More commonly, the Hausman test is automated by software packages to contrast the complete set of common estimates. That is, we carry out a test of a joint hypothesis comparing all the coefficients in Table 15.7, except the intercept, to the corresponding estimates in Table 15.9. If there is no correlation between the error component u_i and the values of x_{kit}, then the six variables common to the two tables (*EXPER, EXPER*^{2}*, TENURE, TENURE*^{2}*, SOUTH*, and *UNION*) will have coefficient estimates with similar magnitudes. The Hausman contrast[13] test jointly checks how close the differences between the pairs of coefficients are to zero. The calculated value of this chi-square statistic is 20.73. We are comparing the values of six coefficients, and the test statistic has an asymptotic chi-square distribution with six degrees of freedom. The 5% critical value for this distribution is 12.592, and the 1% critical value is 16.812. On the basis of the joint test, we reject the null hypothesis that the difference between the estimators is zero even at the 1% level of significance. Again this implies that we should use the fixed effects estimator in this case, or revisit the specification of our model.

The form of the Hausman test in (15.37) and its χ^2 equivalent are not valid for cluster-robust standard errors, because under these more general assumptions, it is no longer true that $\text{var}\left(b_{FE,k} - b_{RE,k}\right) = \text{var}\left(b_{FE,k}\right) - \text{var}\left(b_{RE,k}\right)$.

15.6 The Hausman-Taylor Estimator

The outcome from our comparison of the fixed and random effects estimates of the wage equation poses a dilemma. Correlation between the explanatory variables and the random effects means that the random effects estimator will be inconsistent. We can overcome the inconsistency problem by using the fixed effects estimator, but doing so means that we can no longer estimate the effects of the time invariant variables *EDUC* and *BLACK*. The wage return to an extra year of education, and whether or not there is wage discrimination on the basis of race, might be two important questions that we would like to answer.

To solve this dilemma we ask: How did we cope with the endogeneity problem in Chapter 10? We did so by using instrumental variable estimation. Variables known as "instruments," which are correlated with the endogenous variables but uncorrelated with the equation error, were introduced, leading to an instrumental variables estimator that has the desirable property of consistency. The **Hausman-Taylor estimator** is an instrumental variables estimator applied to the random effects model to overcome the problem of inconsistency caused by correlation between the random effects and some of the explanatory variables. To explain how it works consider the regression model

$$y_{it} = \beta_1 + \beta_2 x_{it,exog} + \beta_3 x_{it,endog} + \beta_3 w_{i,exog} + \beta_4 w_{i,endog} + u_i + e_{it} \tag{15.38}$$

We have divided the explanatory variables into four categories:

$x_{it,exog}$: exogenous variables that vary over time and individuals

$x_{it,endog}$: endogenous variables that vary over time and individuals

$w_{i,exog}$: time-invariant exogenous variables

$w_{i,endog}$: time-invariant endogenous variables

---

[13] Details of the joint test are beyond the scope of this book. For a very advanced reference that contains a careful exposition of the *t*-test, the chi-square test, and a regression-based alternative that may be preferable, see *Econometric Analysis of Cross Section and Panel Data,* 2nd Edition, by Jeffrey Wooldridge (MIT, 2010), p. 328.

Equation (15.38) is written as if there is one variable of each type, but in practice there could be more than one. For the Hausman-Taylor estimator to work the number of exogenous time-varying variables $(x_{it,exog})$ must be at least as great as the number of endogenous time-invariant variables $(w_{i,endog})$.

For the wage equation we will make the following assumptions

$$\begin{aligned}
x_{it,exog} &= \{EXPER,\ EXPER^2,\ TENURE,\ TENURE^2,\ UNION\}\\
x_{it,endog} &= \{SOUTH\}\\
w_{i,exog} &= \{BLACK\}\\
w_{i,endog} &= \{EDUC\}
\end{aligned}$$

The variable *EDUC* is chosen as an endogenous variable on the grounds that it will be correlated with personal attributes such as ability and perseverance. It is less clear why *SOUTH* should be endogenous, but we include it as endogenous because its fixed and random effects estimates were vastly different. Perhaps those living in the South have special attributes. The remaining variables—experience, tenure, *UNION*, and *BLACK*—are assumed uncorrelated with the random effects.

Following Chapter 10, we need instruments for $x_{it,endog}$ and $w_{i,endog}$. Since the fixed effects transformation $\tilde{x}_{it,endog} = x_{it,endog} - \bar{x}_{i,endog}$ eliminates correlation with u_i, we have $\tilde{x}_{it,endog}$ as a suitable instrument for $x_{it,endog}$. Also, the variables $\bar{x}_{i,exog}$ are suitable instruments for $w_{i,endog}$. The exogenous variables in (15.38) can be viewed as instruments for themselves, making the complete instrument set $x_{it,exog}$, $\tilde{x}_{it,endog}$, $w_{i,exog}$, $\bar{x}_{i,exog}$. Hausman and Taylor modify this set slightly using $\tilde{x}_{it,exog}, \tilde{x}_{it,endog}, w_{i,exog}, \bar{x}_{i,exog}$ which can be shown to yield the same results. Their estimator is applied to the transformed generalized least squares model from (15.31)

$$y^*_{it} = \beta_1 + \beta_2 x^*_{it,exog} + \beta_3 x^*_{it,endog} + \beta_3 w^*_{i,exog} + \beta_4 w^*_{i,endog} + v^*_{it} \tag{15.39}$$

where, for example, $y^*_{it} = y_{it} - \hat{\alpha}\bar{y}_i$, and $\hat{\alpha} = 1 - \hat{\sigma}_e / \sqrt{T\hat{\sigma}^2_u + \hat{\sigma}^2_e}$. The estimate $\hat{\sigma}^2_e$ is obtained from fixed-effects residuals; an auxiliary instrumental variables regression[14] is needed to find $\hat{\sigma}^2_u$.

Estimates for the wage equation are presented in Table 15.10. Compared to the random effects estimates, there has been a dramatic increase in the estimated wage returns to education from 7.3% to 17%. The estimated effects for experience and tenure are similar. The wage reduction for *BLACK* is estimated as 3.6% rather than 11.7%, and the penalty for being in the *SOUTH* is also less, 3.1% instead of 8.2%. The instrumental-variable standard errors are mostly larger, particularly for *EDUC* and *BLACK* where the biggest changes in estimates have been observed. Which set of estimates is better will depend on how successful we have been at making the partition into exogenous and endogenous variables in (15.38), and whether the gain from having consistent estimates is sufficiently large to compensate for the increased variance of the instrumental variables estimators.

15.7 Sets of Regression Equations

So far in this chapter, we have considered methods for estimating panel data models when the panel is short and wide: N is large and T is small. We now turn to a model and estimation

[14] Details can be found in the advanced book, Jeffrey Wooldridge (2010), *Econometric Analysis of Cross-Section and Panel Data*, 2nd Edition, MIT Press, pp. 358–361.

Table 15.10 Hausman-Taylor Estimates of Wage Equation

Variable	Coefficient	Std. Error	t-value	p-value
C	−0.75077	0.58624	−1.28	0.200
$EDUC$	0.17051	0.04446	3.83	0.000
$EXPER$	0.03991	0.00647	6.16	0.000
$EXPER^2$	−0.00039	0.00027	−1.46	0.144
$TENURE$	0.01433	0.00316	4.53	0.000
$TENURE^2$	−0.00085	0.00020	−4.32	0.000
$BLACK$	−0.03591	0.06007	−0.60	0.550
$SOUTH$	−0.03171	0.03485	−0.91	0.363
$UNION$	0.07197	0.01345	5.35	0.000

procedures for a panel that is long and narrow: T is large relative to N. If the number of time-series observations is sufficiently large, and N is small, we can estimate separate equations for each individual. These separate equations can be specified as

$$y_{it} = \beta_{1i} + \beta_{2i}x_{2it} + \beta_{3i}x_{3it} + e_{it} \tag{15.40}$$

The i subscript on the β's means that they can be different for each individual. Thus, this model can be used to represent N different equations, one for each individual. There are T observations on each of the N equations. For the short and wide panel considered in earlier sections, T was not sufficiently large to estimate separate equations for each individual. We assumed $\beta_{2i} = \beta_2$ and $\beta_{3i} = \beta_3$; the slope coefficients were the same for all individuals, but the intercept β_{1i} was allowed to vary.

15.7.1 Grunfeld's Investment Data

The example we use for this section is an old but very famous one. The factors affecting the investment behavior by firms was studied by Grunfeld[15] using a panel of data. His example and data, which is simply referred to in the literature as "the Grunfeld data," have been used many times to illustrate the issues involved in modeling panel data.

Investment demand is the purchase of durable goods by both households and firms. In terms of total spending, investment spending is the volatile component. Therefore, understanding what determines investment is crucial to understanding the sources of fluctuations in aggregate demand. In addition, a firm's net fixed investment, which is the flow of additions to capital stock or replacements for worn-out capital, is important because it determines the future value of the capital stock and thus affects future labor productivity and aggregate supply.

There are several interesting and elaborate theories that seek to describe the determinants of the investment process for the firm. Most of these theories evolve to the conclusion that perceived profit opportunities (expected profits or present discounted value of future

[15] Grunfeld, Y. (1958) *The Determinants of Corporate Investment*. Unpublished Ph.D. thesis, Department of Economics, University of Chicago. Grunfeld, Y. and Z. Griliches (1960) "Is Aggregation Necessarily Bad?" *Review of Economics and Statistics*, 42, 1–13.