Chapter 8
TESTING HYPOTHESES

After you have estimated a model using the methods in earlier chapters, you may want to test some hypotheses about the estimated parameters of the model. For example, you may want to test whether one or more variables belong in an equation, or whether the parameters satisfy some linear or nonlinear constraint. TSP has many methods for performing such tests, and in this chapter, we outline some of them; your ingenuity may find others.

Hypothesis testing of linear or nonlinear constraints on the parameters of econometric models can generally be performed using one of three methods. These methods are: the likelihood ratio (which compares constrained and unconstrained estimates), the Wald test (based on the unconstrained estimates), and the Lagrange multiplier, or LM, test (based on the constrained estimates). In different situations, you may find one or the other of these tests easier to compute. Although all are asymptotically equivalent, in finite samples the results will differ (except in very specific simple cases). For example, in the case of linear constraints in a generalized least squares model, the three statistics obey the following inequality:

$$\text{Wald} \geq \text{LR} \geq \text{LM}$$

See the excellent article by Engle (1985) for further information and references.

In this chapter we outline the tests that can be used with the linear estimation methods: the t-test, F-test, and a special case of the F-test called the Chow test. Then we discuss the tests available for testing nonlinear hypotheses, and linear or nonlinear hypotheses about nonlinear models. These fall into two basic classes, depending on the form of the hypothesis: the likelihood ratio and quasi-likelihood ratio test, and the Wald test. The t-test and F-test for linear models, and the Wald test for nonlinear models can all be done with the ANALYZ command. Finally we briefly discuss the Lagrange multiplier (score) test.

The LM test is an example of a type of test called a "specification test", for which a specific alternate hypothesis is often not available. Another large class of specification tests may be generated by a principle enunciated by Hausman (1977). The Hausman test is discussed in section 8.9 and an example is given in Section 13.4.1.

Classical hypothesis testing involves accepting or rejecting decisions based on tabulated significance levels for the theoretical distribution of the test statistic. TSP provides these tables as a procedure (CDF) for the Student's t, F, chi-squared, normal, and bivariate normal distributions so that you can easily find the p-level for any test statistic you calculate. CDF also contains distributions for the Dickey-Fuller tests for unit roots and the Engle-Granger cointegration tests, which are discussed in Chapter 11.

8.1. t-tests

The simplest hypothesis test is the t-test, usually a test for equality of a single coefficient in your model with some prespecified value. The most common value is zero. TSP always prints out the value of the statistic for the t-test of zero in the table of regression results.

In general, t-tests on single coefficients are of the form (estimated coeficient - hypothesized value)/(estimated standard error) . For the standard case of testing against zero, this simplifies to the ratio of the estimated coefficient to its standard error. For example, if you want to test the hypothesis that the coefficient on X1 is one:

```
OLSQ Y C X1 X2 X3;
SET TTEST = (@COEF(2) - 1)/@SES(2);       ? X1 is the second variable in @COEF and @SES
SET DFT = @NOB - @NCID;                   ? degrees of freedom — used to compute the P-value
CDF(T,DF=DFT) TTEST;                      ? print TTEST and its P-value
```

You can do the same thing using the ANALYZ procedure (described below):

```
FRML TEST1 X1-1;              ? Hypothesize that X1=1
ANALYZ TEST1;                 ? Compute F-test (1, .)
```

T-tests of this kind, both simple and complex, may be performed on the results of any TSP linear or nonlinear estimation procedure. For some estimation methods, mostly nonlinear models, the resulting statistic will not be a t-statistic, but an asymptotically normal variable. The distribution of that statistic is almost the same as a two-tailed t-statistic based on more than 50 observations and provides a more conservative test in small samples. For certain time series regressions (those using integrated variables), tests like the above result in a Dickey-Fuller test, rather than a t-test.(See Section 11.6.)

T-tests can also be computed and displayed for estimators you have programmed with matrix procedures. Just supply the TSTATS command with the names of the coefficient vector and the variance-covariance matrix, and TSTATS will print a standard regression output table. For example:

```
TSTATS(NAMES=(BETA1-BETA7)) BETA VARB;
```

prints a table of the seven BETA coefficients, together with their standard errors (the square roots of the diagonal elements of VARB) and t-statistics for the hypothesis that each of them is zero.

8.2. F-tests

F-tests are commonly used to test linear hypotheses which involve more than one coefficient. The simplest way to compute F-tests is to use the ANALYZ command, where you write one FRML for each restriction. ANALYZ will compute the value of each restriction, the F-statistic for all the restrictions as a joint test, and the implied constrained values of the original coefficients. (Note: ANALYZ only computes an F-statistic for linear restrictions following an OLSQ command -- following other procedures ANALYZ computes an asymptotic chi-squared test). For example, to test that the sum of the coefficients on X1 and X2 equals one, use the following commands:

```
OLSQ Y C X1 X3 X3;
FRML SUM1  X1+X2 - 1;   ? write the restriction so that the value would be zero if the restriction is true
ANALYZ SUM1;
```

Another way to test linear hypotheses is to estimate the unconstrained model, then impose the null hypothesis and estimate a constrained version of the model. In this case an F-statistic (with numerator degrees of freedom equal to the number of constraints and denominator degrees of freedom equal to the degrees of freedom in the unconstrained model) can be computed from the sum of squared residuals of the two models. Our example above had one constraint, that the sum of b_1 and b_2 was unity (where b_1 and b_2 are the coefficients of X1 and X2, respectively). We can write this constraint as $b_1=(1-b_2)$, so we can impose this by regressing (Y-X1) on C, (X2-X1), and X3. The commands for this test are:

```
OLSQ Y C X1 X2 X3 ;                           ? Unrestricted Model
SET SSRU = @SSR ; SET DFU = @NOB-@NCID ;      ? Sum of squares and degrees of freedom
GENR YX1 = Y-X1 ; GENR DX = X2-X1 ;           ? for unrestricted model.
OLSQ YX1 C DX X3 ;                            ? Restricted Model.
SET FSTAT = ((@SSR-SSRU)/1)/(SSRU/DFU) ;
CDF(F,DF1=1,DF2=DFU) FSTAT ;                  ? Print out p-value for F-statistic.
```

TSP's internal variable @NOB contains the number of observations in the current SMPL and is always available for use in computations. @NCID is the number of estimated coefficients. Note the use of CDF to print out the significance level associated with this F-statistic.

One advantage of this last technique is that it easily encompasses situations where there may be many constraints, particularly situations consisting of several zero restrictions on variables. It is a variation on the quasi-likelihood ratio techniques discussed in the section on nonlinear two- and three-stage least squares.

8.3. Chow tests

Chow tests are a special form of F-test that check the stability of regression coefficients over two or more subsamples of the data. This is normally done by running a regression for the whole sample, then running the same regression for subsamples, and comparing the sums of squared residuals (SSRs). An F-test for the constraint that the two sets of coefficients are equal can be computed from the SSRs. However, TSP automates this particular form of the F-test, using the options CHOW and CHOWDATE in the REGOPT procedure. Here is a simple example of a Chow test, both automated and manual:

```
SMPL 47:1 80:1;
REGOPT(CHOWDATE=60:2) CHOW;                    ? The two periods are 47:1-60:1; 60:2-80:1
OLSQ Y C X;
```

To compute this F-statistic "by hand", run the unconstrained and constrained regressions, and then form the F ratio:

```
?   estimate for the whole sample.
SMPL 47:1 80:1 ;
SET DF = @NOB-4 ;
OLSQ Y C X ;
SET SSR0 = @SSR ;
?   estimate for subperiod 1.
SMPL 47:1 60:1 ;
OLSQ Y C X ;
SET SSR1 = @SSR ;
?   estimate for subperiod 2.
SMPL 60:2 80:1 ;
OLSQ Y C X ;
SET SSR2 = @SSR ;
?   compute and print the Chow test.
SET CHOW = (SSR0-SSR1-SSR2)*DF/(2*(SSR1+SSR2)) ;
CDF (F,DF1=2,DF2=DF) CHOW ;
```

8.4. Pseudo-F tests for 2SLS

The 2SLS residuals are asymptotically normal, and can be used to form a pseudo-F test that has an approximate F distribution, following Startz (1983). The main difference is that the 2SLS objective function @PHI ($e'P_Z e$) is used in the numerator instead of @SSR. The previous F test example can be used to illustrate the pseudo-F test:

```
2SLS Y C X1 X2 X3 | C Z1-Z3 ;  ? Unrestricted model
SET PHIU = @PHI ;   SET S2U = @S2 ;   SET DFU = @NOB-@NCID ;
GENR YX1 = Y-X1 ;   GENR DX = X2-X1 ;
2SLS YX1 C DX X3 | C Z1-Z3 ;  ? Restricted model
SET PSEUDOF = ((@PHI-PHIU)/1)/S2U ;
CDF(F,DF1=1,DF2=DFU) PSEUDOF ;
```

Chow tests can also be performed with this methodology, but be careful to expand the instrument list for the restricted model (the one estimated on the whole period) by multiplying it by subsample dummies. This makes the instruments the same for all models being compared:

```
?   set up the subsample dummies.
SMPL 47:1 80:1 ;   LOW = 0 ;   HIGH = 0 ;
?   estimate for subperiod 1.
SMPL 47:1 60:1 ;   LOW = 1 ;
2SLS Y C X | C Z ;
SET PHI1 = @PHI ;   SET SSR1 = @SSR ;
?   estimate for subperiod 2.
```

```
SMPL 60:2 80:1 ;   HIGH = 1 ;
2SLS Y C X | C Z ;
SET PHI2 = @PHI ;   SET SSR2 = @SSR ;
?   estimate over the whole sample (restricted model)
SMPL 47:1 80:1 ;
ZLOW = Z*LOW ;   ZHIGH = Z*HIGH ;
2SLS Y C X | LOW HIGH ZLOW ZHIGH ;
SET PHI0 = @PHI ;
?   compute and print the pseudo-F version of the Chow test.
SET DF = @NOB-4 ;
SET PFCHOW = ((PHI0-PHI1-PHI2)/2) / ((SSR1+SSR2) / DF) ;
CDF (F,DF1=2,DF2=DF) PFCHOW ;
```

8.5. Likelihood ratio tests

In many ways, the likelihood ratio test is conceptually the easiest of the tests we consider in this chapter. It can be used for any model that is estimated by maximum likelihood. In TSP, this includes OLSQ, AR1, LSQ without instrumental variables (nonlinear least squares or multivariate regression), FIML, PROBIT, TOBIT, SAMPSEL, LOGIT, and ML. For these methods, TSP prints out a value of the logarithm of the likelihood evaluated at the estimated parameters and stores it under the name @LOGL.

If L_1 is the value of the likelihood function for the maximum of the unconstrained model and L_0 is the value when the constraints are imposed, then the likelihood ratio test is computed as

$$LR = 2(L_1 - L_0)$$

This test is always positive (or zero) since the likelihood of the unconstrained model is at least as high as that of the constrained model. The LR statistic is distributed asymptotically as a chi-squared variable with degrees of freedom equal to the number of constraints. For further information on the asymptotic properties of this statistic in nonlinear models estimated by maximum likelihood, see Gallant and Holly (1980).

To compute this test in TSP, save the two values of @LOGL for the unconstrained and constrained estimates, difference them, and multiply by 2. You can use CDF with the chi-squared distribution to obtain the associated p-value.

8.6. Nonlinear two- and three-stage least squares -- the QLR test

A quasi-likelihood ratio test can be used for hypothesis testing in a nonlinear model estimated by two- or three-stage least squares. This type of test is discussed in Gallant and Jorgenson (1979). The tests are based on the principle that if an estimated function of the data is asymptotically normally distributed with a variance-covariance matrix that can be consistently estimated, it is possible to construct a variable that is chi-squared distributed from this function of the data. In the cases discussed here, the function in question is the vector of differences between the residuals under the null and the residuals under the alternative, which has expectation zero under the null. The appropriate covariance matrix estimate is the minimum distance weighting matrix under the most unrestricted model.

The QLR test which Gallant and Jorgenson propose for three-stage least squares is

$$T = n\,(Q_0 - Q_1)$$

where Q_0 is the value of the minimum distance criterion for the null hypothesis, Q_1 its value for the maintained hypothesis, and n the number of observations.

The formula for the minimum distance function Q(b), b a vector of parameters, is

$$Q = f(b)'(S^{-1} \otimes P_Z)f(b)$$

where f(b) is the stacked vector of residuals from the model, S a consistent estimate of the covariance of the disturbances, and P_Z the projection matrix of the instruments, $Z(Z'Z)^{-1}Z'$. LSQ minimizes nQ(b) when it is doing three-stage least squares and prints it as F =, FNEW =, and E'HH'E. It may be retrieved under the name @PHI after convergence.

A cautionary note on testing with a minimum distance criterion: the statistic can be ill-behaved (have the wrong sign) if the estimate of S is not held constant across the null and maintained hypotheses. To do so, obtain a consistent estimate of S using nonlinear two-stage least squares and then maintain that estimate as a constant weighting matrix while doing three-stage least squares on the two models. For example:

```
3SLS(INST=(Z1,Z2,...),MAXITW=0) MEQ1 MEQ2 ... ;
COPY @COVU S ;
3SLS(INST=(Z1,Z2,...),MAXITW=0,WNAME=S) EQ1 EQ2 ... ;
SET Q0 = @PHI ;
3SLS(INST=(Z1,Z2,...),MAXITW=0,WNAME=S) MEQ1 MEQ2 ... ;
SET Q1 = @PHI ;

SET TEST = Q0-Q1 ;
CDF (CHISQ,DF=# constraints) TEST ;
```

Q0 is the objective function for the null hypothesis and Q1 is the maintained hypothesis. The first estimator is nested within the second, hence Q0 $\geq$ Q1, and the test is necessarily positive. This test is asymptotically chi-squared with degrees of freedom equal to the number of restrictions imposed, that is, the number of parameters in the second model minus the number in the first.

For two-stage least squares the objective function used by TSP is

$$Q(b) = (1/n)\ f'(b)\ P_Z\ f(b)$$

In this case, the same kind of test can be computed as

$$T = n\ (Q_0 - Q_1)/s^2$$

where s^2 is a consistent estimate of the variance of the disturbance (in particular, one may use the standard error squared of the two stage estimate of the maintained hypothesis). In TSP, nQ(b) is stored as @PHI when two stage least squares estimates are obtained with LSQ, so that a QLR test for two-stage least squares can be done in the following way:

```
LSQ(INST=(Z1,Z2, ...)) MEQ1 ;
SET S2 = @S*@S ;   SET Q1 = @PHI ;
LSQ(INST=(Z1,Z2, ...)) EQ1 ;
SET Q0 = @PHI ;
SET TEST = (Q0-Q1)/S2 ;
CDF (CHISQ,DF=# constraints) TEST ;
```

8.7. Wald tests -- testing linear and nonlinear restrictions: ANALYZ

The likelihood ratio and quasi-likelihood ratio tests, which compare the estimates of a constrained and unconstrained version of the model, are best suited to testing hypotheses of the form

$$b = f(g)$$

that is, some larger set of parameters b is expressed as a set of nonlinear functions of a smaller set g. When the constraint is expressed this way, it is usually easy to write the equations for both the unconstrained and constrained models. However, likelihood ratio tests have the disadvantage of requiring the estimation of both forms of the model, which can become expensive (in terms of CPU time) for large nonlinear models.

The Wald test provides an alternate method for performing the same test. Asymptotically it is the same as the likelihood ratio test if the null hypothesis is true, although it may differ quite a bit in practice. Choosing among these tests is not straightforward. See Gallant and Holly (1980) and Berndt and Savin (1977) for further discussion.

In TSP, the Wald test can be done with ANALYZ; it is a generalization of the t- and F-tests described above. This test is also known as the delta method in nonlinear contexts. Suppose the hypothesis to be tested can be written as

$$h(b) = 0$$

where b is the vector of parameters of the unconstrained model and h(b) is a set of m nonlinear constraints on those parameters. Given a set of estimates b and the associated covariance estimate V(b), ANALYZ computes the constraints h(b) (a row vector) and their covariance matrix:

$$V(h(b)) = (\partial h/\partial b)' \, V(b) \, (\partial h/\partial b)$$

all evaluated at the estimated b vector. From h(b) and its variance we form a test statistic

$$T = h(b) \, V[h(b)]^{-1} \, [h(b)]'$$

This test statistic is distributed asymptotically as a chi-squared variable with degrees of freedom equal to m under the null hypothesis (when the constraints hold). The p-value of $\chi^2(m)$ is printed; if you are testing at the .05 significance level, and the p-value is less than .05, the null hypothesis is rejected.

ANALYZ computes this test if you specify the constraints h(b) as FRMLs in unnormalized form (no left-hand side variable or equal sign), estimate the unconstrained model, and then issue an ANALYZ command with the names of FRMLs as arguments.

Gallant and Jorgenson give an example of testing for symmetry in a translog consumer demand system with three goods (two equations to be estimated). The TSP commands to compute this example are:

```
FRML EQ1 Y1 =  (A1 + B11*LNP1+B12* LNP2+B13*LNP3+B1T*TIME)/
               ( -1 + (B11+B21+B31)* LNP1+(B21+B22+B23)*LNP2
                    + (B31+B32+B33)* LNP3+(B1T+B2T+B3T)*TIME) ;
FRML EQ2 Y2 =  (A2 + B21*LNP1+B22* LNP2+B23*LNP3+B2T*TIME)/
               ( -1 + (B11+B21+B31)* LNP1+(B21+B22+B23)*LNP2
                    + (B31+B32+B33)* LNP3+(B1T+B2T+B3T)*TIME) ;
PARAM A1 A2 B11-B13 B21-B23 B31-B33 B1T B2T B3T ;
LSQ (INST=(...list of instrumental variables...)) EQ1 EQ2 ;
FRML SYM1 B12-B21 ;
FRML SYM2 B13-B31 ;
FRML SYM3 B23-B32 ;
ANALYZ SYM1 SYM2 SYM3 ;
```

ANALYZ results consist of a table with the computed differences of the symmetry parameters and their implied standard errors. The value of the Wald test, degrees of freedom, and p-value follow.

8.8. Lagrange Multiplier Tests (Score Tests)

The Lagrange Multiplier test is generally based on the magnitude of the derivatives of the likelihood function with respect to the constraints evaluated at the constrained estimates. Sometimes you may find this test easier to compute, particularly when the alternate hypothesis is not well specified. (See Engle (1985) for background).

As Breusch and Pagan (1979) among others have shown, many LM tests can be computed as the number of observations times the R^2 from a particular regression depending on the hypothesis to be tested. For example, suppose you have estimated an equation (linear or nonlinear) under the assumption that the disturbances are homoskedastic and wish to test for heteroskedasticity as an unknown function of the exogenous variables. This unknown heteroskedasticity can be modeled as a polynomial function of the Xs:

$$\sigma_t^2 = a_0 + a_1X_{1t} + a_2X_{2t} + a_{12}X_{1t}X_{2t} + a_{11}X_{1t}^2 + a_{22}X_{2t}^2 + ...$$

A test for heteroskedasticity of σ_t^2 is a test that $a_1=a_2=a_{12}=...=0$ in this equation. The test is performed by regressing a consistent estimate of σ_t^2on the exogenous variables of the model and their powers, and computing the TR^2 of the regression, where T is the sample size:

```
OLSQ Y C X1 X2 ;
USQ = @RES*@RES ;
X1X2 = X1*X2 ;  X1SQ = X1*X1 ;  X2SQ = X2*X2 ;
OLSQ(SILENT) USQ C X1 X2 X1X2 X1SQ X2SQ ;
SET TRSQ = @NOB*@RSQ ;
CDF (CHISQ,DF=5) TRSQ ;
```

This particular LM test has been automated and is more easily computed as follows:

```
REGOPT(PVPRINT) WHITEHT;
OLSQ Y C X1 X2;
```

In the case of qualitative dependent variable models, the same method can be used with the derivative of the model with respect to σ^2 substituted for the dependent variable USQ. Here is a more complex example based on the sample selection model described in Section 9.3 of this manual:

Suppose that you wish to test for heteroskedasticity of the disturbance e_2 in the sample selection model, using the form

$$\sigma_t^2 = G(\alpha + X_t\gamma)$$

where G is an arbitrary monotonic transformation. Then (see Lee and Maddala 1985, Poirier and Ruud 1983) a simple LM test can be obtained by regressing the partial of the likelihood function with respect to σ^2 (evaluated at the maximum likelihood estimates obtained under homoskedasticity) on the X_t's and their powers and examining the TR^2 from that regression. In this case, the partial of the log likelihood function for a single observation with respect to σ^2 can be shown to be proportional to:

$$\omega = (e_{2t}/\sigma)^2 - 1 - \rho(e_{2t}/\sigma)\lambda_t$$

where λ_t is the inverse Mills ratio for the observation, stored by the SAMPSEL procedure as a series. The LM test for $\gamma = 0$ is a test for:

$$(\partial \log L/\partial\sigma^2)G'(\alpha)X_t = 0$$

where the degrees of freedom for the test are the number of regressors in X_t and all quantities are evaluated at the maximum likelihood estimates obtained under the null hypothesis. Note that since $G'(\alpha)$ does not vary across observations, it drops out of the R^2 in the test, so that the exact monotonic transformation does not matter. The test is obtained by regressing ω on X_t and X_t^2 and examining the TR^2 from that regression. Here is an example of how to code this test in TSP:

```
SAMPSEL IN79 C LOGE76 | DLE7679 C LOGE76 ;
LIST SSCOEF DO D1 BO B1 SIGHAT RHOHAT ;        ? Names of sample selection coefficients.
UNMAKE @COEF SSCOEF ;
SMPLIF IN79 ;                                  ? Choose observed data.
UHAT = @RES/SIGHAT ;                           ? Standardized residuals.
DLOGL = UHAT*UHAT - 1 - RHOHAT*UHAT*@MILLS ;
LOGESQ = LOGE76*LOGE76 ;
OLSQ DLOGL C LOGE76 LOGESQ ;                   ? Compute LM test.
SET TRSQ = @NOB*@RSQ ;
CDF (CHISQ,DF=2) TRSQ ;
```

8.9. Hausman Specification Tests

The Hausman specification test is based on the comparison of an estimator which is efficient (or more efficient) under the null hypothesis but inconsistent under the alternative with an estimator which is consistent (and less efficient) under both hypotheses. This principle may be applied in many settings; useful applications are testing for the exogeneity of instruments or testing for random versus fixed effects in panel data. When applied to the simultaneous equations setting to test for exogeneity, it is often called a Hausman-Wu test.

A general example of a Hausman test in TSP is given in Section 13.4.1, after we have presented matrix operations and matrices. Some classic applications of the Hausman test are tests for random versus fixed effects in panel data now automated in the PANEL procedure, and tests for the independence of irrelevant alternatives in the Logit model (Hausman-McFadden (1984)).