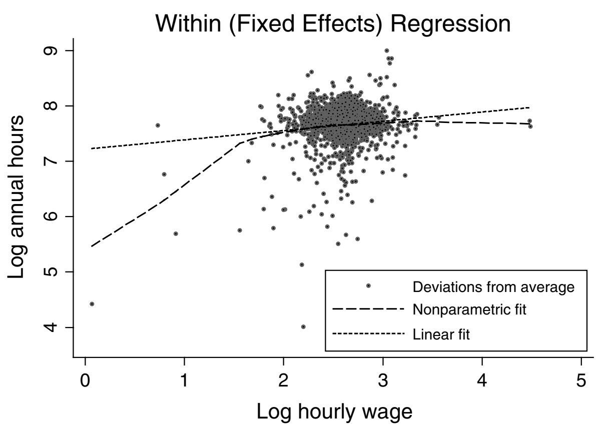

Figure 21.3: Hours and wages: within (fixed effects) regression. Deviation of log hours from ten-year average plotted against deviation of log wage from ten-year average using ten years of data for 532 men. Same sample as Figure 21.1.

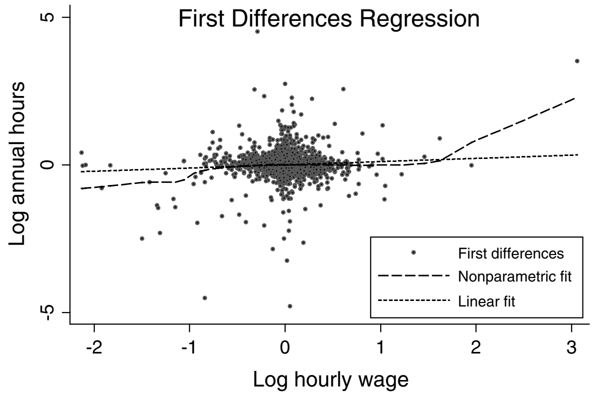

Figure 21.4: Hours and wages: first differences regression. First difference of log hours plotted against first difference of log wage using ten years of data for 532 men. Same sample as Figure 21.1.

Table 21.3 gives the residual autocorrelations after pooled OLS regression of lnhrs on lnwg. The autocorrelations generally lie between 0.2 and 0.4 for data two to nine periods apart. The decay rate is very slow, and the autocorrelations appear closer to a random effects model that assumes that $\mathrm{Cor}[u_{it}, u_{is}]$ is constant for $t \neq s$ than to an AR(1) model that has exponential decay.

The autocorrelations for lnhrs before regression are very similar to those given in Table 21.3, since $\widehat{u}_{it} \simeq y_{it}$ as evident from the poor explanatory power of pooled OLS with $R^2 = 0.015$. The autocorrelations for the regressor lnwg, not tabulated here, are much higher, ranging from approximately 0.9 at one lag, to 0.7 at nine lags.

The correlations of the residuals from the within regression are given in Table 21.4. If the original errors ε_{it} in (21.3) are iid then it can be shown that the transformed errors $\varepsilon_{it} - \bar{\varepsilon}_i$ have autocorrelations at all lags equal to $-1/(T-1) = -0.11$. There is some departure from this here, particularly for the first lag, which is always positive.

Table 21.3. *Hours and Wages: Autocorrelations of Pooled OLS Residuals*[a]

	u79	**u80**	**u81**	**u82**	**u83**	**u84**	**u85**	**u86**	**u87**	**u88**
upols79	1.00									
upols80	.33	1.00								
upols81	.44	.40	1.00							
upols82	.30	.31	.57	1.00						
upols83	.21	.23	.37	.47	1.00					
upols84	.20	.23	.32	.34	.64	1.00				
upols85	.24	.32	.41	.35	.39	.58	1.00			
upols86	.20	.19	.28	.25	.31	.35	.40	1.00		
upols87	.20	.32	.33	.29	.31	.34	.39	.35	1.00	
upols88	.16	.25	.30	.26	.21	.25	.34	.55	.53	1.00

[a] Note: Autocorrelations of residuals are from pooled OLS regression of lnhrs on lnwg for 532 men in 10 years. The autocorrelations die slowly.