Chapter 5
ESTIMATION OF LINEAR EQUATIONS

This chapter discusses several simple statistical models that involve linear structural equations. The models include: descriptive univariate and bivariate statistics (MSD), ordinary least squares (OLSQ), two-stage least squares (2SLS), limited information maximum likelihood (LIML), least squares with correction for serial correlation (AR1), and least absolute deviations estimation (LAD).

In these procedures the equation is defined implicitly by listing its variables. The observations used for estimation are defined by the current sample (the most recent SMPL, SELECT, or SMPLIF). Any observations containing missing values are temporarily dropped from the sample for all procedures except AR1 with the ML option. Most of the models allow the WEIGHT option for general least squares estimates, and the ROBUST option for standard error computations robust to heteroskedasticity. They also allow distributed lag variables (polynomial, Shiller, or unconstrained). You can use REGOPT (described fully in the *Reference Manual*) to control the calculation and printing of numerous diagnostic statistics.

Chapter 9 discusses estimation with qualitative dependent variables (PROBIT, TOBIT, LOGIT, SAMPSEL). Estimation methods for nonlinear and simultaneous equation (linear and nonlinear) models are covered in Chapter 7 (LSQ, FIML, GMM) and in Chapter 9 (ML). Chapter 10 describes constructing hypotheses tests for linear and nonlinear models. Chapter 11 discusses estimation of time series models (ARIMA models, VAR and GARCH-M estimation).

5.1. Descriptive Statistics: MSD, CORR

Before doing any regressions or using the more expensive (in terms of computer time) estimation procedures, check your input and transformed data: you can look for bad observations or outliers introduced by data entry error and check that the data transformations did what you expected. If you don't have much data, you can check it by using the PRINT command, but this method may be overwhelming if you have several hundred observations.

TSP has several descriptive statistics procedures that conveniently summarize large amounts of data for you. MSD prints a table of univariate statistics. MSD (CORR) computes and prints a correlation matrix. MSD(COVA) produces a covariance matrix. MOMENT computes an uncentered sum of squares and cross products matrix. See Chapter 6 for a description of the graphics procedures, including HIST, which prints frequency bar charts.

MSD followed by a list of series will produce a table of means, standard deviations, sums, variances, skewness, kurtosis, and maximum and minimum values for all the series. Example:

```
MSD LSALES,LGNP;
```

The CORR, COVA, and MOMENT options work the same way, except that they produce a matrix for the list of series provided. The moment of two variables is their uncentered covariance -- the sum of cross products divided by the number of observations.

```
MSD(CORR) LSALES,LGNP ;
```

would store the 2x1 vectors @MEAN, @STDDEV, @MIN, @MAX, @SUM, @VAR, and the 2x2 matrix @CORR. (See section 6.1 for more information about using @ variables.) The results computed by these commands are stored for use in further analysis.

You can use the WEIGHT= option in MSD to provide a weighting variable for the computations. (See Section 5.8.)

To obtain the median and interquartile range of the series use the ALL option of the MSD command:

```
MSD (ALL)  LSALES LGNP ;
```

```
                                    Equation   1
                                    ============
                        Method of estimation = Ordinary Least Squares

Dependent variable: CONS
Current sample:  1949 to 1975
Number of observations:  27

       Mean of dep. var. = 519.033      LM het. test = .783303 [.376]
  Std. dev. of dep. var. = 147.459     Durbin-Watson = .461739 [.000,.000]
Sum of squared residuals = 3958.17  Jarque-Bera test = 3.29859 [.192]
   Variance of residuals = 158.327   Ramsey's RESET2 = 16.7953 [.000]
Std. error of regression = 12.5828   F (zero slopes) = 3545.79 [.000]
               R-squared = .992999    Schwarz B.I.C. = 5.23184
      Adjusted R-squared = .992719    Log likelihood = -105.645

           Estimated    Standard
Variable  Coefficient     Error       t-statistic   P-value
C          -17.8024      9.33496       -1.90707      [.068]
GNP        .633919       .010646       59.5465       [.000]

ID      ACTUAL(*)   FITTED(+)                          RESIDUAL(O)
                                                                  O
1949     320.0000   293.2618    +*                        26.7382     +  |   +   O
1950     338.1000   320.3936     +*                       17.7065     +  |   +O
1951     342.3000   347.6521      +                       -5.3521     +O|   +
1952     350.9000   361.5983      *+                     -10.6983     O  |   +
1953     364.2000   376.3686       *+                    -12.1686     O  |   +
1954     370.9000   371.2339       +                      -0.3339     +  O   +
1955     395.1000   397.2880        +                     -2.1879     +  O   +
1956     406.3000   406.1628         +                     0.1372     +  |O  +
1957     414.7000   413.8333         +                     0.8667     +  |O  +
1958     419.0000   412.9458         +                     6.0542     +  |  O+
1959     441.5000   438.8731          +                    2.6269     +  |O  +
1960     453.0000   449.2693           +                   3.7307     +  |O  +
1961     462.2000   460.9968           +                   1.2032     +  |O  +
1962     482.9000   488.7625            +                 -5.8625     +O|   +
1963     501.4000   508.7944             +                -7.3944     +O|   +
1964     528.7000   536.4966              +               -7.7966     +O|   +
1965     558.1000   569.1435               *+            -11.0435     O  |   +
1966     586.1000   604.0724                *+           -17.9724    O+  |   +
1967     603.2000   620.9981                 *+          -17.7981    O+  |   +
1968     633.4000   648.9539                  *+         -15.5539    O+  |   +
1969     655.4000   666.0698                   *+        -10.6697     O  |   +
1970     668.9000   663.8510                    +          5.0490     +  |  O+
1971     691.9000   684.2632                     +         7.6368     +  |  O+
1972     733.0000   724.5805                      +*       8.4195     +  |  O+
1973     766.3000   764.0737                         +     2.2263     +  |O  +
1974     759.8000   749.6836                        +*    10.1163     +  |   O
1975     766.6000   734.2795                        +*    32.3205     +  |   +   O
```

Example 5.1: Ordinary Least Squares

5.2. Ordinary least squares: OLSQ

The first variable specified in the OLSQ command is the dependent variable and the rest are the independent variables. Recall, for example, the consumption function in the illustrative model:

CONS=A+B*GNP

A is a parameter that multiplies an implied variable, the constant, which always has the value 1. In TSP, the constant has the special name, C (or CONSTANT). It may be used in OLSQ or elsewhere without loading it as data. The OLSQ statement for our consumption function is

OLSQ CONS C,GNP ;

OLSQ computes the least squares regression coefficients and a variety of associated statistics. These statistics include the standard error of the residuals from the regression, the Durbin-Watson statistic, and the mean and standard deviation of the dependent variable. The R-squared, R-bar-squared, and F-statistic for the hypothesis that all the coefficients except the constant are zero are also printed.

You can plot the actual and fitted values of the dependent variable by including PLOTS earlier in the program.

5.3. Regression output

All regression procedures in TSP produce output in the same format. We describe the output here using the OLSQ command as an example (**Example 5.1**). In later sections of the manual we will point out any differences you may expect when using other estimation methods.

Using the standard textbook notation where y_t is the dependent variable and X_t is the vector of independent variables, e_t (the residual) is defined as

$$e_t = y_t - \hat{y}_t \qquad \text{where } \hat{y}_t = X_t b$$

and where b denotes the estimated regression coefficients. To give the exact formulas for the regression statistics in the TSP output, we define the sum of squared residuals **SSR** and the total sum of squares **SST**:

$$SSR = \Sigma\, e_t^2$$
$$SST = \Sigma\, (y_t - \bar{y})^2$$

The standard error of the regression is then given by

$$s = \sqrt{SSR/(T-K)}$$

where T is the number of observations in the current sample and K is the number of independent variables.

The **R-squared** is defined as the squared correlation coefficient between y and $\hat{y}$:

$$R^2 = [\mathrm{cov}(y, \hat{y})]^2 / [\mathrm{var}(y) * \mathrm{var}(\hat{y})]$$

R-squared is defined the same way for the 2SLS, AR1, LIML and PROBIT commands; only the formulas for $\hat{y}$ and b are different. For an OLSQ regression with a constant term, this squared correlation is equal to 1-**SSR**/**SST**. If $\hat{y}$ is constant, the R-squared will be zero. On the other hand, if the regression yields a perfect fit of the dependent variable, the correlation and the R-squared will be one; thus you can interpret the R-squared as the fraction of total variance that is "explained" by the variables in your regression other than the constant. It is also the squared cosine of the angle between the actual and predicted y and $\hat{y}$, once the means have been removed.

The **adjusted R-squared** is the quantity called R-bar-squared in some texts. The formula for it is given by:

$$\bar{R}^2 = (R^2(T-1) + 1 - K)/(T-K)$$

For a regression with a constant term this can be written as

$$\bar{R}^2 = 1 - [SSR/(T-K)]/[SST/(T-1)]$$

This quantity has the advantage that it does not automatically increase as variables are added to the regression since the numerator includes an adjustment for the number of estimated coefficients (K). As more variables are added to the regression with little or no additional explanatory power, it is possible for this quantity to become negative.

The next few statistics test the assumptions of the linear regression model: homoskedasticity, lack of serial correlation, correct functional form, and normality.

The **LM heteroskedasticity test** is a test for homoskedasticity of the residual variances. It is computed by regressing the squared residuals on the squared fitted values of the regression. The resulting TR^2 has a chi-squared distribution with one degree of freedom.

The **Durbin-Watson statistic** is described in Durbin and Watson (1951). A useful discussion of its properties and how to interpret it is in Pindyck and Rubinfeld (1991). The formula used by TSP to compute this statistic is

$$DW = \sum_t (e_t - e_{t-1})^2 / SSR$$

This statistic is valid even if there are gaps in the SMPL (missing observations), although in that case, it will be somewhat conservative. The p-values printed next to the Durbin-Watson statistic are exact finite sample values when the data are time series, and are an asymptotic approximation when the data have frequency N. See REGOPT in the *Reference Manual* for details.

If a lagged dependent variable is included in the right-hand-side variables, the Durbin-Watson statistic is biased, so TSP prints two alternate statistics described in Durbin (1970). These two statistics are labeled "Durbin's h" and "Durbin's h (alt.)" respectively and they are asymptotically equivalent. Sometimes h cannot be computed due to a negative square root, while the h alternative may have little power in small samples.

Ramsey's RESET test is a Lagrange multiplier test of functional form, computed by regressing the residuals on the independent variables and the square of the fitted dependent variable. This version of the test has a chi-squared distribution with one degree of freedom. The RESET test can also be significant if there is a single outlier that is fit well by a quadratic function of X. The **Jarque-Bera test** is a joint test for skewness and kurtosis of the disturbances, asymptotically distributed as chi-squared random variable with two degrees of freedom under the null of normality.

The individual t-statistics produced by TSP correspond to variable-by-variable tests of exclusion restrictions: they are computed as the value of the coefficient divided by its estimated standard error. This quantity is distributed as a t variable with (T-K) degrees of freedom. The t-statistic produced by TSP represents a test of only one of many possible hypotheses about the coefficient estimate: that it is zero. You may wish to test a different hypothesis, i.e., that the coefficient is equal to a_0. To construct the t-statistic for this test, compute

$$t = (\text{coef} - a_0)/(\text{std. err.})$$

If the HI option is used with OLSQ, the series $h_i = \text{diag}(X(X'X)^{-1}X')$ is stored under the name @HI. This series is useful for detecting "influential observations" (data errors, outliers, etc.). See the *Reference Manual* for details.

In the regression output, p-values are shown next to all test statistics produced. These p-values are the probability of seeing a test statistic at least this large when the null hypothesis is true. That is, small p-values imply regection. With the REGOPT procedure, you can obtain additional regression diagnostics, such as: Breusch-Godfrey LM test for autocorrelation, Ljung-Box Q-statistics, the augmented Dickey-Fuller test, ARCH(1) test, recursive residuals, Chow test for parameter stability, White or Breusch-Pagan heteroskedasticity tests, Shapiro-Wilks normality test, and Akaike and Schwarz Information Criteria (used for selecting lag length or sets of regressors). Try the REGOPT (PVPRINT, STARS, LMLAGS=2, QLAGS=2, BPLIST=C) ALL; command before using OLSQ to see the range of available diagnostics. See the *Reference Manual* for further details.

5.4. Two-stage least squares: 2SLS, INST

In the 2SLS statement, the dependent variable and independent variables are listed exactly as for OLSQ. However, they are preceded by the 2SLS options, which include the required list of instrumental variables. For example, in the illustrative model, government expenditures, G, the logs of the money supply, LM, and the trend variable, TIME, are considered as exogenous and may serve as instruments. The appropriate statement is

```
2SLS(INST=(C,G,TIME,LM)) CONS C,GNP;
```

Another way to write this is with the INST command (it is an alias of 2SLS). The INST command uses the keyword INVR or the bar | to mark the beginning of the list of instruments:

```
                                    Equation   4
                                    ============
                        Method of estimation = Instrumental Variable

Endogenous variables: GNP
Included exogenous variables: C
Excluded exogenous variables: G LM TIME

Dependent variable: CONS
Current sample:  1949 to 1975
Number of observations:  27

       Mean of dep. var. = 519.033           R-squared = .992999
  Std. dev. of dep. var. = 147.459  Adjusted R-squared = .992719
Sum of squared residuals = 3965.29       Durbin-Watson = .464998 [.000,.000]
   Variance of residuals = 158.612             E'PZ*E = 2908.96
Std. error of regression = 12.5941

            Estimated    Standard
Variable  Coefficient      Error       t-statistic   P-value
C          -19.7149      9.40495       -2.09622      [.036]
GNP        .636178       .010731       59.2860       [.000]
```

Example 5.2: Instumental Variable Estimation.

INST CONS,C,GNP INVR C,G,TIME,LM ; or INST CONS C GNP | C G TIME LM;

The constant C should always be used as an instrument. Two-stage least squares is 2SLS with all of the exogenous variables in the complete model (and no other variables) listed as instruments. Valid estimation can be based on a smaller number of instruments when the complete model involves many exogenous variables. Valid estimation can also be performed even when the rest of a simultaneous model is not fully specified. In these cases, the estimator is instrumental variables, but not really two-stage least squares. (The IV command is the same as 2SLS and can be used for these cases). Of course, there must be at least as many instruments as the number of right-hand side variables.

The output of 2SLS for the consumption function in the illustrative model is shown in **Example 5.2**. All statistics relating to residuals are calculated from the same formulas as OLSQ. The residuals are the "structural residuals", e = ŷ-Xb, not the "second-stage residuals", which would be obtained by doing two-stage least squares literally in two stages (this would involve replacing some column(s) of X with predicted values from a first stage).

Here are the equations used by 2SLS to compute the estimated coefficients and their standard errors. Let **Z** be the matrix of values of the instruments, and **y** and **X** be the dependent and independent variables. The coefficient estimates are

$$b = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y$$

The variance-covariance matrix of these coefficient estimates is given by

$$V(b) = s^2[X'Z(Z'Z)^{-1}Z'X]^{-1}$$

If the number of instruments Z is equal to the number of right-hand-side variables X, the classical instrumental variable estimator results:

$$b = [Z'X]^{-1}Z'y$$

This follows easily from the previous formula for b, since for this case

$$[X'Z(Z'Z)^{-1}Z'X]^{-1} = (Z'X)^{-1}(Z'Z)(X'Z)^{-1}$$

and four of the matrices cancel.

```
                                    Equation   5
                                    ============
                      Method of estimation = LIML

Endogenous variables: GNP
Included exogenous variables: C
Excluded exogenous variables: G LM TIME

Dependent variable: CONS
Current sample:  1949 to 1975
Number of observations:  27

       Mean of dep. var. = 519.033           R-squared = .992999
  Std. dev. of dep. var. = 147.459  Adjusted R-squared = .992719
Sum of squared residuals = 4063.75       Durbin-Watson = .465556 [.000,.000]
   Variance of residuals = 162.550  F (over-id. rest.) = 31.0720 [.000]
Std. error of regression = 12.7495  Lambda (eigenval.) = 3.70192

           Estimated    Standard
Variable  Coefficient     Error       t-statistic   P-value
C         -25.1644      9.69649       -2.59521      [.009]
GNP       .642613       .011077       58.0112       [.000]
```

Example 5.3: LIML

The objective function for 2SLS is not the ordinary sum of squared residuals, but rather the sum of squared residuals after projection onto the instruments (labeled E'PZ*E in the output):

$$e'P_Z e = e'Z(Z'Z)^{-1}Z'e$$

5.5. Limited information maximum likelihood: LIML

The LIML command computes maximum likelihood estimates for a linear model with endogenous variables on the right-hand side (and normally distributed disturbances). It is specified like the 2SLS command -- instruments are given in the options list. The advantages of LIML over 2SLS are asymptotic efficiency and a small sample distribution with less bias. The FULLER option allows for additional small-sample corrections (see the *Reference Manual* for details). The disadvantage of LIML is that the estimates are more sensitive to specification error in the lists of included and excluded variables. (See Section 7.3 for estimation of nonlinear LIML models.) For the illustrative model, we have:

```
LIML(INST=(C,G,TIME,LM)) CONS C,GNP;
```

The output from this command is shown in **Example 5.3**. The only difference from the output for OLSQ and INST is the presence of the F-statistic for overidentifying restrictions (two in this case, since there are three excluded exogenous variables and only one right-hand side endogenous variable) and the estimate of k, the variance ratio.

5.6. First-order serial correlation: AR1

When the disturbances of a linear regression model are correlated, the coefficient estimates of ordinary least squares are inefficient, and the standard error estimates are biased. The AR1 procedure provides several methods to obtain efficient estimates of an equation whose disturbances display first order serial correlation, that is, the observed disturbance is

$$u_t = e_t + \rho u_{t-1}$$

where e_t is uncorrelated. For a discussion of the efficiency gains from the serial correlation correction and some Monte-Carlo evidence, see Rao and Griliches (1969). To estimate the consumption function in the illustrative model with a correction for serial correlation, replace OLSQ with AR1:

```
AR1 CONS C,GNP ;
```

Beach and MacKinnon's (1978) estimation method is used by default. Called ML by TSP, this method includes the first observation with a special weight, rather than simply dropping it, and imposes the stationarity requirement on the estimate of the serial correlation parameter by requiring $\hat{\rho}$ to be between -1.0 and 1.0. The small sample properties of this estimator, particularly its RMSE, are preferable to those of the conventional GLS procedures.

The ML method cannot be used when there is a WEIGHT in the regression or when there are gaps in the sample. For these cases and for comparability with earlier versions of TSP or other programs, you may wish to use the CORC or HILU methods invoked as follows:

```
AR1(METHOD=CORC) CONS,C,GNP;

AR1(METHOD=HILU) CONS,C,GNP;
```

The CORC method uses a two stage iteration suggested by Cochrane and Orcutt (1949). This method estimates ρ from ordinary least squares residuals, transforms the dependent and independent variables so that the residuals from the transformed equation will be roughly serially uncorrelated, and then runs a regression using the transformed variables (the transformation is $\bar{x}_t = x_t - \hat{\rho} x_{t-1}$).This process repeats until $\hat{\rho}$ converges or the maximum number of iterations is reached.

Variations on the CORC method can be used to estimate nonlinear AR1 equations directly or AR(p) (p>1) processes. The FORM(NAR=p) and LSQ commands make this easy (see Section 7.2.1 for more details on LSQ and the *Reference Manual* for details on FORM). Here is an example of AR(2) estimation:

```
FORM(NAR=2) CAR2 CONS C GNP;
SMPL 51,75;                          ? DROP FIRST 2 OBSERVATIONS
LSQ CAR2;
```

The Cochrane-Orcutt procedure is asymptotically equivalent to the maximum likelihood procedure described earlier, but estimates may differ in small samples. When there are gaps in the sample (a SMPL of the form 1, 10, 21, 30, etc.), the CORC procedure drops the first observation in each group. Otherwise, it would use the wrong lagged variables for that observation. This technique may be used in estimating with a pooled time series-cross section sample by creating artificial gaps so that one unit variable is not used as a lagged value for the next unit. See AR1 in the *Reference Manual*, especially the TSCS option. See also Chapter 15, "Panel Data," in this manual.

The HILU method, Hildreth and Lu (1960), and the MLGRID method take a different approach to the AR estimation problem: They search over a user-specified set of ρ values for the one that gives the minimum standard error (maximum log likelihood in MLGRID) for the transformed equation. HILU drops the first observation in each sample group when computing the standard error. MLGRID retains it with a special weight, in a manner analogous to the ML method. Normally the values searched over are on the interval (- 0.9,0.9) in equidistant .1 steps. This may be modified by the user, as in this command:

```
AR1 (METHOD=HILU,RMIN=0.0,RMAX=0.5,RSTEP=.02) CONS,C,GNP;
```

If you wish to compute the regression for only one value of ρ, use the RSTART option and the HILU or MLGRID method. This option and others are described more fully in the *Reference Manual.*

All these methods produce the usual regression output; the summary statistics are based on the $\hat{\rho}$-transformed variables, while the plot displays the actual dependent variable (untransformed) and the corresponding fitted value, i.e.

$$\hat{y}_t = X_t * b + \hat{\rho}(y_{t-1} - X_{t-1} * b) = E(y_t | X_t, y_{t-1}, X_{t-1}, b, \hat{\rho})$$

Note that this is the same equation used to do a static forecast of the AR1 model (see Chapter 14). If the model is true, the residuals, $y_t - \hat{y}_t$, will be uncorrelated. These actual and fitted values may be displayed in a plot with the PLOTS option. When the CORC and HILU methods are used, the sample of observations is automatically adjusted to delete the first observation.

In addition to the coefficient estimates, $\hat{\rho}$ and its asymptotic standard error are also presented. The asymptotic covariance of the estimates of $\hat{\rho}$ and the regression coefficients is zero, unless lagged dependent variables appear on the right-hand side. In this case, TSP estimates ρ jointly with the other coefficients, so that the coefficient standard errors will not be biased. The estimation method is changed to CORC, since the lagged dependent variable would be correlated with the residual in the ML first observation. This technique is not implemented for models with weighting, instruments, or sample gaps.

The CORC and HILU methods may be combined with the weighted regression option described in Section 5.8. All of the results presented will be for a combined GLS transformation using both $\hat{\rho}$ and the WEIGHT variable, except the plot, which is based on the original data with a $\hat{\rho}$ transformation only (i.e., the residuals will be uncorrelated, but heteroskedastic).

5.6.1. Instrumental variable estimation in AR1

Two-stage least squares and instrumental variable estimation with a serial correlation correction may be obtained using AR1 with the INST option, although only the CORC and HILU methods of estimation are available. Several considerations in obtaining consistent estimates in this case are discussed by Fair (1970). To estimate the consumption function in our model by two-stage least squares, we use

```
AR1(INST=(C,G,LM,TIME)) CONS C,GNP;
```

Because Fair originally claimed that the lagged dependent and independent variables must be in the instrument list to obtain consistent estimates, AR1 automatically adds CONS(-1) and GNP(-1) to the list before estimating. To override this option and specify a complete list of instruments yourself, use the NOFAIR option:

```
AR1(NOFAIR,INST=(C,G,LM,TIME,GNP(-1),CONS(-1))) CONS C,GNP;
```

Fair retracted his claim in 1984; it has since been disproved by Buse (1989), but the alternate instruments required for consistency involve pseudo-differencing with the estimated ρ, which is tedious to do by hand. Buse also showed that the most efficient estimator asymptotically in this case also includes the lagged excluded exogenous variables, but he cautions that in small samples this may quickly exhaust the degrees of freedom.

5.7. Distributed lags

A distributed lag refers to the inclusion of more than one lag of a regressor (X variable) in a regression function. Many lags of a single variable may be easily included in the linear estimation procedures (OLSQ, 2SLS, LIML, AR1, PROBIT, TOBIT, but not LAD) of TSP. For example,

```
OLSQ CONS C  GNP GNP(-1)-GNP(-15) ;
```

specifies a regression of consumption on GNP and 15 quarters of GNP lagged for the illustrative model. Note that OLSQ prints the Schwarz Bayesian Information Criterion and/or the Akaike Information Criterion, which may be used to select the length of an unconstrained distributed lag. A regression like this (with many unconstrained lags) frequently produces coefficient estimates that are highly imprecise and jump around from period to period. This has led econometricians to invent ways of smoothing the coefficient estimates. TSP supports two of these techniques: polynomial distributed lags, invented by Almon (1965) and Shiller lags, due to Shiller (1973).

5.7.1. Polynomial distributed lags

You can include PDL variables in any linear equation in TSP by including a variable or variables with a lag specification in parentheses in the list of independent variables. This lag specification has the form (n,p,constraint), where

n = number of terms in the polynomial (degree plus one)

p = number of lags in the lag distribution
constraint = a keyword specifying the endpoint constraints:
BOTH (2 constraints), FAR (1 constraint), NEAR (1 constraint), NONE (no constraints)

The number of parameters associated with the distributed lag is n less the number of constraints; this must be positive and less than or equal to p. Note that the number of lags includes the current observation, i.e., the lagged variables are (t,t-1,t-2,..., t-p+1).

Examples:

```
OLSQ CONS C,GNP(4,16,FAR) ;
```

specifies a regression of CONS on a distributed lag of GNP covering 16 quarters, having terms up to the third power, and obeying the far endpoint constraint. The output is shown on the next page (**Example 5.4**).

```
OLSQ CONS C,GNP(4,16,FAR),R(4,24,NEAR) ;
```

specifies an additional distributed lag on R covering 24 quarters and a near endpoint constraint.

```
OLSQ CONS C,GNP(-1)(4,16,NONE);
```

forms a PDL of GNP that starts with the first lagged observation instead of the current observation.

When you use PDL variables in an instrument list, the choice of instruments depends on whether the variable is treated as exogenous or endogenous. If it is treated as exogenous, you want the scrambled variables described in the next section as instrumental variables. These could be specified by listing the names of the variables created by the "scrambling" process, i.e., in the above example, @PDL1, @PDL2, @PDL3. If PDL were treated as endogenous, it would suffice to ensure that there were enough instrumental variables in the list to meet the rank condition for identification -- at least as many as the number of variables in the regression after scrambling.

5.7.2. What PDL does

Each PDL term stands for a distributed lag of the form

$$b_0x_t + b_1x_{t-1} + \ldots + b_{p-1}x_{t-p+1}$$

The polynomial specification requires that b_i lie on a polynomial in i:

$$b_i = a_i + a_2\frac{(i+1)}{(p+1)} + \ldots + a_n\left[\frac{(i+1)}{(p+1)}\right]^{(n-1)}$$

This implies constraints on the coefficients of x and its lags imposed by defining a new set of n variables z_t that are linear functions of the x variables. The actual functions differ slightly from what is implied by the above equations; see the Cooper (1972) article for details (Lagrangian interpolation polynomials are used). The variables $z_{1t},\ldots,z_{nt}$ enter the regression as ordinary independent variables (labeled @PDL1, @PDL2, and so forth in the regression results). Their coefficients are estimates of the parameters $a_1,\ldots,a_n$. After running this regression, TSP computes the corresponding estimates of the lag distribution coefficients b_0, $b_1,\ldots,b_{p-1}$ from the polynomial. The standard errors of the b_i are computed from the covariance matrix of the a_i. In addition, the mean lag and the sum of the lags are presented along with their standard errors. The expressions defining them are

Example 5.4 Polynomial distributed lags:

See *****linear-output-graphs.5.wpd

sum lag = $b_0 + b_1 + ... + b_{p-1}$
mean lag = $[b_0 0 + b_1 1 + b_2 2 + ... + b_{p-1}(p-1)]/(\text{sum lag})$

If the lag coefficients change sign, the mean lag computed by this formula is meaningless.

If endpoint constraints are imposed, the procedure is modified slightly. The near constraint sets a hypothetical b_{-1} to zero. TSP imposes this by dropping the first variable from the regression. The far constraint sets a hypothetical b_p to zero. TSP imposes this by subtracting the last variable from each of the others and dropping the last variable from the regression.

5.7.3. Shiller lags

Constraining lag coefficients to lie exactly on a low-order polynomial often seems quite arbitrary, since there is no *a priori* reason for the functional form. The Shiller lag technique is a method for imposing smoothness on the lag coefficients without using an exact polynomial. The idea of the method is to impose a "smoothness" prior on the coefficients and estimate using Bayesian techniques. For a full discussion of the method, see Shiller (1973).

To use the Shiller lag procedure, you must specify a degree of differencing (this is similar to the polynomial degree specified in PDL), the number of lags of the X variable to include in the regression, the endpoint restrictions, and the value of the smoothness prior that you wish to use. For example, the command

```
OLSQ CONS C,GNP(3,16,NONE,.0025)
```

specifies a 2nd degree Shiller lag for GNP with no endpoint restrictions and a prior variance on the differenced coefficients equal to .0025. The output from this command is shown on the next page (**Example 5.5**).

The prior variance on the differenced lag coefficients is what controls the smoothness of the distribution; a prior variance equal to zero yields PDL estimates, while a very large prior variance will give unconstrained lag coefficients. Other values yield answers between the smoothness of a polynomial and the jagged nature of unconstrained lags. One possibility is to use the sample variance of the unconstrained lag coefficients as a prior; this option is provided if you use a -1 as the prior variance.

One way to interpret Shiller lags is as a generalization of the PDL method. For example, the command

```
OLSQ CONS C,GNP(4,16,FAR,0)
```

estimates the same regression as the first example in section 4.7.1, since it specifies that the fourth degree differences of the coefficients should be precisely zero (with no error). This guarantees that all the coefficients will lie on a third degree polynomial, as for a PDL.

5.8. Weighted regression: the WEIGHT option

The weighted least squares estimator is used when the variance of the disturbances in a regression differs across observations. It is most frequently used in cross-section regressions where the units of observation differ in scale or size -- for example, a cross section of states in the United States.

You can obtain weighted least squares in TSP by using the WEIGHT option in any of the linear regression procedures (OLSQ, INST, or AR1). WEIGHT= should appear as an option in parentheses between the command name and the name of the dependent variable. After the = sign, put the name of a series whose values are proportional to the inverses of the variances of the disturbances in the regression. In cross section across political units, the weight variable should be the reciprocals of populations if the variables are sums or aggregates, and should equal population if the variables are measured per capita. For example, if YOUNG is the fraction of young adults living by themselves (a per capita variable) and POP is population,

Example 5.5 Shiller lags

See ******linear-output-graphs2.5.wpd

weighted least squares estimates to the weighted data, and unweighted residuals, computed by applying the weighted least squares estimates to the original data.

Note that both are based on the weighted least squares coefficients, which are minimum variance among all linear unbiased estimators if the weights are correct. The weighted residuals satisfy the usual properties of residuals and are the proper basis for statistical testing. The unweighted residuals describe the departure of the actual data from the regression function in their original units. TSP computes all the standard regression statistics for the weighted residuals and a subset of them for the unweighted residuals.

5.8.1. Normalization of weights

TSP divides the weights provided by the user by a constant so that the sum of the weights equals the number of observations. Observations given weight zero are not counted. This normalization does not affect the regression coefficients or most of the statistics. It leaves the magnitudes of the weighted data and weighted residuals the same, on average, as the unweighted data and residuals. Normalization may be suppressed with the UNNORM option. With UNNORM, if the weights are all large, the residuals, sum of squared residuals, residual variance, standard error, and implied number of observations will all be correspondingly large.

5.8.2. Weighted descriptive statistics

The WEIGHT= option may be used with any MSD statement to obtain weighted means, standard deviations, covariances, correlations, and moments. The data are multiplied by the square roots of the weight variable before the statistics are computed. The weights are normalized so that they sum to the number of observations; the UNNORM option is not available. Here is an example of using the WEIGHT= option:

```
MSD (WEIGHT=POP,CORR) YOUNG,C,URBAN,CATHOLIC,ABORTION,SERVEMP,SOUTH ;
```

5.9. Robust standard errors in the regression procedures

When you know the form and size of the heteroskedasticity in your data, you can use WEIGHT to obtain consistent estimates of the standard errors of your model. However, this is often not the case. The ROBUST option causes TSP to compute standard errors that are consistent even in the presence of unknown heteroskedasticity, by using the data to estimate its magnitude. Econometric references for this technique are White (1980a), White (1982), and Chamberlain (1984). Note that in general, White gives two terms in his formulae, the second of which vanishes if the model is correctly specified up to an additive error. TSP computes only the first term.

To obtain these heteroskedastic-consistent estimates of the standard errors of your model, include the option ROBUST on an OLSQ, 2SLS, or LSQ statement (see Chapter 7). For example,

```
OLSQ(ROBUST) CONS C GNP ;
```

computes the standard errors and variance estimate shown below for our illustrative example (compare to the conventional estimates in **Example 5.1**).

```
            Estimated    Standard
Variable  Coefficient      Error       t-statistic   P-value
C          -17.8024       11.3535       -1.56800      [.129]
GNP        .633919        .013560       46.7507       [.000]
Standard Errors are heteroskedastic-consistent (HCTYPE=2).
```

For standard errors that are robust to autocorrelation, use the GMM(NMA=n) command. For example:

```
FORM HAC CONS C GNP;
```

```
GMM(HET,NMA=2,INST=(C,GNP)) HAC;
```

5.10. Least absolute deviations regressions (LAD)

The ordinary least squares estimator is optimal when the disturbance in the equation is normally distributed. But when the disturbance is not normally distributed, other estimators are better. If the distribution is known, the efficient estimator is maximum likelihood with the correct distribution function. However, in many cases, you may suspect that your data distribution is "fat-tailed" or contains outliers, without knowing exactly its form. In this setting, the LAD estimator, which minimizes the sum of absolute deviations of the residuals, may be more efficient. The LAD estimator is also known as the L1 regression, least absolute residual (LAR), least absolute error (LAE), and minimum absolute deviation (MAD). See Chapter 22 of Judge et al (1988) for a discussion of its properties.

To estimate by least absolute deviations in TSP, use the LAD command like the OLSQ command. For example,

```
LAD CONS,C,GNP ;
```

estimates the consumption function using L1 regression instead of L2 (least squares) regression. The estimates are efficient if the disturbances have a Laplace distribution; and have a smaller variance than least squares if the disturbances have any of the fat-tailed distributions, such as Cauchy and Student's t. The standard error estimates are computed as though the true distribution of the errors was Laplace. See the *Reference Manual* for details of the computations and options for quantile regression.