Chapter 6

Linear estimation and testing

The previous chapter outlined various estimators that could be extracted from a properly constituted X'X matrix. These estimators have been implemented in a GAUSS program called XPReg.GP, and this chapter describes briefly the development and working of the program, the extension for instrumental variables, and the hypothesis tests implemented.

6.1XPReg.GP: Estimation basics

6.1.1Development of the software

The regression program was first implemented in Autumn 1991, and has passed through various stages reflecting the uses made of it. It was designed originally as a simple (and necessary) tool to obtain statistics and estimates from cross-products, and although in the subsequent development of the model the coding, capabilities and features have changed enormously, the basic principles have changed relatively little.

The five stages of the model have been, roughly,

Version 1Autumn 1991Simple OLS cross-section

Version 2 Spring 1992Time-specific intercepts; analysis of covariance; residual variance analysis

Version 3Autumn 1992Fixed-effects (balanced panels only)

Version 4Spring 1993Instrumental variables

Version 5Spring 1994Proper fixed-effects model; time-differencing; instrumental variable and joint significance tests; complete internal rewrite

The original fixed-effects model was a relatively simple extension to deal with individual heterogeneity in the manner of section 5.2. However, a flaw in the mathematics meant that the program only dealt with heterogeneity properly for balanced panels. For unbalanced panels the "fixed effects" model merely applied a meaningless transformation to the data; this was corrected in Version 5.

The original extraction software was completed and used before the first version of the regression program (the early X'X matrices being used as cross-tabulations), and has remained essentially the same although the speed and efficiency of the programs have improved[1]. In line with Version 5 of the program the extraction software was completely rewritten, the intention being to integrate more fully the complete process from collection of data to analysis of results. Although extraction and analysis are separate tasks, the choice of models available is obviously dependent upon the type of cross-product matrix created, and the type of matrix created depends on the model to be estimated.

In early 1994, the University of Stirling agreed with the Department of Employment to provide extraction software enabling general access to the NES in the form of cross-product matrices. Requests to the DE would list the data to be collected, interpreting software would produce extraction software, and the extraction software would produce an X'X matrix to be returned to the researcher. The researcher could then analyse the data using some provided software or his own tools. Under the initial specification the software was designed to produce input for the simple instrumental-variables version of XPReg.GP (Version 4) and a basic working suite of programs was developed. However, the opportunity was taken to reconsider completely the type and nature of potential models, with a view to implementing those that were both feasible and desirable in the context of an X'X dataset. The result was Version 2 of the extraction software and Version 5 of the analysis program.

[1] The original extraction routine was written by Elizabeth Roberts at the University of Stirling.

The extraction software is documented elsewhere, in the NES user instructions to be issued by the DE. This software is the intellectual property of the DE. Stirling University retains control over the regression program code and distribution. Some more basic analysis programs have also been developed and, with a restricted version of XPReg.GP, given to the DE for distribution to users.

It should be noted that, while the extraction software is to some extent specific to the NES, the analytical software is independent of the source of the data. A properly constructed cross-product matrix and some locational information is the sole data requirement.

6.1.2Collinearity amongst the time dummies

Estimation proceeds using the arithmetic and notation of the previous chapter. The models that can be estimated depend upon the matrices created. Clearly the balanced time differenced model can only be run on a balanced full-size dataset as outlined in section 5.3. However, this same dataset could also be used for the fixed-effects model (if T_i=T for all individuals, there is no need for a separate means matrix) and the cross-section models, which treat matrices separately. The actual combinations of matrices and models are described in a user guide to the software[2].

When the full fixed effects models are being estimated there is a problem of multicollinearity between the time dummies and the individual dummies. Differencing or taking deviations will remove the individual dummies, but will not restore the X matrix to full column rank. This can easily be seen if we consider the deviations transformation on a group of time dummies for T=4:

[2] Initial draft available from the DE or the author.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1-\frac{1}{T} & -\frac{1}{T} & -\frac{1}{T} & -\frac{1}{T} \\ -\frac{1}{T} & 1-\frac{1}{T} & -\frac{1}{T} & -\frac{1}{T} \\ -\frac{1}{T} & -\frac{1}{T} & 1-\frac{1}{T} & -\frac{1}{T} \\ -\frac{1}{T} & -\frac{1}{T} & -\frac{1}{T} & 1-\frac{1}{T} \end{bmatrix} \qquad \textbf{(6.1)}$$

The rank of the transformed matrix is T-1 and not T, and so this matrix is not invertible. However, suppose there are only three observations for one individual. It may thought that this leads to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1-\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & 1-\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1-\frac{1}{3} \end{bmatrix} \qquad \textbf{(6.2)}$$

which has full rank, but this is not the case. In the previous chapter the first matrix in (6.2) was depicted with no zero columns or rows to simplify exposition; in other words the data was packed and so the correct version of (6.2) is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1-\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & 1-\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & 1-\frac{1}{3} \end{bmatrix} \qquad \textbf{(6.3)}$$

where the transformed matrix has rank T-1 again. More correctly, note that the proper form of Q_i in (5.48) is always a TxT matrix, but with zeros on the appropriate rows and columns:

$$Q_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \qquad \textbf{(6.4)}$$

If these zeros did not appear in these places, then the transformation matrix would not sweep out the heterogeneity: spurious values of $-\alpha_i$ would appear in previously blank lines. This then leads to the transformation:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1-\frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} \\ -\frac{1}{3} & 1-\frac{1}{3} & 0 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & 0 & 1-\frac{1}{3} \end{bmatrix} \qquad \textbf{(6.5)}$$

which again is of rank 2. Dropping one time dummy (one column) will leave this particular matrix still with rank 2. However, this is one individual's record; when the matrix in (6.5) is stacked with the records for other individuals (whose patterns of observations differ) then the overall matrix will be of rank three. Therefore when the moment is taken of every individual's records to produce a 3x3 matrix, it will have full rank and so be invertible.

This makes no qualitative difference to the algebra, and so it was ignored in the previous chapter. As far as estimations goes, the program will automatically drop the first time dummy in a fixed-effects or balanced full-size differencing model to make the matrix invertible. In the case of the pooled models, this amounts to dropping the constant completely[3].

The value of this missing constant term can be recovered from the means. In all cases,

[3] The use of categorical variables also leads to collinearity problems. Selection of these other dummies to be dropped is up to the user.

$$\lambda_1 = \bar{y} - \bar{x}\hat{\beta} \qquad \textbf{(6.6)}$$

where the means are taken over the whole regression. This mean will also incorporate any dummy variables dropped; in other words, it is the expected value of the dependent variable for a "representative individual" - including the mean of the fixed-effects.

The standard errors given in section 5.2 and 5.3 have to be emended for these adjustments. In both cases, one is taken from the denominator of the estimated standard error, reflecting the fall in the number of variables used. Corrected standard errors are given in Appendix A6.

6.1.3Collinearity between time dummies and incremental variables

In a recent paper, Bell and Ritchie (1995a) have shown that allowing coefficients to vary over time has a hitherto unreported side-effect. When variables which increment or decrement periodically over time (such as age, tenure, age of youngest child, et cetera) are included in a regression which has time-varying coefficients, the coefficients on the incrementing variables are poorly identified because of collinearity with the time dummies and any other incrementing variables.

The reason is that the addition of an incrementing variable effectively amounts to the inclusion of a person-specific numerator variable and either a trend (in the case of an incrementing cardinal variable) or a secondary set of time dummies (in the case of qualitative variables). This combines with the time dummies (and any other incrementing variables) to make identification of the particular coefficients difficult.

This effect is specific to models where the coefficients are estimated jointly over time; thus the cross-section model of section 5.1 and the unbalanced differenced model of section 5.4 are

unaffected because of their block-diagonal nature. However, the unrestricted models of section 5.2 and 5.3 potentially have this identification problem, and so the interpretation of some coefficients requires some care. Chapter nine discusses a specific example of this identification problem.

6.1.4Single observations

The fixed effects covariance estimator takes deviations from individual means, and so clearly individuals with only one observation play no significant part in the estimation of the coefficients (although they will affect the calculation of λ_1 in (6.6)). The extraction software creates the main and mean matrices separately, for the reasons of practicality and flexibility discussed at the end of section 5.2.3, and it cannot take account of single-observation cases.

The effect of including single-observation cases and then excluding them is relatively minor, and does not affect the calculation of the coefficients. It does mean that the calculation of the estimated variance in, for example, (5.91) will have values for ΣT_i and N different to those arising from a model in which the single-observations are initially excluded. However, the number of single observations in any year is only around 3-5% of the total observed, so, while N may have 20% of single observations (and so be roughly 20% "too big") over the full sixteen years of the survey, ΣT_i is only around 4% "too big". As ΣT_i is easily the dominant term in the calculations for all but very short study periods, it seems likely that the estimated variance is slightly underestimated in the fixed-effects models.

The other area where having single observations upsets the results is in the displayed means, which include single observations in the calculations as they represents the mean values of each variable for a particular period. However, they play no part in the fixed effect calculations.

The single observation issue does not affect the cross-section studies, as these are only

concerned with observations within a period and not the correlation between observations over time. The differencing calculations likewise are unaffected: for the balanced panel a single observation period is not feasible, and for the unbalanced panel the extraction software rejects single observations.

6.2Hypothesis Testing

One of the more serious limitations of the regression program is the area of hypothesis testing. Many of the more informative tests are based on an analysis of the residual errors (serial correlation, heteroscedastic-consistent errors, et cetera). The relevant statistics would have to be calculated by sending a program to the DE offices, and so such statistics are not provided by the program. This is not a very satisfactory solution, but at present there is no alternative.

The analytical features generated automatically by the program are limited to what is available under the X'X format: essentially anything involving the total, estimated and residual sums of squares and other linear combinations of the variables. These are all available from the cross-product matrix by some method or other, and so some useful tests and statistics may be produced.

Given the TSS, ESS and RSS, then R^2 and R^2 adjusted for degrees of freedom may be calculated. The estimates of the variance lead to the t-statistics via the variance of β:

$$Var(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} \qquad \textbf{(6.7)}$$

and F-tests for the general significance of the regression are available from

$$F = \frac{ESS/dof1}{TSS/dof2} \qquad \textbf{(6.8)}$$

where dof1 and dof2 are the appropriate degrees of freedom. As noted in chapter five, F-tests

for choosing between the unrestricted, pooled and restricted models can also be calculated. Because the degrees of freedom are more complex for panel models (especially unbalanced ones), these are given in full in Appendix A6.

Equation (6.8) is a restricted form of a more general hypothesis-testing framework whereby sets of hypotheses may be tested jointly. The program can automatically calculate one set. Variables may be defined as parts of "groups"; usually, each of the different dummy variable groupings is generally treated as a set of related variables. The program then tests for the joint significance of all the groups which have two or more members; for example, the program will report an F-ratio for whether the occupation dummies as a whole contribute anything significant to the model, as well as the usual t-ratios for each individual occupational dummy.

These F tests are all based on the assumption of normality in the error term. There are as yet no tests in the model for this assumption, as most tests are based on an analysis of the residuals, and these are unavailable to the regression program.

6.3Estimated variance analysis

The program provides a breakdown of the variance by variable grouping, following a suggestion of Blackburn (1990). First, note that

$$TSS = ESS + RSS = \hat{\beta}'X'y + RSS = \hat{\beta}'X'X\hat{\beta} + RSS \qquad \textbf{(6.9)}$$

Assume that the X variables can be organised into M groups of variables. The number of elements in each group may vary; for example, a set of occupation dummies may count as one group, whereas a wage variable may be thought of as a one-element group. Then

$$X \equiv [\, X_1 \;\; X_2 \ldots X_M \,] \qquad \beta \equiv [\, \beta_{1'} \;\; \beta_{2'} \ldots \beta_{M'} \,]' \qquad \textbf{(6.10)}$$

with

$$X'X = \begin{bmatrix} X_{1'}X_1 & X_{1'}X_2 & & \\ X_{2'}X_1 & X_{2'}X_2 & & \\ & & \vdots & \\ & & & X_{M'}X_M \end{bmatrix} \qquad \textbf{(6.11)}$$

Using these definitions gives

$$ESS = \hat{\beta}'X'X\hat{\beta} = \sum_{m=1}^{M} \hat{\beta}_{m'}X_{m'}X_m\hat{\beta}_m + 2\sum_{m=1}^{M}\sum_{n=m+1}^{M} \hat{\beta}_{m'}X_{m'}X_n\hat{\beta}_n \qquad \textbf{(6.12)}$$

In other words, the explained sum of squares can be broken down into two components: firstly, the contribution of each group to the total explained variance; and secondly, the explained covariance between groups, which may be positive, zero, or negative.

This information is useful as it gives an indication of how the groups interact with one another; more importantly, it weights the results by the estimated coefficients. Thus it can be shown not whether two variables interact (which could be found simply from the covariance matrix of X), but whether that interaction is important to the relationship being studied.

This is perhaps most useful when regressions are run with time-varying coefficients. If, for example, the contribution of age to the variance in wages declines over time, this could be attributable to a decline in the variance of ages (the population is more homogeneous and so age has less chance of explaining wage differentials); a decline in the coefficient values (reflecting a decline in the return to age); or changes in both, not necessarily in the same direction.

One simple way to test this is by studying how the age variance changes over time. However, this does not take account of any scale effects. An alternative suggested by Blackburn (1990) is to apply the coefficients for one "base" year to the covariance matrices for each year in turn.

The result is effectively an index of the relevance of a variable group in the regression[4].

If a large part of the variance of y is explained by the own-variance terms, then this suggests that the various influences on the dependent variable are largely independent of one another. this can be seen by noting that if the variables are independent then as the number of observations becomes large

$$\hat{\beta}_{m'} X_{m'} X_m \hat{\beta}_m \to +\infty \quad \hat{\beta}_{m'} X_{m'} X_n \hat{\beta}_n \to \hat{\beta}_{m'} \overline{X}_{m'} \overline{X}_n \hat{\beta}_n \qquad \textbf{(6.13)}$$

where the covariances converge to the separate means of each group of variables. If the variables are not independent then

$$\hat{\beta}_{m'} X_{m'} X_m \hat{\beta}_m \to +\infty \quad \hat{\beta}_{m'} X_{m'} X_n \hat{\beta}_n \to \pm\infty \qquad \textbf{(6.14)}$$

However, in most of the models the analysis is done on deviations from either time or individual means. Thus the means in (6.13) will be the means of the transformed variables. Clearly the mean value of these transformed variables will be zero; moreover, the variables will converge to the sum of their mean values if the number of observations becomes large and the variables are not independent. Thus (6.13) and (6.14) become

$$\hat{\beta}_{m'} X_{m'} X_m \hat{\beta}_m \to +\infty \quad \hat{\beta}_{m'} X_{m'} X_n \hat{\beta}_n \to 0 \qquad \textbf{(6.15)}$$

when the variable groups are independent, and

$$\hat{\beta}_{m'} X_{m'} X_m \hat{\beta}_m \to +\infty \quad \hat{\beta}_{m'} X_{m'} X_n \hat{\beta}_n \to \pm\infty \qquad \textbf{(6.16)}$$

when they are not. For the balanced time-differenced model of section 5.3, the sum of all the transformed X variables is equal to the sum of the first observations for each individual, and so (6.13) and (6.14) will reflect the averages of these first observations.

Note that using (6.15) and (6.16) as indicators of independence will depend to some extent on

[4] The usual index number problem arises. With no preferences for one particular base over another, the simplest one to implement was chosen.

the scaling of the variables, particularly when categorical and continuous variables are mixed.

6.4Instrumental Variables

6.4.1Instrumental variable regression

One simple extension to the models outlined in the previous chapter is to allow for the use of instrumental variables. The linear generalised instrumental variables estimator (GIVE) can be derived in a number of ways; the GMM interpretation is given below (Hall(1993); see Bowden and Turkington(1984) for a 'traditional' derivation). Let Z be a matrix of instruments uncorrelated with the error vector u. Then

$$\begin{aligned} E(X'(X\beta - y)) &= E(X'u) \neq 0 \\ E(Z'(X\beta - y)) &= E(Z'u) = 0 \end{aligned} \qquad \textbf{(6.17)}$$

by assumption. Defining a quadratic form for the sample condition

$$S \equiv (\frac{1}{n}Z'u)'W(\frac{1}{n}Z'u) = (\frac{1}{n}Z'(y - X\beta))'W(\frac{1}{n}Z'(y - X\beta)) \qquad \textbf{(6.18)}$$

where W is a weighting matrix not dependent upon β which converges in probability to a positive definite matrix. Differentiating to find the value of β which minimises this expression,

$$\begin{aligned} \frac{\partial S}{\partial \hat{\beta}} = 0 &= -\frac{2}{n^2}X'Z'WZy + \frac{2}{n^2}X'Z'WZX\hat{\beta} \\ \hat{\beta} &= (X'Z'WZX)^{-1}X'Z'WZy \end{aligned} \qquad \textbf{(6.19)}$$

The optimal choice of weighting matrix is $W=n(Z'Z)^{-1}$, and so the linear GIVE is

$$\hat{\beta} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y \qquad \textbf{(6.20)}$$

Where Z and X have the same rank (that is, Z'X is square), (6.20) collapses to

$$\hat{\beta}_{iv} = (Z'X)^{-1}Z'y \qquad \textbf{(6.21)}$$

which is exactly the same form as the OLS estimator. All the matrix arithmetic of the previous

chapters therefore still holds, with the obvious proviso that the rows and columns selected from the $\Sigma v_i'v_i$ matrix will differ if $Z \neq X$[5]. This holds for the means matrix calculations too.

Where Z and X have the same rank, the calculations for standard errors are

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta}_{iv})'(y - X\hat{\beta}_{iv})}{dof} = \frac{y'y - 2\hat{\beta}_{i'v}X'y + \hat{\beta}_{i'v}X'X\hat{\beta}_{iv}}{dof}$$

$$Var(\hat{\beta}_{iv}) = \hat{\sigma}^2(Z'X)^{-1}(Z'Z)(X'Z)^{-1} \quad \textbf{(6.22)}$$

where dof is the appropriate number of degrees of freedom for the model (Johnston (1984), p366). These are the same as for the OLS estimator, as the transformation matrices, the number of observations and restrictions and the number of periods remain the same in the two models.

The instrumental variables estimator therefore requires five cross-product matrices (X'X, X'y, Z'Z, Z'X, Z'y) in contrast to the two needed for OLS (X'X, X'y). However, these can all be created from the raw cross-product matrix as long as the instruments already exist in the matrix. This is a significant disadvantage in that it greatly limits the options for two-stage solutions. For example, to run a two stage least squares regression would involve creating a dataset; running the first stage regression; writing a new extraction program using the estimated coefficients; creating a new dataset; and running the second stage regression. This may be tedious and threatens significant time penalties for a poor choice of regressors for the first round coefficients[6].

It was noted in section 2.3 that dynamic models could be consistently estimated by the use of

[5] Other minor changes from the programming point of view are that the matrix is not symmetric, and it is no longer necessarily positive definite.

[6] This does not weaken the claim that the cross-product matrix is an effective tool for linear regression; in fact, the case is more compelling for IV regressions, as the cross-product could contain numerous first-round estimates for a relatively small increase in matrix size.

lagged dependent variables as instruments. The extraction software does allow for lagged variables (including in the differenced format). Therefore dynamic models are feasible, although as the software currently stands this would involve the loss of a number of observations, and the estimator is unlikely to be very efficient. In the longer term a more flexible and efficient approach to dynamic models would be desirable and there are no theoretical difficulties to replicating, for example, the simpler Arellano and Bond (1988) estimators (that is, those with spherical errors).

If the number of instruments exceeds the number of regressors, then no new conceptual or practical difficulties arise. All the data in (6.20) is available from the cross-product matrix and the standard errors of the coefficients in (6.22) also need to be amended:

$$\begin{aligned} Var(\hat{\beta}_{iv}) &= \hat{\sigma}^2\left[('Z(Z'Z)^{-1}Z'X)'(Z(Z'Z)^{-1}Z'X)\right] \\ &= \hat{\sigma}^2\left[('X'Z(Z'Z)^{-1}Z')(Z(Z'Z)^{-1}Z'X)\right] \end{aligned} \qquad \textbf{(6.23)}$$

It is clear that the non-square z'x matrix does not require any additional information: all the data needed is somewhere in the cross-product matrix. This is a consequence of the linear nature of the IV estimator used here.

This has not been implemented in the program to date, purely from an operational point of view as it complicates the selection of variables somewhat (and this facility has not been needed so far). Thus the current version of the program does not cater for non-square Z'X matrices. However, as Bowden and Turkington (1984, pp29-30) show, only the minimum number of instruments play a significant part in the regression; in other words, the effective Z'X matrix is square[7].

6.4.2Testing the IV specification

[7] This does not mean that choosing minimal instruments will necessarily be an efficient IV solution, but that choosing more instruments than the minimum effective set will increase the variance of the estimates. See Bowden and Turkington (1984) pp32-36.

Two tests on the instruments are easily implemented using the cross-product matrix.

The first test statistic is a Hausman test for regressor-disturbance independence. This relies on the potential inconsistency of the OLS estimator compared to the supposed consistency of the IV estimator to provide a testable distance measure.

The test is between two hypotheses:

$$\begin{aligned} H_0: &\quad plim\ X'u = 0 \quad plim\ Z'u = 0 \\ H_1: &\quad plim\ X'u \neq 0 \quad plim\ Z'u = 0 \end{aligned} \tag{6.24}$$

Under the null hypothesis, OLS estimates of the coefficients are consistent and efficient, whereas the IV estimates are consistent but inefficient. However, under the alternative hypothesis, OLS is inconsistent. Define

$$\hat{q} \equiv \hat{\beta}_{iv} - \hat{\beta}_{ols} \tag{6.25}$$

Then the Hausman test is whetherq is significantly different from 0; that is, whether

$$n\hat{q}'\Omega\hat{q} = 0 \tag{6.26}$$

where Ω is a weighting matrix. The obvious choice for this weighting matrix is the inverse covariance ofq , and it can be shown (Hausman (1978); Bowden and Turkington (1984)), that under a fairly general set of assumptions, an asymptotic test statistic for (6.20) is

$$n\hat{q}'\hat{V}ar(\hat{q})^{-1}\hat{q} \_ \chi^2(K) \tag{6.27}$$

where K is the number of variables, n the number of observations, and Var(q) is a consistent estimate of the variance ofq ,

$$Var(\hat{q}) = Var(\hat{\beta}_{iv}) - Var(\hat{\beta}_{ols}) \tag{6.28}$$

Under the null hypothesis, Var(q) should be large (as the IV estimate of β is inefficient) andq

small, and so a large value for (6.23) indicates rejection of H_0[8].

It should be noted that this test is predicated on the assumption that the IV estimate is consistent even if H_0 is rejected. Without this assumption, this merely amounts to a test for the relative independence of Z compared to X. In other words, the Hausman test compares the relative performance of two estimators, both potentially erroneous. Thus this test does require a degree of confidence about the consistency of the IV estimator[9].

The second test is a general one for the validity of the instruments used, the Sargan test[10]. The test statistic is simply

$$q = \frac{1}{n\hat{\sigma}^2}(Z'e)'W(Z'e) = \frac{(y - X\hat{\beta})'Z(Z'Z)^{-1}Z'(y - X\hat{\beta})}{\hat{\sigma}^2} \sim \chi^2(p - k) \qquad \textbf{(6.29)}$$

where p and k are the number of columns in Z and X respectively. The basic idea behind the test is that, if the instruments are uncorrelated with the error terms, then $e'Z(Z'Z)^{-1}Z'e$ should converge to n independent squared errors, and so q should be small. The adjustment for degrees of freedom reflects the fact that, of the p columns in Z, k will be constrained by the action of setting $\partial S/\partial\beta=0$ in (6.19).

If p=k, then clearly the Sargan test is not appropriate. The weighting matrix is irrelevant, and q in (6.29) collapses to $\sigma^2/\sigma^2 = 1$. The reason is that the coefficient vector uses all the information in Z by construction, whereas in the overidentified case only the most effective columns of Z are significant (Bowden and Turkington (1984), p29). This test is not currently

[8] The "n" in (6.22) and (6.23) relates to the number of observations used to calculate the coefficient estimates, for the Hausman test is based on N repeated observations on a parameter set. For our purposes, n is N_t for the cross-section models, and $\Sigma_i T_i$ for the fixed effects models.

[9] See Bowden and Turkington (1984, pp52-55) for a discussion of what the Hausman test actually measures.

[10] The Sargan test as presented here can be seen as a particular form of Hansen's test of "overidentifying restrictions" in the GMM methodology (see Hall (1993) for the derivation of the general test statistic).

implemented as only square Z'X matrices are employed.

Appendix A6Degrees of Freedom in the Linear Model

The general solution for degrees of freedom are

$$dof = n - k - r \qquad \textbf{(A6.1)}$$

where n is the total number of observations, k is the number of variables, and r is the number of other restrictions: for example, taking deviations from time means (as in the cross-section case) means an additional T restrictions in the unrestricted model as the sum of the variables for each of the T periods must sum to zero. The appropriate calculations for these results are

$$R^2 = \frac{RSS}{TSS} = \frac{RSS}{TSS} = \frac{ESS}{TSS} \qquad \textbf{(A6.2)}$$

$$\frac{\hat{\sigma}^2}{\sigma^2} = \chi^2(n - k - r) \qquad \textbf{(A6.3)}$$

$$\bar{R}^2 = 1 - \frac{RSS/(n - k - r)}{TSS/(n - r)} \qquad \textbf{(A6.4)}$$

A general test of q linear restrictions of the form $R\beta=r$ has a χ^2 form:

$$(R\beta - r)'\left[\sigma^2 R(X'X)^{-1}R'\right]^{-1}(R\beta - r) \sim \chi^2(q) \qquad \textbf{(A6.5)}$$

which, combined with (A6.3) to remove the unknown σ, gives the F-test

$$\frac{(R\beta - r)'\left[R(X'X)^{-1}R'\right]^{-1}(R\beta - r)/q}{RSS/(n - k - r)} \sim F(q, n - k - r) \qquad \textbf{(A6.6)}$$

For a test of q variables being zero, this test collapses to

$$\frac{(RSS_r - RSS_u)/q}{RSS_u/(n_u - k_u - r_u)} \sim F(q, n - k - r) \qquad \textbf{(A6.7)}$$

where the r subscript refers to the RSS from some restricted regression and u the unrestricted or "base" regression. The value of q is slightly complicated if the number of other restrictions on the regression changes; for example, in the cross-sectional regressions the number of restrictions "r" may be T or 1, although in the fixed effects regressions the number of restrictions is always N as deviations are taken around the individual mean. This will change the calculations for the estimated variance used to derive (A6.7). The correct value for "q" in

(A6.7) is

$$q = (n_u - k_u - r_u) - (n_r - k_r - r_r) \qquad \textbf{(A6.8)}$$

For the joint significance test ($\beta \neq 0$), q will clearly be k: all the model results show deviations from some mean value, and so the appropriate test from (A6.6) is for all remaining variables being zero.

Let T be the number of periods under study, N be the number of individuals observed (not observations), T_i and N_t the observations for individual i and per period t, respectively, and K the number of variables in the X matrix, excluding the constant term. One time dummy is dropped in the fixed-effects and balanced differenced models. Then the degrees of freedom for the reported statistics are given in Table A6.1, overleaf.

Table A6.1Degrees of freedom

		Unrestricted	Pooled	Restricted
σ^2	Cross-section	ΣN_t - T - TK	ΣN_t - 1 - K	ΣN_t - T - K
	Fixed-effects	ΣT_i - N - TK - (T-1)	ΣT_i - N - K	ΣT_i - N - K - (T-1)
	Balanced differenced	$2\Sigma T_i$ - 2N - TK - (T-1)	$2\Sigma T_i$ - 2N - K	$2\Sigma T_i$ - 2N - K - (T-1)
	Unbalanced differenced	$2(\Sigma N_t$ - T - TK)	$2(\Sigma N_t$ - 1 - K)	$2(\Sigma N_t$ - T - K)
R^2	Cross-section	ΣN_t - T - TK ΣN_t - T	ΣN_t - 1 - K ΣN_t - 1	ΣN_t - T - K ΣN_t - T
	Fixed-effects	ΣT_i - N - TK - (T-1) ΣT_i - N	ΣT_i - N - K ΣT_i - N	ΣT_i - N - K - (T-1) ΣT_i - N
	Balanced differenced	$2\Sigma T_i$ - 2N - TK - (T-1) $2\Sigma T_i$ - 2N	$2\Sigma T_i$ - 2N - K $2\Sigma T_i$ - 2N	$2\Sigma T_i$ - 2N - K - (T-1) $2\Sigma T_i$ - 2N
	Unbalanced differenced	$2(\Sigma N_t$ - T - TK) $2(\Sigma N_t$ - T)	$2(\Sigma N_t$ - 1 - K) $2(\Sigma N_t$ - T)	$2(\Sigma N_t$ - T - K) $2(\Sigma N_t$ - T)
Joint F	Cross-section	TK ΣN_t - T - TK	K ΣN_t - 1 - K	K ΣN_t - T - K
	Fixed-effects	TK + T - 1 ΣT_i - N - TK - (T-1)	K ΣT_i - N - K	K + T - 1 ΣT_i - N - K - (T-1)
	Balanced differenced	TK + T - 1 $2\Sigma T_i$ - 2N - TK - (T-1)	K $2\Sigma T_i$ - 2N - K	K + T - 1 $2\Sigma T_i$ - 2N - K - (T-1)
	Unbalanced differenced	TK $2(\Sigma N_t$ - T - TK)	K $2(\Sigma N_t$ - 1 - K)	K $2(\Sigma N_t$ - T - K)
Specification Tests	CS Pooled v	(T-1)(K+1) ΣN_t - T - TK		
	CS Restricted v	K(T-1) ΣN_t - T - TK	T-1 ΣN_t - T - TK	
	FE Pooled v	(T-1)(K+1) ΣT_i - N - TK - (T-1)		
	FE Restricted v	K(T-1) ΣT_i - N - TK - (T-1)	T-1 ΣT_i - N - K - (T-1)	
	BD Pooled v	(T-1)(K+1) $2\Sigma T_i$ - 2N - TK - (T-1)		
	BD Restricted v	K(T-1) $2\Sigma T_i$ - 2N - TK - (T-1)	T - 1 $2\Sigma T_i$ -2 N - K - (T-1)	
	UD Pooled v	(T-1)(K+1) ΣN_t - T - TK		
	UD Restricted v	K(T-1) ΣN_t - T - TK	T-1 ΣN_t - T - TK	