III. The Generalized Regression Model and Related Topics

1. The Model

2. Estimators of the coefficient vector (β)
- **a. Least squares**
  - **(1) Sufficient condition**
  - **(2) Distribution of OLS estimator**
- **b. Maximum likelihood**
  - **(1) Sufficient conditions for a maximum**
  - **(2) Relationship to BLUE**
  - **(3) Distribution of the MLE of β**
  - **(4) Cramer-Rao matrix**
  - **(5) Efficiency**
  - **(6) MLE using OLS**
  - **(7) Likelihood Ratio tests**
- **c. Alternative estimators**

3. Forecasting

4. A review of some important special cases
- **a. No autocorrelation with homoskedastic disturbances**
  - **(1) Coefficient estimators**
  - **(2) Variance estimators**
  - **(3) Hypothesis testing**
    - **(a) Simple t-tests**
    - **(b) F-tests**
    - **(c) Chow tests**
    - **(d) Likelihood ratio tests**
  - **(4) Forecasting**
  - **(5) Consequences of using least squares when $\Sigma \neq \sigma^2 I$**
- **b. Heteroskedasticity**
  - **(1) Definition**
  - **(2) Statistical tests**
  - **(3) Estimation**
- **c. Autocorrelation**
  - **(1) Definition**
  - **(2) Statistical tests**
  - **(3) Estimation**
  - **(4) Forecasting**

5. Autoregressive conditional heteroskedasticity (ARCH) models
- **a. Introduction -- an example**
- **b. Estimation**
- **c. Test for ARCH disturbances**

6. **Stochastic Regressors**

7. **Instrumental variables**
   a. **Background**
   b. **Instrumental variables estimator**
   c. **Some special cases**
      (1) **Least squares**
      (2) **Generalized least squares**
      (3) **Projection of X on Z**
      (4) **Stochastic regressors**
   d. **Selection of instrumental variables**
   e. **Instrumental variables in quantile regression models**
   f. **Some other issues–weak instruments, application references**

8. **Specification tests**
   1. **Background**
   2. **Hausmann test**
   3. **Applications**

9. **Seemingly unrelated regression models (Sure-Zellner)**

10. **Models for Panel Data: cross sectional and time series**
    a. **OLS and GLS**
    b. **Random and fixed effects models**
    c. **Differences in differences**
    d. **Statistical inference**

11. **Regression discontinuities**

12. **Exercise set**

James McDonald
Brigham Young University
12/2011

The Generalized Regression Model

1. The Model

Let Y_t, X_{ti} denote endogenous (dependent) and exogenous (independent) variables.

Consider the relationship

$$Y_t = \beta_1 + \beta_2 X_{t2} + \ldots + \beta_K X_{tK} + \varepsilon_t \qquad (1.1)$$

$$t = 1, 2, \ldots, N, \; N > K;$$

(1.1) can be represented in terms of matrices as

$$Y = X\beta + \varepsilon \qquad (1.1)'$$

where

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & X_{12} & \cdots & X_{1K} \\ \vdots & \vdots & & \vdots \\ 1 & X_{N2} & \cdots & X_{NK} \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} \text{ and } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}.$$

It will also be assumed that

(A.1) $\varepsilon \sim N[0,\Sigma]$,

(A.2) the X's are nonstochastic and

$$\underset{N\to\infty}{\text{Limit}} \frac{(X'\Sigma^{-1}X)}{N} = \Sigma_x$$

which is nonsingular.

(A.1) implies the random disturbances are distributed normally, each with mean zero and

$$\mathrm{Var}(\varepsilon) = \begin{bmatrix} \mathrm{Var}(\varepsilon_1) & \mathrm{Cov}(\varepsilon_1,\varepsilon_2) & \cdots & \mathrm{Cov}(\varepsilon_1,\varepsilon_N) \\ & \mathrm{Var}(\varepsilon_2) & \cdots & \mathrm{Cov}(\varepsilon_2,\varepsilon_N) \\ & \ddots & & \vdots \\ & & & \mathrm{Var}(\varepsilon_N) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix}$$

$$= \Sigma.$$

Note that this specification doesn't require that the random disturbances are independent of each other (autocorrelation is possible) nor that the variances of the random disturbances are the same (heteroskedasticity is possible); however this specification allows for homoskedastic and nonautocorrelated errors if $\Sigma = \sigma^2 I$.

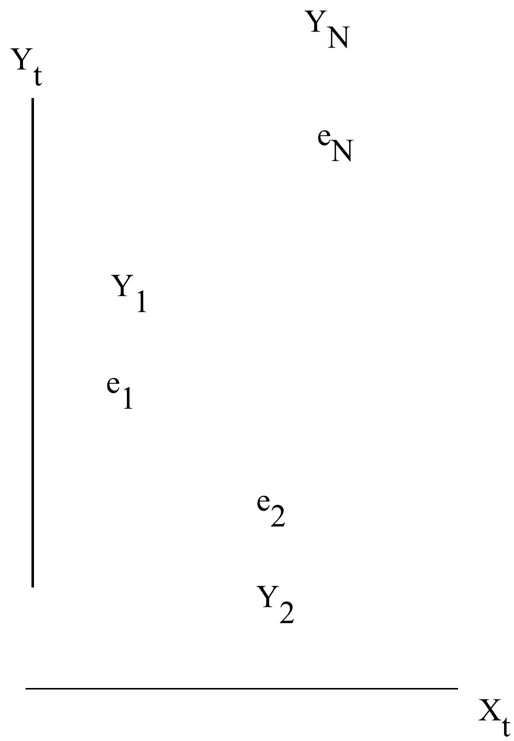
The unknown parameters in the generalized regression model are

(1) the coefficients β

(2) the variance-covariance matrix Σ.

2. Estimators of the Coefficient Vector (β)

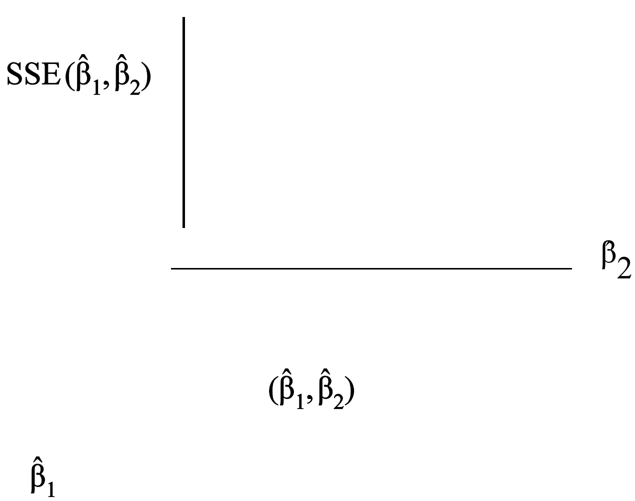
a. Least Squares. The least squares technique is based upon the principle of selecting an estimator of β, which minimizes the associated sum of squared errors--vertical deviations between the observed dependent variables and the estimated regression line (plane).



$$Y = X\beta + \varepsilon$$

$$= X\hat{\beta} + e$$

$$= \hat{Y} + e$$



The sum of squared errors can be expressed as

$$SSE(\hat{\beta}) = \Sigma e_t^2 \qquad \text{(2.1 a-d)}$$

$$= (e_1,\ldots,e_N) \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix}$$

$$= e'e$$

$$= (Y - X\hat{\beta})' \; (Y - X\hat{\beta})$$

depends upon the estimators of β. The least squares estimators are obtained by minimizing $SSE(\hat{\beta})$ with respect to the vector $\hat{\beta}$, i.e.,

$$\underset{\hat{\beta}}{\text{minimize}} \; SSE\,(\hat{\beta})$$

$$= \min_{\hat{\beta}} \; (Y - X\hat{\beta})' \; (Y - Y\hat{\beta}) \; .$$

After expanding the expression (2.1d) for the sum of squared errors, we obtain

$$SSE(\hat{\beta}) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \; .$$

Differentiating this expression with respect to $\hat{\beta}$ yields the <u>necessary</u> conditions for a solution to (2.2):

$$\frac{dSSE(\hat{\beta})}{d\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

or

$$\boxed{(X'X)\hat{\beta} = X'Y} \qquad (2.3)'$$

(2.3)' is also referred to as the system of normal equations.

If $|X'X| \neq 0$, (2.3)' can be solved to yield the least squares estimator

$$\boxed{\hat{\beta} = (X'X)^{-1}X'Y} \qquad (2.4)$$

Note: (1) The sufficient condition for (2.4) to be a solution to (2.2) is that

$$\frac{d^2SSE}{d\hat{\beta}d\hat{\beta}'} = 2(X'X)$$

is positive definite.

(2) The least squares estimator is distributed normally as

$$\boxed{\hat{\beta} \sim N(\beta;(X'X)^{-1}X'\Sigma X(X'X)^{-1})}$$

Proof. (a) See I.B.2 for a "useful" theorem

(b) Alternatively,

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'(Y) = (X'X)^{-1}X'(X\beta+\varepsilon)\\ &= \beta + (X'X)^{-1}X'\varepsilon.\end{aligned}$$

Therefore, $E(\hat{\beta}) = \beta$

$$Var(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$

$$= E(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}$$

$$= (X'X)^{-1}X'\Sigma X(X'X)^{-1} .$$

b. <u>Maximum Likelihood Estimators</u>

The likelihood function associated with (1.1)' is given by

$$L(Y;\beta,\Sigma) = \frac{e^{(Y-X\beta)'\Sigma^{-1}(Y-X\beta)/2}}{(2\pi)^{N/2}\ |\Sigma|^{1/2}}$$

and the log likelihood function is given by

$$\ell(Y;\beta,\Sigma) = \ln(L(Y;\beta,\Sigma)) \tag{2.7}$$

$$= (-1/2)(Y - X\beta)'\Sigma^{-1}(Y - X\beta) - (½)[N \ln(2\pi) + \ln|\Sigma|]$$

$$= (-1/2)(Y'\Sigma^{-1}Y - 2\beta'X'\Sigma^{-1}Y + \beta'X'\Sigma^{-1}X\beta) - (½)[N\ln(2\pi) + |\Sigma| \ln].$$

The maximum likelihood estimator (MLE) of β is obtained by maximizing (2.6) or

(2.7) with respect to β, i.e.,

$$\underset{\beta}{\text{Max}}\ \ell(Y;\beta,\Sigma). \tag{2.8}$$

The necessary conditions are given by

$$\frac{d\ell}{d\beta} = -[-X'\Sigma^{-1}Y + X'\Sigma^{-1}X\tilde{\beta}] = 0$$

or

$$\boxed{X'\Sigma^{-1}X\tilde{\beta} = X'\Sigma^{-1}Y} \quad .$$

(2.10) is sometimes referred to as the <u>system of modified normal equations</u>.

If $|X'\Sigma^{-1}X| \neq 0$, the maximum likelihood estimator is the solution to (2.10),

$$\boxed{\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y}$$

.

This estimator is also referred to as Aitken's estimator and also as the generalized or weighted least squares (GLS) estimator.

Note: (1) The sufficient condition for β to yield a maximum of (2.7) is for

$$\frac{d^2\ell}{d\beta d\beta'} = -X'\Sigma^{-1}X$$

to be negative definite.

(2) $\tilde{\beta}$ can also be shown to be the best linear unbiased estimator (BLUE) of β.

(3) $\tilde{\beta}$ is distributed normally and

$$\boxed{\tilde{\beta} \sim N(\beta; (X'\Sigma^{-1}X)^{-1});}$$

hence, the least squares and MLE are both unbiased estimators.

The variances of $\hat{\beta}$ and $\tilde{\beta}$ may be different if $\Sigma \neq \sigma^2 I$.

(4) The Cramer-Rao matrix is given by

$$\Sigma^*_{\tilde{\beta}} = -\left[E\frac{\partial^2 \ell}{\partial\beta\partial\beta'}\right]^{-1}$$

$$= (X'\Sigma^{-1}X)^{-1}$$

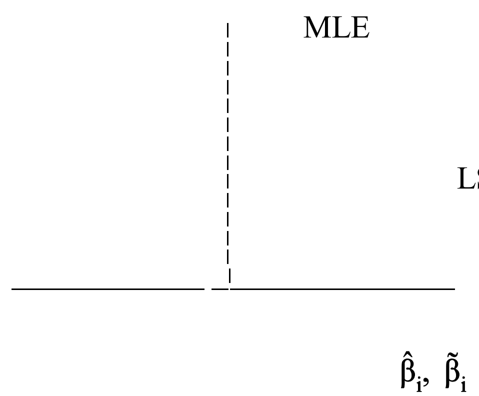
which is the variance covariance matrix of $\tilde{\beta}$.

(5) $\mathrm{Var}(\hat{\beta}) - \mathrm{Var}(\tilde{\beta})$

$$= (X'X)^{-1}X'\Sigma X(X'X)^{-1} - (X'\Sigma^{-1}X)^{-1}$$

is a positive semi-definite matrix. This implies that the variance of

$\mathrm{Var}(\hat{\beta}_i) \geq \mathrm{var}(\tilde{\beta}_i)$.



(6) MLE Using OLS

If a matrix T can be found such that $T\Sigma T' = \sigma^2 I$ $(T'T = \Sigma^{-1})$, then

the MLE of β can be obtained by estimating

$$TY = TX\beta + T\varepsilon$$

using least squares.

(7) Likelihood Ratio (LR) test

MLE lends itself to testing hypotheses of the form $q(\beta) = 0$ where $q(\beta)$ denotes an r x 1 vector of continuous functional constraints on the vector β, e.g.

$$q(\beta) = \begin{pmatrix} \beta_2\beta_3 - 1 \\ \beta_4 \\ \beta_1 + \beta_2 - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This hypothesis can be tested by estimating the model with and without the constraints imposed. Let the corresponding log likelihood values be denoted by ℓ^* and ℓ.

$$LR = 2(\ell - \ell^*)$$

provides the basis for testing H_o: $q(\beta) = 0$ and is asymptotically distributed as a chi-square,

$$\boxed{LR = 2(\ell - \ell^*) \overset{a}{\sim} \chi^2 (r)}$$

c. <u>Alternative estimators</u>

There are many alternatives to least squares and MLE. These include, among others:

Least absolute deviation (LAD) estimators

$$\underset{\beta}{\text{Min}} \sum_{t=1}^{n} |\varepsilon_t|;$$

L_p estimators

$$\min_{\beta} \sum_{t=1}^{n} |\varepsilon_t|^p;$$

M estimators

$$\min_{\beta} \Sigma\ \rho(\varepsilon_t);$$

among others. Each of these estimators works well for errors from a particular distribution, but not necessarily for others. There is considerable interest in estimators which are "robust" over many distributions. See the Appendix to this section for additional discussion of this material and some examples.

3. Forecasting

Recall that the model under consideration is given by (1.1).

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{t2} + \ldots + \beta_k X_{tk} + \varepsilon_t \qquad (3.1) \\ &= (1, X_{t2}, \ldots, X_{tk})\, \beta + \varepsilon_t \\ &= X_t\, \beta + \varepsilon_t. \end{aligned}$$

Goldberger (1962, JASA) demonstrated that the minimum variance h-period ahead unbiased predictor of Y is given by

$$\boxed{Y_N(h) = \tilde{Y}_{N+h} = X_{N+h}\, \tilde{\beta} + W'\Sigma^{-1}e} \qquad (3.2)$$

where

- $\tilde{\beta}$ denotes the MLE of β

- $e = Y - X\tilde{\beta}$, the estimated residual vector and

- $W' = E(\varepsilon_{N+h}\,\varepsilon)$

$$= E\begin{bmatrix} \varepsilon_{N+h} & \varepsilon_1 \\ \vdots & \\ \varepsilon_{N+h} & \varepsilon_N \end{bmatrix} = \begin{bmatrix} \mathrm{cov}(\varepsilon_1,\varepsilon_{N+h}) \\ \vdots \\ \mathrm{cov}(\varepsilon_N,\varepsilon_{N+h}) \end{bmatrix}$$

- N = Sample size associated with the sample used in estimating β.

Note that if the error terms are uncorrelated, $E(\varepsilon_t\varepsilon_s) = 0$ for $t \neq s$, W = 0; hence (3.2) simplifies to

$$\boxed{Y_N(h) = \tilde{Y}_{N+h} = X_{N+h}\tilde{\beta}} \tag{3.3}$$

for random disturbances which are not autocorrelated.

4. Some Important Special Cases

a. <u>No autocorrelation with homoskedastic random disturbances</u>

The model can then be written as

$$Y = X\beta + \varepsilon \tag{4.1}$$

(A.1) $\varepsilon \sim N(0,\sigma^2 I)$

(1) Coefficient Estimators

Exercise: Demonstrate that the least squares and MLE of β are identical in this case and

$$\boxed{\hat{\beta} = \tilde{\beta} = (X'X)^{-1}X'Y \sim N(\beta; \sigma^2 (X'X)^{-1})}$$

This model is generally referred to as the classical normal linear regression model. The least squares estimators of β, $\hat{\beta}$, will be

- unbiased
- minimum variance of all unbiased estimators (hence BLUE)
- consistent
- asymptotically efficient
- normally distributed

(2) Variance Estimators

- An unbiased estimator of σ^2 is given by

$$s^2 = \sum e_t^2 / (N - K) \tag{4.4}$$

$$= e'e / (N - K)$$

Exercise: $\dfrac{(N - K)s^2}{\sigma^2} \sim \chi^2(N - K).$

- An unbiased estimator of var($\hat{\beta}$) = $\sigma^2(X'X)^{-1}$ is given by

$$s^2(X'X)^{-1} \tag{4.5}$$

(3) Hypothesis Testing

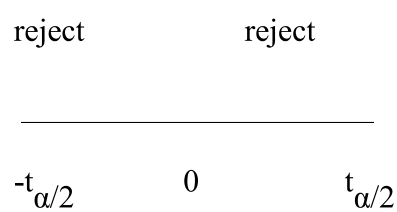
This form of the model facilitates testing numerous hypotheses of interest, e.g.

(a) Ho: $\beta_i = \beta_i^o$

This hypothesis can be tested using a t-test

$$\frac{\hat{\beta}_i - \beta_i^0}{s_{\hat{\beta}_i}} \sim t(N - K)$$

where the pdf of the t-statistic and corresponding critical values (2-tailed test) appear as in the following figure:

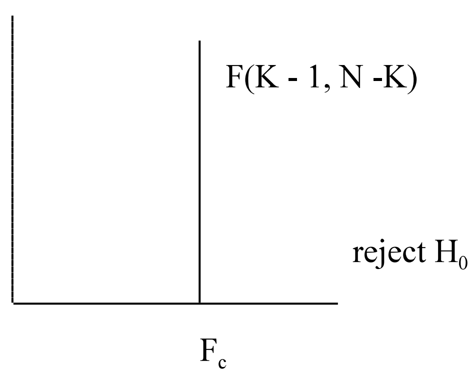


(b) H_o: $\beta_2 = \beta_3 = \ldots = \beta_K = 0$

This test of the overall "explanatory" power of the model can be

performed with an F test.

$$F = \frac{SSR/K - 1}{SEE/N - K} = \left(\frac{R^2}{1 - R^2}\right)\left(\frac{N - K}{K - 1}\right)$$

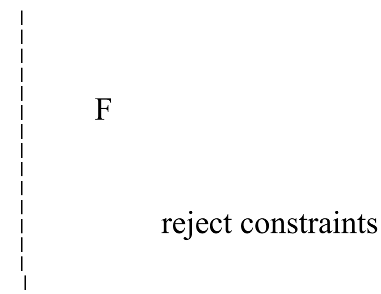


(c) The <u>Chow test</u> can be used to test these hypotheses as well as more

general constraints on the coefficient vector. The test statistic is based upon the results of estimating the model without the constraints imposed and again with the constraints imposed.

The test statistic is defined by

$$\frac{\dfrac{SSE* \; - \; SSE}{r}}{\dfrac{SSE}{N \; - \; K}} \sim F(r, \; N \; - \; K)$$



where the * denotes the results from estimating the constrained model, and r = (N-K)* - (N-K) is the number of independent restrictions imposed by the hypotheses.

(d) Likelihood Ratio Test

The log likelihood value for the special case of homoskedastic and independent (not autocorrelated) residuals is given by

$$\ell(\beta,\sigma^2) = -SSE/2\sigma^2 - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2)$$

The concentrated (over σ^2) likelihood function is obtained from (4.9) by replacing σ^2 by its MLE, SSE/N, to obtain

$$\ell_c(\beta) \;=\; -\frac{N}{2} \;-\; \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(SSE/N)$$

$$\boxed{\ell_t(\beta) \;=\; -\frac{N}{2}\left(1 + \ln(2\pi) + \ln\left(\frac{SSE}{N}\right)\right).}$$

The likelihood ratio (LR) test is obtained by estimating the model with and without the constraints imposed to yield ℓ^* and ℓ, respectively. The LR test statistic is constructed by taking twice the difference of ℓ and ℓ^* and has an asymptotic χ^2 distribution with degrees of freedom equal to the number of independent constraints, i.e.,

$$\boxed{LR \;=\; 2(\ell - \ell^*) \overset{a}{\sim} \chi^2(r)}$$

For normally distributed errors and

• σ known $\quad LR \;=\; \dfrac{SSE^* - SSE}{\sigma^2}$

• σ^2 unknown

$$LR = N \ln(SSE^*/SSE). \tag{4.10''}$$

(4.10') corresponds to the case of known variance and is quite similar in structure to the Chow test in (4.8) (recall SSE/N-K is an unbiased estimator of σ^2).

(4) Forecasting (conditional on given X's)

From (3.2) and (3.3) we see that the best linear unbiased forecasts of Y, given X's, are given by

$$\tilde{Y}_{N+h} \;=\; X_{N+h}\tilde{\beta}. \qquad (4.$$

$\tilde{Y}_{N+h}$ is distributed normally with mean $X_{N+h}\beta$ and variance

$$\mathrm{var}(\hat{Y}_{N+h}) = \sigma^2_{\hat{Y}} = X_{N+h}\sigma^2(X'X)^{-1}X'_{N+h}.$$

which can be estimated by

$$s^2_{\hat{Y}} = X_{N+h}s^2(X'X)^{-1}\ X'_{N+h}.$$

The Forecast error is defined to be the difference between the predicted and observed values for Y, i.e.,

$$FE = Y_{N+h} - \hat{Y}_{N+h}$$

The forecast error is distributed normally with mean 0 and variance

$$\sigma^2_{FE} = \underbrace{\sigma^2}_{\substack{\text{measure of}\\ \text{uncertainty}\\ \text{of } \varepsilon_t}} + \underbrace{\sigma_{\hat{Y}}^2}_{\substack{\text{measure of}\\ \text{uncertainty}\\ \text{of } X_{N+h}\beta}} \tag{4.15}$$

σ^2_{FE} can be estimated by

$$s^2_{FE} = s^2 + s^2_{\hat{Y}}. \tag{4.}$$

In summary,

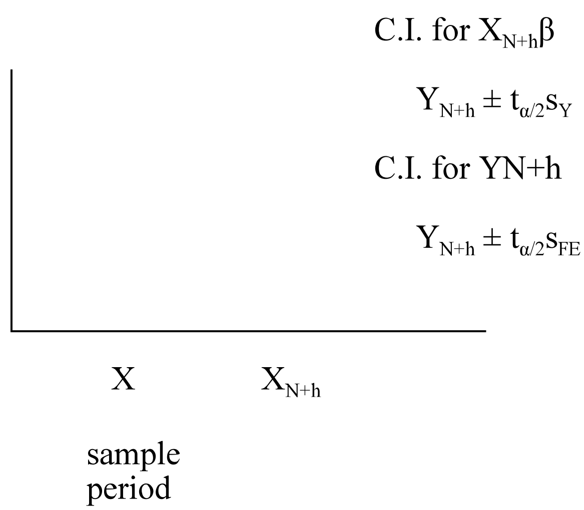
$$\boxed{\begin{array}{l} \tilde{Y}_{N+h} \sim N(X_{N+h}\beta,\ \sigma_{\hat{Y}}^2) \\ FE = Y_{N+h} - \hat{Y}_{N+h} \\ \sim N(0,\ \sigma_{FE}^2 = \sigma^2 + \sigma_{\hat{Y}}^2) \end{array}} \tag{4.18}$$

Confidence intervals for the regression line, $X_{N+h}\beta$, and for the actual value of Y, Y_{N+h} respectively are given by

$$\boxed{X_{N+h}\beta:\ X_{N+h}\hat{\beta} \pm t_{\alpha/2}\, s_Y}$$

$$\boxed{Y_{N+h}:\ X_{N+h}\hat{\beta} \pm t_{\alpha/2}\, s_{FE}}$$

where $t_{\alpha/2}$ denotes the critical value for a t-statistic with N-K degrees of freedom at the α-level of significance. This can be graphically depicted as follows:



(5) Consequences of using least squares estimation when $\Sigma \neq \sigma^2 I$.

(a) Least squares estimators will still be unbiased, consistent, and normally distributed, but will not be minimum variance estimators (∑-known).

(b) The t and F statistics reported will not be distributed as t(N - K) and F(K - 1, N - K) because the wrong formulas $(s^2(X'X)^{-1})$ will be used to estimate the variance $(X'X)^{-1}X'\Sigma X(X'X)^{-1}$ $\hat{\beta}$. Hal White has proposed a consistent estimator of the var-covariance matrix which will yield asymptotically appropriate t-statistics in this case. The command is operational in Shazam by listing HETCOV as an option to OLS.

(c) $\text{Var}\,(\hat{Y}_{OLS}) \geq \text{Var}\,(\tilde{Y}_{MLE})$

*<u>For these reasons and others it is very important to perform tests of the assumptions of the model</u>.

b. <u>Heteroskedastic Random Disturbances</u>

(1) <u>Definition</u>:

The heteroskedastic regression model corresponds to the situation in which the variances of the random disturbances are not constant. This situation frequently arises when working with cross-sectional data.

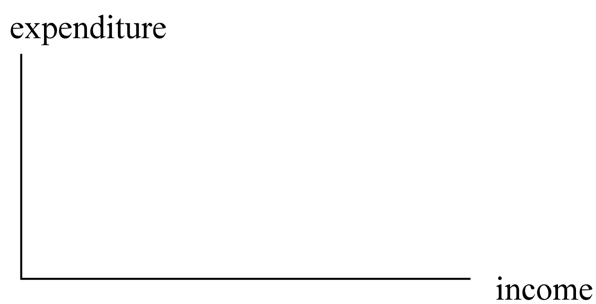
The model can be written as

$$Y = X\beta + \varepsilon$$

where $\varepsilon \sim N\,[0,\Sigma]$

$$\text{and} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \sigma_n^2 \end{bmatrix} .$$

As an example, cross sectional analysis of consumption patterns often appear in the form:



Larger income levels are frequently seen to be associated with greater variation in expenditure levels.

(2) Statistical tests

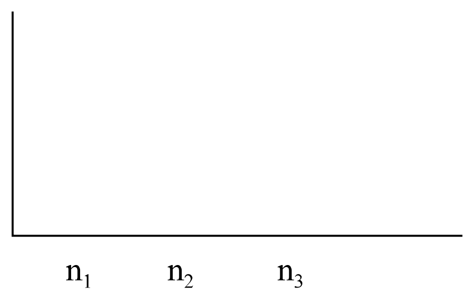
There are many tests of the hypothesis of homoskedastic random disturbances,

$$H_0: \ \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_N^2$$

These are based upon an analysis of estimated random disturbances (see exercise set number 1.b for a precautionary note).

The tests check for systematic behavior in the magnitudes of the random disturbances (or variances) and include the Goldfeld Quandt test, the Bartlet test, graphical analysis, Park test, rank correlation tests, Glejser test and White test.

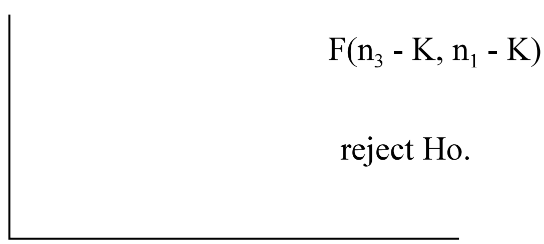
The Goldfeld Quandt test is an exact test (many of the others are asymptotic tests). The data is divided into three groups of approximately equal size.



Separate regressions are run on the first and third groups, yielding s_1^2 and s_3^2.

Under the hypothesis of constant variances (homoskedasticity)

$$\boxed{\frac{s_3^2}{s_1^2} \sim F(n_3 - K, n_1 - K)}$$



The second group isn't used for statistical reasons (power of the test) and separate regressions are used so that s_3^2 and s_1^2 will be distributed independently and the test statistic will be distributed as an F statistic. The larger of s_1^2 and s_3^2 is placed in the numerator of the test statistic.

Other tests for heteroskedasticity are based on an attempt to estimate relationships of the form $\sigma_t^2 = f(X_t)$. Recall that homoskedasticity implies that $\sigma_t^2 = f(X_t) = \sigma^2$ and σ_t^2 (even if observed) will not depend on X_t. Recall that $\sigma_t^2 = E(\varepsilon_t^2)$ and neither ε_t, σ_t^2 nor f( ) are observed.

- $e_t^2 = (Y_t - X_t\hat{\beta})^2$ is often used as a proxy for σ_t^2 (see eqn. for the notation used here).
- Alternative function forms for $f(X_t)$ have been considered in the literature.

The White test for heteroskedasticity [Econometrica, 1980, pp. 817-38] is based on using a second order Taylor Series approximation for the unknown function $f(X_t)$. This test is performed by regressing the squares of the OLS residuals (e_t^2) on an intercept, each of the X's, the squares and cross products of the X's. A Lagrangian multiplier (LM) test is used to test for the collective explanatory power of the X's, i.e.,

$$\boxed{LM = nR^2 \overset{a}{\sim} \chi^2\left(\frac{(K-1)(K+2)}{2}\right)}.$$

For example, if the original model was

$$Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + \varepsilon_t$$

the White test would involve using the OLS residuals

$$e_t^2 = (Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{t2} - \hat{\beta}_3 X_{t3})^2$$

in the regression

$$e_t^2 = \delta_1 + \delta_2 X_{t2} + \delta_3 X_{t3} + \delta_4 X_{t2}^2 + \delta_5 X_{t3}^2 + \delta_6 X_{t2} X_{t3}$$

and using

$$LM = nR^2 \overset{a}{\sim} \chi^2(5).$$

A lack of statistical significance of LM is consistent with the null hypothesis of homoskedasticity.

The Modified White Test is similar to the White Test except that the squares of the estimated residuals are regressed on the predicted Y's and the squares of the predicted Y's. The corresponding LM test is distributed as a Chi-square with two degrees of freedom. An obvious advantage of the Modified White Test, over the White Test, is in applications characterized by many explanatory variables.

The Park and Glejser tests correspond to using $|e_t|$ as a proxy for σ_t (or e_t^2 for σ_t^2) and then relationships of the form $|e_t| = f(X_t)$ or $e_t^2 = g(X_t)$ are estimated for various functions f( ) or g( ).

Rank correlation tests are based upon correlation between the magnitude of $|e_t|$ and the explanatory variables.

In Shazam, the command "DIAGNOS/HET CHOWTEST", immediately following the OLS estimation command, performs several tests for heteroskedasticity.

(3) Estimation.

Least squares estimation attributes equal weights to each observation and does not yield minimum variance estimators in the case of heteroskedastic errors. MLE attributes less "weight" to observations associated with large variances and can be obtained by applying least squares to

$$Y_t/\sigma_t = \beta_1(1/\sigma_t) + \beta_2(X_{t2}/\sigma_t) + \ldots + \beta_K(X_{tK}/\sigma_t) + \varepsilon_t/\sigma_t$$

whose errors have constant variances.

The most difficult task is generally involved with the determination of the form of σ^2_t. Note the variance of the transformed error term is constant.

Weighted least squares in STATA and Shazam perform this estimation.

STATA:

vwls dep_var indep_vars,std (where std is the name of the standard error estimated before running vwls which stands for variance weighted least squares)

c. Autocorrelation

(1) Definition.

Autocorrelation exists if the random disturbances corresponding to different observations are correlated, i.e., the variance covariance matrix (∑) is not diagonal. This situation frequently arises when working with time series. Autocorrelation can also arise from the use of an incorrect functional form or from the deletion of a relevant variable. It will be assumed that an "appropriate" functional form has been selected and no relevant variables have been deleted.

The model can be written as

$$Y = X\beta + \varepsilon$$

where

$$\varepsilon \sim N[O,\Sigma]$$

and

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2N} \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma^2 \end{bmatrix}.$$

The form of the σ_{ij} will depend upon the nature of the correlations between ε_i and ε_j, $\sigma_{ij} = \mathrm{corr}(\varepsilon_i,\varepsilon_j)\sigma^2$.

<u>One of the most commonly adopted</u> models for autocorrelation is

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

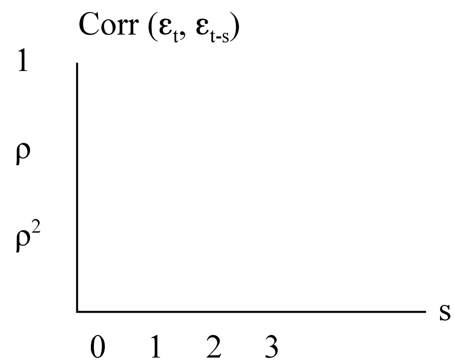
where $-1 < \rho < 1$, $u = (u_1, \ldots , u_n)' \sim N(0,\sigma_u^2 I)$

This model is referred to as a <u>first order autoregressive process</u>, AR(1).

The corresponding Σ matrix can be shown to be

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \vdots & \vdots & & & \vdots \\ \rho^{N-1} & \rho^{N-2} & & \cdots & 1 \end{bmatrix}$$

where $\sigma^2 = \sigma_u^2/(1-\rho^2)$. The correlation between ε_i and ε_j is given by $\rho^{|i-j|}$ which approaches zero as the observations get further apart. This might be graphically depicted with a <u>correlogram</u>.



The STATA command to plot the correlogram is

corrgram e

where "e" denotes the estimated errors from the regression model.

The reasons for the dominance of the AR(1) model in applications may be due to the availability of computer software and the widespread usage of annual data. It is no longer uncommon to have quarterly, monthly, or daily observations on some series. Consequently, other models for the behavior of autocorrelated error terms have been considered. These include:

<u>Autoregressive models of order p</u>, AR(p)

$\varepsilon_t = \varphi_1\varepsilon_{t-1} + \ldots + \varphi_p\varepsilon_{t-p} + u_t$;

<u>Moving average model of order q</u>, MA(q)

$\varepsilon_t = u_t - \theta_1 u_{t-1} - \ldots - \theta_q u_{t-q}$; and

<u>Autoregressive moving average process or order</u> (p,q)

ARMA (p,q)

$\varepsilon_t = \varphi_1\varepsilon_{t-1} + \ldots + \varphi_p\varepsilon_{t-p} + u_t - \theta_1 u_{t-1} - \ldots - \theta_q u_{t-q}.$

These models will be considered in more detail in the section on time series analysis.

(2) Tests for Autocorrelation

Many tests for autocorrelation have been developed. These include a "signs test" which is based upon the number of times that the sign of the random disturbances (estimated) changes.

STATA reg y x's; predict e, resid; runtest e

Wooldridge proposes regressing the OLS residuals on the lagged estimated residuals and then using a "t-test" to determine the statistical significant of the estimated coefficient. This has problems if the estimated value of ρ is near one.

Probably the most common test for autocorrelated error terms of the AR(1) form is the Durbin Watson test statistic.

The Durbin Watson test statistic is defined by

$$\text{D.W.} = \sum_{t=2}^{N} (e_t - e_{t-1})^2 \Big/ \sum_{t=1}^{N} e_t^2$$

and can be rewritten as

$$\text{D.W.} = \frac{2\sum_{t=1}^{N} e_t^2 - 2\sum_{t=2}^{N} e_t e_{t-1}}{\sum_{t=1}^{N} e_t^2} - \frac{e_1^2 + e_N^2}{\sum_{t=1}^{N} e_t^2}$$

$$= 2(1 - \hat{\rho}) - \frac{e_1^2 + e_N^2}{\sum_{t=1}^{N} e_t^2}$$

$$\doteq 2(1 - \hat{\rho})$$

where $\hat{\rho} = \sum e_t e_{t-1} / \sum e_t^2$ is an estimator of the correlation between ε_t and ε_{t-1}. "Critical values" for D.W. are available and depend upon the "α level," sample size N and the number of slope coefficients (K-1 = k').

The model is assumed to include an intercept and not include any lagged dependent variables. For a given N, K-1 and α-level we obtain (d_L,d_U).

The following figure is useful in testing for autocorrelation.

Fail to reject (H_o: $\rho = 0$)

Inconclusive Region   Inconclusive Region

Rejection Region   Rejection Region

0	d_L	d_U	2	$4-d_U$	$4-d_L$	4	D.W.
1			0			-1	ρ

This test statistic is not appropriate if the model includes a lagged dependent variable. In that case Durbin has proposed the h-test

$$h = \hat{\rho} \sqrt{\frac{N}{1 - N\, s^2_{\hat{\beta}_1}}}$$

where $s^2_{\beta_1}$ denotes the least squares estimate of the variance of the coefficient of Y_{t-1} on the right hand side of the equation. The asymptotic distribution of h is N(O,1).

The Breusch-Godfrey test can be used to test for higher autocorrelation.

The command, estat dwatson, following a STATA regression command will calculate the value of the Durbin Watson test statistic. The statistical significance of an exact D.W. Test (independent of the X matrix) is available in some other programs, but is not a Stata option.

(3) Estimation

Least squares estimators do not take account of the correlations between the random disturbances in the estimation process and will not be minimum variance; however, least squares estimators will still be unbiased and consistent.

If $\sum$ is known the MLE, BLUE, GLS estimator of β is given by

$$\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \ .$$

For random disturbances which are AR(1)

$$\Sigma = \frac{\sigma_u^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \cdots & \rho^{N-1} \\ \rho & 1 & \cdots & \rho^{N-2} \\ \vdots & \vdots & & \vdots \\ \rho^{N-1} & & \cdots & 1 \end{bmatrix}$$

and it can be shown that $\sum^{-1}$ is equal to

$$\Sigma^{-1} = \left(\frac{1}{\sigma_u^2}\right)\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix} .$$

Let

$$T = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}.$$

Since

$$T'T = \sigma_u^2 \sum{}^{-1},$$

an application of least squares to the transformed model

$$Y^* = X^*\beta + \varepsilon^*$$

$$TY = TX\beta + T\varepsilon$$

$$\begin{bmatrix} \sqrt{1-\rho^2} & & Y_1 \\ Y_2 & - & \rho Y_1 \\ & \vdots & \\ Y_n & - & \rho Y_{N-1} \end{bmatrix} = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}X_{12} & \cdots & \sqrt{1-\rho^2}X_{1K} \\ 1-\rho & X_{22}-\rho X_{12} & \cdots & X_{2K}-\rho X_{1K} \\ \vdots & \vdots & & \vdots \\ 1-\rho & X_{N2}-\rho X_{N-12} & \cdots & X_{NK}-\rho X_{N-1K} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \sqrt{1-\rho^2} & \varepsilon_1 \\ \varepsilon_2 - \rho\varepsilon_1 & \\ \vdots & \\ \varepsilon_N - \rho\varepsilon_{N-1} & \end{bmatrix}$$

yields $\hat{\beta}_T = (X^{*\prime}X)^{-1}X^{*\prime}Y^* = (X'T'TX)^{-1}X'T'TY$

$\hat{\beta}_T$ is the maximum likelihood or generalized least squares estimator,

$\hat{\beta}_T = (X'\Sigma^{-1}X'\Sigma^{-1}Y$ if T'T is constructed to be $\sum^{-1}$.

The Prais-Winsten approach is based upon using the N observations in Y*, X*. The Cochrane-Orcutt approach is based upon all but the first observation on Y* and X* and yields an approximation to the MLE. When $\sum$ (or ρ) is unknown (almost always) various iterative techniques are available to estimate ρ and β. For more general models for the random disturbances alternative estimation techniques are available. **STATA** will perform the Prais-Winsten or Cochran-Orcutt estimation, respectively, using the commands

tsset time_var

prais dep_var indep_vars

prais dep_var indep_vars, corc

(4) Forecasting (AR(1))

The interdependence between the random disturbances can be used in forming predictions to reduce the variance of conditional forecasts. Recall

$Y_t + X_t\beta + \varepsilon_t$

where $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$.

From (3.2) the minimum variance unbiased predictor is given by

$$Y_N(h) = X_{N+h} = X_{N+h}\tilde{\beta} + W'\Sigma^{-1}e$$

where

$$W = E(\varepsilon_{N+h}\varepsilon) = E\begin{bmatrix} \varepsilon_{N+h}\varepsilon_1 \\ \vdots \\ \varepsilon_{N+h}\varepsilon_N \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} \rho^{N+h-1} \\ \vdots \\ \rho^h \end{bmatrix}$$

for an AR(1) model. Consequently, this simplifies to

$$\boxed{\tilde{Y}_{N+h} = X_{N+h}\tilde{\beta} + \rho^h e_N} \qquad .$$

The second term takes account of the dependency of the error terms and approaches zero as h increases.

5. Auto Regressive Conditional Heteroskedasticity (ARCH) Models

a. Introduction and an example

Some times series are characterized by "clumps," "clusters," or groups of large residuals and groups of small residuals. This pattern of residuals has led to the development of autoregressive conditional heteroskedasticity (ARCH) models.

Graph:

ARCH models are extremely popular. A Google search for "ARCH models"on 2/3/2011 reported 7,240,000 results.

ARCH models attempt to model the behavior of the mean and variance. Many of the basic properties of ARCH models can be introduced by considering a simple ARCH model. Assume that

$$Y_t = X_t \beta + \varepsilon_t \tag{5.1}$$

where $\varepsilon_t = u_t [\alpha_0 + \alpha_1 \varepsilon_{t-1}^2]^{.5}$, u_t is iid N[0,1]. (5.2)

It follows that $E[\varepsilon_t \mid \varepsilon_{t-1}] = 0$.

$$\begin{aligned} E[\varepsilon_t] &= \int E[\varepsilon_t|\varepsilon_{t-1}]\, f(\varepsilon_{t-1})\, d\varepsilon_{t-1} \\ &= 0 \end{aligned} \tag{5.3}$$

from the relationship of the joint and conditional pdf's.

Similarly,

$$\sigma_t^2 = \quad Var[\varepsilon_t|\varepsilon_{t-1}] = E[\varepsilon_t^2|\varepsilon_{t-1}] \tag{5.4}$$

$= E(u_t^2)\,[\alpha_0 + \alpha_1\,\varepsilon_{t-1}^2]$

$= [\alpha_0 + \alpha_1\,\varepsilon_{t-1}^2]$

$= Var[Y_t|Y_{t-1}];$

hence, conditional on ε_{t-1}, ε_t is heteroskedastic. However, the unconditional variance of ε_t is

$$Var[\varepsilon_t] = E[Var[\varepsilon_t|\varepsilon_{t-1}] \tag{5.5}$$

$$= \alpha_0 + \alpha_1\,E[\varepsilon_{t-1}^2\,]$$

$$= \alpha_0/[1-\alpha_1]$$

if the underlying process is variance stationary.

(5.4) can be written as

ARCH(1): $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$ and more generally

ARCH(q): $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_q\varepsilon_{t-q}^2$

GARCH(q, p) : $\sigma_t^2 - \delta_1\sigma_{t-1}^2 - \ldots - \delta_p\sigma_{t-p}^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_q\varepsilon_{t-q}^2$

Engle (1982) and Bollersliv (1986).

b. Estimation

Since assumptions A.1 - A.5 still hold, OLS will still be the minimum variance linear unbiased estimator. However, there will be a more efficient nonlinear estimation- the maximum likelihood estimator. The log-likelihood function for this model, given by

$$\ell = -\frac{1}{2}\sum_{t=1}^{n} \ell n\left[(2\pi)(\sigma_t^2)\right] - \frac{1}{2}\sum_{t=0}^{n}\frac{\varepsilon_t^2}{\sigma_t^2} \tag{5.6}$$

$$= -\frac{1}{2}\sum_{t=1}^{n}\ell n\left[\left(2\pi\right)\left(\alpha_0+\alpha_1\varepsilon_{t-1}^2\right)\right]-\frac{1}{2}\sum_{t=0}^{n}\frac{\left(Y-X_t\beta\right)^2}{\left(\alpha_0+\alpha_1\varepsilon_{t-1}^2\right)}$$

is maximized over the regression parameters $\left(\beta\right)$ and the ARCH parameters

$\left(\alpha_i\text{'}s\right)$

STATA will perform this estimation with the commands

tsset time_variable

arch depvar indep_vars, arch(order of ARCH, number of lagged errors squared) garch(order of GARCH process, number of lagged conditional variances), e.g., *arch y x's, arch(1) garch(1)*

Additonal model flexibility is obtained by selecting alternative distributions for the standardized error, u_t . The normal is the default, the student t and generalized exponential distributions can be used. The corresponding commands are

arch y x's, arch(p) garch(q) dist(t)

arch y x's, arch(p) garch(q) dist(ged)

c. A test for ARCH disturbances can be performed by regressing the square of the OLS residuals on lagged values of the same variables and using and using a LM test $\left(nR^2\right)$ to test the null hypothesis of no ARCH effects

$\left(H_0:\ \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0\right)$ $\sigma_t^2 = \alpha_0 u_t$ which implies that

asymptotic chi square distribution (χ^2(df=p)).

STATA can perform this test using the commands we have discussed earlier, "archlm", "archlm, lags(p)", or "archlm, lags(1/p)" following the reg y x's command**.**

d. Generalizations of the (G)ARCH models

Numerous generalizations of (G)ARCH models have been proposed in the literature which allow for asymmetries, nonlinearities, and other variations of the basic model. Some of these specifications have catchy acronyms such as AARCH, SAARCH, TARCH, NARCH, PARCH, ABARCH, EGARCH, ATARCHE, among others.

6. Stochastic Regressors

This classical normal linear regression model is given by

$$Y = X\beta + \varepsilon \tag{6.1}$$

where

(A.1)' $\varepsilon \sim N(0, \sigma^2 I)$

(A.2)' X_t (rows of X) are nonstochastic and

$$\underset{N\to\infty}{\text{Limit}} \left(\frac{X'X}{N} \right) = \Sigma_{xx} \qquad \text{is nonsingular}$$

We have already discussed variations in the first assumption (A.1)'.

This assumption provides a convenient foundation to develop the classical model. However, situations in which we can assume the X's to be fixed in repeated samples are rare in economic modeling. This assumption is particularly problematic if any of the X's are correlated with the error, as might be the case with endogenous regressors. We will look at a simple macro model containing an endogenous regressor and then formally outline the consequences of relaxing/violating A.5.

a. A simple macro model .

$$\boxed{\begin{aligned} C_t &= \alpha + \beta Y_t + \varepsilon_t \\ Y_t &= C_t + Z_t \end{aligned}} \qquad \text{simple macro model}$$

Dependent variables: C, Y

Independent variable(s): Z=I+G+X

Note: Y on the right handside of the consumption function is an endogenous variable and is referred to as an endogenous regressor

The corresponding reduced form equations, expressing each endogenous variable in terms of the exogenous or independent variables, are given by:

$$C_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta}Z_t + \frac{\varepsilon_t}{1-\beta}$$

$$Y_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta}Z_t + \frac{\varepsilon_t}{1-\beta}$$

Note: The endogenous regressor (Y_t) in the consumption function and ε_t are not independent since

$$\operatorname{cov}(Y_t, \varepsilon_t) = \frac{\sigma^2}{1-\beta}$$

Therefore, $$\operatorname{plim} \hat{\beta}_{OLS} = \beta + \frac{(1-\beta)\sigma^2}{\sigma_Z^2 + \sigma^2}.$$

This is also an example of the simultaneous equation problem where least squares estimators are biased and inconsistent.

One of the main lessons to be learned from this example is that if any of regressors (variables on the right hand side of the equation) have nonzero correlation with the random disturbances, <u>then</u> least squares estimators can be <u>biased</u> and <u>inconsistent</u>.

b. Formal analysis of the consequences of having stochastic X's.

We will consider two cases, (1) where the X's are stochastic, but uncorrelated with the disturbances and (2) where the X's are correlated with the disturbances.

(1) <u>Case 1 of relaxing (A.2)'</u>

(A.2)* The X_t's are stochastic.

X_t and ε_t are stochastically independent.

$\underset{N\to\infty}{\operatorname{plim}} (X'X/N) = \Sigma_{XX}$ is nonsingular

The least squares estimator can be written as

$$\hat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \varepsilon)$$

$$= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon$$

$$= \beta + (X'X)^{-1}X'\varepsilon$$

Taking the expected value of $\hat{\beta}$ yields

$$E(\hat{\beta}) = \beta + E(X'X)^{-1}X'E(\varepsilon)$$

$= \beta$, hence $\hat{\beta}$ is unbiased.

The <u>variance</u> of $\hat{\beta}$ is given by

$$\mathrm{Var}(\hat{\beta}|X) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$

$$= (X'X)^{-1}X'\varepsilon\varepsilon'\, X(X'X)^{-1}$$

$$= ((X'X)^{-1}X'\, E(\varepsilon\varepsilon')X(X'X)^{-1})$$

$$= (X'X)^{-1}X'\, \sigma^2 I\, X(X'X)^{-1}$$

$$= \sigma^2(X'X)^{-1}$$

The <u>consistency</u> of $\hat{\beta}$ can be proven as follows:

$$\operatorname*{plim}_{N\to\infty} \hat{\beta} = \beta + \operatorname*{plim}_{N\to\infty} (X'X)^{-1}X'\varepsilon$$

$$= \beta + \operatorname*{plim}_{N\to\infty}\left(\frac{X'X}{N}\right)^{-1}\frac{X'\varepsilon}{N}$$

$$= \beta + \operatorname{plim}\left(\frac{X'X}{N}\right)^{-1}\operatorname{plim}\frac{X'\varepsilon}{N}$$ Slutsky's Theor

$$= \beta + \Sigma_{XX}^{-1}\, 0$$

$$= \beta$$

Consequently, if the regressors are stochastic but distributed independently of the random disturbances, the least squares estimators are still unbiased and consistent. However, the t and F statistics need not be valid for small samples since $\hat{\beta}$ will no longer be distributed normally due to the X's being stochastic.

(2) Case 2 of relaxing (A.2)

(A.2)** (a) X_t is stochastic

(b) X_t and ε_t are stochastically dependent and $\text{cov}(X_t, \varepsilon_t) \neq 0$.

(c) $\underset{N \to \infty}{\text{plim}}\, [(X'X)/N] = \Sigma_{XX}$ is nonsingular.

Thus,

$$E(\hat{\beta}) = \beta + E\{(X'X)^{-1}X'\varepsilon\}$$

$$\neq \beta$$

$$\underset{N \to \infty}{\text{plim}}\ (\hat{\beta}) = \beta + \underset{N \to \infty}{\text{plim}}\ (X'X)^{-1}X'\varepsilon$$

$$= \beta + \Sigma_{xx}^{-1}\,\text{Cov}(X\varepsilon)$$

$$\neq \beta$$

- **Thus, if the regressors and errors have nonzero correlation then the least squares estimators will be biased and inconsistent.**

- **We now discuss one of the most common methods of obtaining consistent estimators in the presence of stochastic X's.**

7. Instrumental Variables

a. Some background

Consider the generalized regression model

$$Y = X\beta + \varepsilon \tag{7.1}$$

where

(A.1) $\varepsilon \sim N(0,\Sigma)$

Suppose that there exists a set of variables (N observations on K instrumental variables) Z_t, $Z = (Z'_1,\ldots,Z'_N)'$ such that

$$\operatorname*{plim}_{N\to\infty}(Z'X/_N) = \Sigma_{ZX}$$

is nonsingular and

$$(Z'\varepsilon/_N) \underset{\to}{p} N(0,\psi)$$

where $\underset{\to}{p}$ means converges in probability to a normally distributed vector with mean zero (null vector) and variance ψ.

This implies that

$$\operatorname{plim}(Z'\varepsilon/_N) = 0.$$

b. The instrumental variables estimator (same number of Z's as X's, Z and X are NxK))

Now consider the estimator defined by the modified normal equations:

$$Z'Y = Z'X\beta_{IV} \qquad \text{or } Z'e=0 \qquad (7$$

$$\beta_{IV} = \left(Z'X\right)^{-1}Z'Y = \beta + \left(Z'X/n\right)^{-1}\left(Z'\varepsilon/n\right)$$

is referred to as the instrumental variables estimator of β based upon the instruments Z.

The motivation for the modified normal equations can be seen by multiplying equation (7.1) by $Z'/_N$, i.e.,

$$\frac{Z'Y}{N} = \frac{Z'X\beta}{N} + \frac{Z'\varepsilon}{N}$$

with the last term converging to zero as N increases, leaving (7.4).

The asymptotic distribution of $\tilde{\beta}_{IV}$ is

$$\boxed{\tilde{\beta}_Z \overset{a}{\sim} N(\beta;\ \Sigma_{ZX}^{-1}\ \psi\ (\Sigma'_{ZX})^{-1}/N)}$$

(7.7)

If $Var(\varepsilon) = \sigma^2 I$, then the varian β_{IV} ɔvariance of can be written as

$$\sigma^2 (Z'X)^{-1}(Z'Z)(X'Z)^{-1}$$

(7.8)

The bias of the IV estimator is given by

$$E((Z'X/n)^{-1}(Z'\varepsilon/n)) .$$

The consistency of $\tilde{\beta}z$ follows from

$$\text{plim } \tilde{\beta}_Z = \text{plim } (\beta) + \text{plim}\left(\frac{Z'X}{N}\right)^{-1} \text{plim}\left(\frac{Z'\varepsilon}{N}\right)$$

$$= \beta + \Sigma_{ZX}^{-1} 0$$

$$= \beta.$$

c. Some special cases

Instrumental variables can be used to circumvent a number of problems in econometrics as well as providing a unified approach to many estimation problems.

(1) Least squares

If Z=X, then the corresponding estimator is the least squares estimator

$$\tilde{\beta}_X = (X'X)^{-1}X'Y = \hat{\beta}$$

(2) Generalized least squares

If Z is selected to be equal to $\sum^{-1}X$, then the associated estimator is given by

$$\tilde{\beta}_Z = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$$

which is the generalized least squ

estimator.

(3) "Projections of X on Z"

(where we have possible more instruments (m) than X's, m $\geq$ K)

If the selected instruments are the projections of X on the variables Z,

$\left(Z(Z'Z)^{-1}Z'\right)X$, then the IV estimator will be given by

$$\tilde{\beta}_{IV} = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'y$$

which is asymptotically normal with variance-covariance matrix

$$\sigma^2\left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}$$

- Not surprisingly, if Z=X this estimator simplifies to the regular OLS estimator. This estimator can also be looked at as having resulted from a generalized least squares estimator of the parameters in equation (7.6).
- Also, if Z is NxK, then $\hat{\beta}_{IV}$ simplifi $\left(Z'X\right)^{-1}Z'Y$ which agre with the result from the previous section, but differs from the projection result when m > K.

- IV estimators of the parameters in the model

$$y_1 = \beta_1 y_2 + \beta_2 y_3 + \gamma_1 x_1 + \gamma_2 x_2 + \varepsilon$$

can be obtained in STATA (11) using the command:

ivregress 2sls **y1 (y2 y3 = z1 z2 z3) x1 x2**

where y2 and y3 are endogenous regressors and z1, z2, and z3 are instrumental variables. There must be at least as many instrumental variables as endogenous regressors

Generalized method of moments (gmm) and limited information maximum likelihood (liml) are alternative estimators and may be used instead of 2sls.

The following postestimation commands may be useful in your analysis

estat firststage is used to explore the correlation of the instruments

with the endogenous regressors

estat overid **is used to test whether extra instruments are correlated with the error term.**

This particular estimator is an instrumental variables estimator and is also referred to as the two-stage least squares estimator which will be discussed in greater detail in another section.

d. Selection of instrumental variables. A valid instrumental variable (Z) is not included in the equation being estimated and should satisfy two conditions: relevance and exogenous

(1) Relevance: $corr(X, Z) \neq 0$

(2) Exogenous: $corr(\varepsilon, Z) = 0$

A **necessary condition** to perform IV estimation is that there must be at least as many instrumental variables (m) as there are endogenous regressors (k), $m \geq k$. The regression coefficients are said to be exactly identified if the m=k. IV estimators can't be obtained if m<k (underidentified). The model is said to be overidentified if m>k. The validity of the overidentifying assumptions is tested using the "estat overid" command.

It is common practice to use an F-test to test for the "relevance" of the instruments. The endogenous regressor(s) is (are) regressed on the independent and instrumental variables. An F-test is then used to test for statistical significance of the instruments. The instruments are said to be "weak" if the F statistic $\left(\tilde{F}\right)$ is statistically small. Weak instruments can result in instrumental variables being worse than OLS. For one endogenous regressor, a rule of thumb is that the maximal bias of the IV estimator (2SLS) will not be greater than 10% of the OLS bias if $10 \leq \tilde{F}$.

A Hausman test (discussed in the next section) compares the OLS and instrumental (or other consistent estimator) variables estimator to test the endogeniety of the endogenous regressor.

e. Instrumental variables estimation in quantile regression models. Chernozhukov and Hansen (Econometrica, 2005, 245-261) outline an approach for Instrumental Variables to be applied to quantile regression models which include endogenous regressors. Recall that quantile regression models attempt to model the impact that different variables will have on the distribution of a variable of interest such as what is the impact of the number of years of schooling on the distribution of income.

f. Some other issues

- Structural vs. experimental camps in econometrics. M. Keane (2010) wrote an interesting article in the Journal of Econometrics ("Structural vs. atheoretic approaches to econometrics," 156 (1), pp. 3-20) making a case for the importance of making the underlying assumptions explicit. Insightful comments on this survey paper are provided by J. Rust, R. Blundell, and J. Heckman and Urzua.
- Numerous examples of instrumental variables can be found in recent literature. One example in modeling housing values with violent crime as an explanatory variable, some authors have used the number of murders as an instrumental variable. Some other references are given.
- Angrist, J.D., K. Graddy, and G.D. Imbens (2000). "The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish." *Review of Economic Studies* 67, 499-527.
- Levitt, S. (1997). "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review*, June 1997.
- If the instruments are *weak* (not highly correlated with the rhs endogenous variables), problems can arise. In the case of endogenous regressors, OLS is inconsistent. IV estimation (or a variant) is often suggested in this case. However if the instrument is weakly correlated with the rhs endogenous variable, then even a small correlation between the instrument and the error can produce a larger inconsistency in the IV estimate of beta than in the OLS

estimator. An instrumental variable is often said to be *invalid* if it is correlated with the error term, regardless of the its correlation with the endogenous regressor. Bound, Jaeger, and Baker (1996, JASA, 443-450). This can be seen from an inspection of the bias of the IV estimator given after equation (7.8).

- Bekker (1994, Econometrica, 657-681) provides some approximations to the distributions of IV estimators which are useful in exploring the properties of alternative IV estimators.
- Donald, S.G and W. K. Newey, "Choosing the Number of Instruments," Econometrica, 2001, 1161-1191. They use asymptotic expansions of the MSE of linear combinations of the coefficient estimators to explore the question of the optimal number of valid instruments which will minimize the MSE. They apply their methodology to the widely explored Angrist and Krueger data (1991, QJE, 979-1014) to estimate returns to schooling in the model

$$\ln(w) = \beta_0 + \beta_1(yearsofschool) + \gamma's(9 - yearsofbirth, 50 - stateofbirth) + \varepsilon$$

With instruments equaling subsets of

$$Z = (quarterofbirth, QOB * SOB, QOB * YOB)$$

$\beta_1 \doteq .10$, for 2SLS, LIML with various combinations of instruments whereas

the OLS estimates are between .6 and .7. The use of quarter of birth and interactions with year of birth is an interesting twist of this paper. Angrist

(1990, AER, 313-335) used lottery numbers as an instrument for military service.

- If the instruments are weak (not highly correlated with the rhs endogenous variables, then GMM and IV statistics are non-normal and hypothesis tests are unreliable. Stock, Wright, and Yogo (2002, JBES, 518-529). This paper also provides an excellent survey of related issues and alternatives.
- Kleibergen (September 2002, Econometrica, 1781-1803). Pivotal statistics for testing structural parameters in instrumental variables regression. Bad instruments not only lead to imprecise estimates of the structural parameters but also imply that the standard statistics we use to assess these estimates are unreliable.
- Angrist, Imbens, and Krueger (Journal of Applied Econometrics, 1999, 57-67) propose using Jackknife instrumental variables which will be, by construction, independent of the error terms in finite samples.

(g) Stochastic Regressors and IV estimation

The consumption function from the section on stochastic regressors can be written as

$$C = X\beta + \varepsilon$$

or $$\begin{pmatrix} C_1 \\ \vdots \\ C_N \end{pmatrix} = \begin{pmatrix} 1 & Y_1 \\ \vdots & \vdots \\ 1 & Y_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

Recall that the Y_t are correlated with the random disturbances; hence, the least squares estimators will be biased and inconsistent.

It can be shown that selecting

$$Z = \begin{pmatrix} 1 & \hat{Y} \\ \vdots & \vdots \\ 1 & \hat{Y}_N \end{pmatrix}$$

where $\hat{Y}_t = \hat{\Pi}_1 + \hat{\Pi}_2 Z_t$ (from macro example), the least squares predictions will yield consistent and asymptotically normal estimators of β_1 and β_2. These estimators can be thought of as IV estimators and are frequently referred to as two stage least squares (2SLS) estimators.

8. Specification Tests in Econometrics

Jerry Hausman, Econometrica, 46(1978), pp. 1251-1270.

a. Background

Consider the model

$$Y = X\beta + \varepsilon$$

where

(A.1)' $\mathrm{var}(\varepsilon \mid x) = \sigma^2 I$ (8.1)

(A.2)' $E(\varepsilon|x) = 0 \text{ or } \underset{N}{\mathrm{plim}} \left(\frac{X'\varepsilon}{N} \right) = 0$ (8.1)

Violations of (A.1)' result in least squares estimators being unbiased, consistent, but not minimum variance. Violations of (A.2)' result in least squares being biased and inconsistent.

(A.2)' is very important and a few tests have been developed to investigate its validity: Wu, Econometrica (1973); Ramsey, Frontiers in Econometrics; Hausman, Econometrica (1978).

b. Outline of the Hausman test

Let $\hat{\beta}_0$ denote a consistent, asymptotically normal, asymptotically efficient estimator of β if (A.1)' and (A.2)' are satisfied, but biased and inconsistent if (A.2)' is violated.

Let $\hat{\beta}_1$ denote an alternative estimator which is consistent under both the null and alternative hypothesis. Under these conditions the Hausmann test is based upon the difference,

$$\boxed{\hat{q} = \hat{\beta}_1 - \hat{\beta}_0 .}$$

If (A.1)' and (A.2)' are satisfied, then plim $\hat{q} = 0$. If (A.2)' is not satisfied, then plim $\hat{q} \neq 0$ and its ($\hat{q}$) asymptotic distribution provides the basis for rejecting (testing) (A.2)'. This is formalized by the theorem

(1) Let $\hat{\beta}_0$ $\hat{\beta}_1$ denote consistent asymptotically normally estimato $\hat{\beta}_0$ ith attaining the Cramer Rao bound

$$\sqrt{n}\left(\hat{\beta}_0 - \beta\right) \sim^a N\left(0, V_0\right)$$
$$\sqrt{n}\left(\hat{\beta}_1 - \beta\right) \sim^a N\left(0, V_1\right)$$

The variance of the asymptotic distribution of $\hat{\beta}_i$ is V_i / n , which goes to zero as n increases because of being a consistent estimator if the null hypothesis is valid.

The asymptotic distribution of $\hat{q}$ ($\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$) is given by

$$\sqrt{N}(q - 0) \overset{a}{\sim} N(0, V(\hat{q}))$$

where $V(\hat{q}) = V_1 - V_0$.

The statistic

$$\boxed{N\hat{q}'\, V(\hat{q})^{-1}\hat{q} \overset{a}{\sim} \chi^2 (\#\ \text{par})}$$

provides the basis for testing for specification error. **Perhaps an easier form for the Hausman Test is as follows:**

$$\left(\hat{\beta}_1 - \hat{\beta}_0\right)' \left(Var\left(\hat{\beta}_1\right) - Var\left(\hat{\beta}_0\right)\right)^{-1} \left(\hat{\beta}_1 - \hat{\beta}_0\right) \overset{a}{\sim} \chi^2 \text{ (\# slope pars)}$$

In this form you don't need to worry about the sample size being in the formula.

Proof.

It can be shown that

- $[\hat{q} - 0]' \, Var^{-1}(\hat{q}) \, [\hat{q} - 0] \sim \chi$ (# parameters)
- $\hat{q}$ and $\hat{\beta}_0$ are uncorrelated.
- $\hat{\beta}_1 \quad \hat{\beta}_0 \qquad + \hat{q}$
- $Var(\hat{\beta}_1) = Var(\hat{\beta}_0) + Var(\hat{q})$
- $Var(\hat{q}) = Var(\hat{\beta}_1) - Var(\hat{\beta}_0)$

$$= N\,[\,V_1 - V_0\,]$$

c. Applications: simultaneous equations, measurement error, panel data

(1) - Simultaneous equations

Simultaneous equations: Let $\beta_0 = \beta_{OLS}$ and $\beta_1 = \beta_{2SLS \text{ or } 3SLS}$.

β_{OLS} - efficient if A.1 - A.5 are valid

Inconsistent - if A.5 is violated

β_{2SLS} - consistent even if A.5 is violated

(2) Measurement error

Let $\beta_0 = \beta_{OLS}$ and $\beta_1 = \beta_{IV}$. Under the assumption of no measurement error, both estimators are consistent estimators of β with the OLS estimator being efficient. However in the presence of measurement error (in the X's) the OLS estimator is inconsistent and an appropriate IV estimator is consistent.

9. Seemingly Unrelated Regression Models (SURE-Zellner)

Consider the problem of estimating the coefficients in G separate regression equations.

$$\begin{aligned} Y_1 &= X_1\beta_1 + \varepsilon_1 \\ Y_2 &= X_2\beta_2 + \varepsilon_2 \\ &\vdots \\ Y_G &= X_G\beta_G + \varepsilon_G \end{aligned} \tag{9.1}$$

where Y_i denotes an Nx1 vector of observations on the dependent variables in the i^{th} regression equation. X_i denotes the NxK_i matrix of observations on the explanatory variables and β_i denotes the K_ix1 vector of associated coefficients.

Assume that

$$\mathrm{Var}(\varepsilon_i) = \sigma_{ii}^2 I_N \tag{9.2}$$

i.e., the random disturbances in each equation are uncorrelated over time and characterized by homoskedasticity.

If the random disturbances in each equation are independent of the random disturbances in all other equations, then least squares estimators of the β_i in each equation in (9.1) will yield MLE and BLUE of the β_i if

$$\varepsilon_i \sim N[0, \sigma_{ii}^2 I_N] \ .$$

$$\hat{\beta}_i = (X'_i X_i)^{-1} X'_i Y_i$$

$$\sim N[\beta_i; \sigma_{ii}(X'_i X_i)^{-1}] \ .$$

If the covariances between contemporaneous random disturbances in the i^{th} and j^{th} equation are given by σ_{ij}, i.e.,

$$E(\varepsilon_i \varepsilon_j') = E \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{\in} \end{bmatrix} \begin{bmatrix} \varepsilon_{j1} & \dots & \varepsilon_{jN} \end{bmatrix}$$

$$= E \begin{bmatrix} \varepsilon_{i1}\varepsilon_{j1} & \cdots & \varepsilon_{i1}\varepsilon_{jN} \\ \varepsilon_{iN}\varepsilon_{j1} & \cdots & \varepsilon_{iN}\varepsilon_{jN} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \cdots & 0 \\ 0 & & & \sigma_{ij} \end{bmatrix}$$

$$= \quad \sigma_{ij} \, I.$$

THEN, the Zellner Seemingly Unrelated Estimator (SURE) can yield unbiased estimators with smaller variances than the least squares estimators.

The SURE estimators can be defined in terms of an alternative representation of (9.1):

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_G \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & X_G \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_G \end{bmatrix}$$

or

$$Y = X\beta + \varepsilon \tag{9.7}$$

where Y, X, β and ε are implicitly defined in (9.6) and have dimensions NGx1, NGx (K_1 + ... + K_G), (K_1 + ... + K_G) x1 and NGx1.

From (9.3) and (9.5),

$$\varepsilon \sim N\,(0, \Omega = \sum \otimes I_N),$$

i.e., $\mathrm{var}(\varepsilon) = \sum \otimes I_N = \Omega$

$$= \begin{bmatrix} \sigma_{11} I_N & \sigma_{12} I_N & \cdots & \sigma_{1N} I_N \\ \sigma_{21} I_N & \sigma_{22} I_N & \cdots & \sigma_{2N} I_N \\ \vdots & \vdots & & \vdots \\ \sigma_{N1} I_N & \sigma_{N2} I_N & \cdots & \sigma_{NN} I_N \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{var}(\varepsilon_1) & \mathrm{cov}(\varepsilon_1 \varepsilon_2) & \cdots & \mathrm{cov}(\varepsilon_1 \varepsilon_N) \\ \mathrm{cov}(\varepsilon_2 \varepsilon_1) & \mathrm{var}(\varepsilon_2) & \cdots & \mathrm{cov}(\varepsilon_2 \varepsilon_N) \\ \vdots & \vdots & & \vdots \\ \mathrm{cov}(\varepsilon_N \varepsilon_1) & \mathrm{cov}(\varepsilon_N \varepsilon_2) & \cdots & \mathrm{var}(\varepsilon_N) \end{bmatrix}.$$

The generalized least squares estimator (Zellner SURE) is given by

$$\boxed{\begin{aligned} \tilde{\beta} &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \\ \tilde{\beta} &\sim N(\beta, (X'\Omega^{-1}X)^{-1}) \end{aligned}} \quad . \tag{9.10}$$

The least squares estimator (9.4) can be rewritten in the notation (9.7) by

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\hat{\beta} \sim N(\beta, (X'X)^{-1}X'\Omega X(X'X)^{-1}) \ .$$

The least squares estimator has larger variances than $\tilde{\beta}$ unless $\sigma_{ij} = 0$ for $i \neq j$ or $X_1 = X_2 = ... = X_G$, i.e., the random disturbances in different equations are mutually independent or the independent variables and related observations are identical in each equation.

ESTIMATION:

The STATA commands for Zellner's Seemingly Unrelated Regression Estimation is

sureg (dep_var1 ind_vars1) (dep_var2 ind_vars2) or

sureg (dep_var1 ind_vars1) (dep_var2 ind_vars2), isure

where the option "isure" iterates the estimation process until convergence is achieved.

Comments:

1. If $\sum$ is unknown, then consistent estimators of σ_{ij} can be obtained using

$$s_{ij} = \frac{1}{N - K} e_i' e_j$$

where $e_i = Y_i - X_i\hat{\beta}_i$ denotes the least squares estimates of the residuals.

2. The seemingly unrelated regression estimator will be the same as the least squares estimator if either

(a) $X_i = X_j$ for all i, j <u>or</u>

(b) $\sigma_{ij} = 0$ for $i \neq j$

3. If $\sigma_{ij} \neq 0$ and the X_i matrices are not identical, then the variances of the least squares estimators will have variances which are at least as large as Zellner's SURE of β.

4. A similar approach can be used to combine cross sectional and time series data.

5. This approach can be modified to handle autocorrelation.

10. Models for independent cross sectional data and for panel data

Independent cross sectional data over time is obtained by conducting random samples at different points in time. For example, random samples of demographic data (age, household size, income, employment status, and educational attainment) over time would be referred to as being an independent cross sectional data set. The sample sizes might be the same or different.

Panel data refers observational data on individuals (i, i= 1, 2, ... m) over time(t=1,2,..., T_i) (two dimensions) and might be denoted as $\left(Y_{it}\right)$. The panel data set is referred to as balanced if every individual is observed for every point of time ($T_1 = T_2 = ... = T_m = T$). Otherwise panel data set is referred to as unbalanced. Observations for a given individual over time are time series; whereas, cross sectional data are observations for different individuals at a given point in time. In many applications, the data are for short periods of time, but include many individuals.

a. Models for independent cross sections over time

Models for these data sets can take a number of different forms. Perhaps the simplest representation is given by

$$Y_{it} = X_{it}\beta + \varepsilon_{it} \tag{1}$$

where X_{it} denotes a *1xk* vector of observations on k-exogenous variables for i^{th} individual in the t^{th} time period and where the marginal impact of the X's on Y is assumed constant over individuals and time (including the intercept). This specification is sometimes called the **pooled**

model and may include binary variables to represent the time period. Let the model be rewritten in matrix form as

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_m \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_m \end{bmatrix}$$

or

$$Y = X\beta + \varepsilon$$

OLS estimates of β $\hat{\beta} = \left(X'X\right)^{-1} X'Y$, can be obtained with the command

reg y x's

Recall, that in the presence of heteroskedasticity OLS estimators are inefficient and have invalid t-statistics. The same tests (White, modified White, Breusch-Pagan, etc.) used in the regular regression model can be used to test for the present of heteroskedasticity in the pooled regression model. Robust standard errors can be used to obtain appropriate t-statistics using the command:

reg y x's, vce(robust, <u>boot</u>strap, or <u>jack</u>knife)

b. Models for Panel data

With panel data the same individuals (persons, firms, countries, etc.) are followed over time. Thus in the model

$$Y_{it} = X_{it}\beta + \varepsilon_{it}$$

X_{it} and Y_{it}, for a given value of "I", correspond to the observations on the dependent and independent variables over time for a given person.

Regular least squares can be used to estimate the coefficients which are assumed to be invariant over time and for different individuals. As we showed in a previous chapter, OLS will not be efficient if either autorcorrelation or heteroskedasticity exists. Generalized least squares (GLS), BLUE, or MLE can provide more efficient estimators in this case. We may also want to allow for different intercepts for individuals or time periods and this leads to random or fixed effects specifications.

(1) GLS (generalized least squares estimators) can provide more efficient estimators than OLS. The formulas for the GLS estimators and corresponding variance-covariance matrix are given by

$$\tilde{\beta} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}Y$$

$$Var\left(\tilde{\beta}\right) = \left(X'\Omega^{-1}X\right)^{-1}$$

where $\mathrm{Var}(\varepsilon) = \Omega \quad \Omega = \Sigma_{mxm} \otimes I_{T_i x T_i} \quad T_i \geq m$, .

In order to obtain GLS (generalized least squares) estimators, simplifying assumptions about the variance of $\varepsilon, \Omega,$ need to be made and the nature of the longitudinal/panel data must be provided to STATA with the **"xtset"** command as follows:

xtset panel_var or

xtset panel_var time_var.

where *panel_var* denotes the individual identification code or group variable name and *time_var* is an index which represents the time variable which defines the panels being used.

This is similar to using the *tsset time_variable* command to alert Stata that time series are being used. To clear the xt settings, use the command **xtset, clear**

An alternative approach is to list *i(panel_var) t(time_var)* after the first *xt* commands that will be discussed next.

Different intercepts for the different panels and or time periods can be obtained using the command **xtreg y x's i.panel i.time**.

Various generalized least squares estimators of β , depending on the form of the variance-covariance of the error term, can be obtained with the "***xtgls***" command.

If there is heteroskedasticity across panels,

$$\Omega = \begin{bmatrix} \sigma_1^2 I & 0 & . & . & . & 0 \\ 0 & \sigma_2^2 I & . & . & . & 0 \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ 0 & 0 & . & . & . & \sigma_m^2 I \end{bmatrix},$$

corresponding GLS estimators can be obtained using the command

xtgls y x's, panels(hetero)

If there is correlation across panels (cross-sectional correlation) of the form

$$\Omega = \begin{bmatrix} \sigma_1^2 I & \sigma_{1,2} I & . \; . \; . & \sigma_{1,m} I \\ \sigma_{2,1} I & \sigma_2^2 I & . \; . \; . & \sigma_{2,m} I \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ \sigma_{m,1} I & \sigma_{m,2} I & . \; . \; . & \sigma_m^2 I \end{bmatrix},$$

the GLS estimator is obtained with the command (this can only be applied to balanced panels)

xtgls y x's, panels(correlated)

The command

xtgls y x's, igls

iterates the generalized least squares procedure until convergence is obtained.

STATA allows for autocorrelation within the panels. The STATA manual, (Lognitudinal/Panel Data, version 10, p. 150) states that three options are allowed:" corr(independent) or no autocorrelation, corr(ar1) (serial correlation where the correlation parameter is common for all panels), or corr(psar1) (serial correlation where the correlation parameter is unique for each panel)." A couple of observations are in order: (1) xtgls y X's, panels(iid) corr(independent) is equivalent to regress y X's; (2) when corr(ar1) or corr(psar1) are specified the iterated GLS estimator does not converge to the MLE.

Some examples and variations include:

xtgls y x's, panel(hetero) **fit panel-data model with hetero across panels**

xtgls y x's, panels(correlated) **correlation and hetero across panels**

xtgls y x's, panels(correlated) igls **uses iterative gls**

xtgls y x's, panels(hetero) corr(ar1)   hetero across panels and auto within panels

xtgls y x's,panels(iid) corr(psar1)

Tests for heteroskedasticity and autocorrelation

An LR test can be used to test for heteroskedasticity across panels.

The unrestricted (with hetero) model is estimated using the command

xtgls y X's, igls panels (hetero)

estimates store hetero

The restricted model (assuming homoskedasticity) is estimated using the command

xtgls Y X's

The degrees of freedom is e(N-g)-1, so LR test is obtained by typing

local df = e(N_g)-1

lrtest hetero ., df(`df')

Testing for autocorrelation.

A user-written program to test for autocorrelation was written by David Drukker.

This program, called xtserial, can be downloaded using the program

findit xtserial

net sj 3-2 st0039

net install st0039

The program is then executed by typing

xtserial y X's

A significant value of the test statistic suggests serial correlation

(2). Fixed and random effects specifications

The fixed and random effects representations are a little different than the form just considered, in particular, they can be represented as:

$$Y_{it} = X_{it}\beta + \alpha_i + \varepsilon_{it}$$

where the marginal impact of changes in the X's are still assumed to be constant across individuals, i.e. the β 's are the same for each individual. The only difference in the relationship across firms is in the intercept term. In fixed effects (fe) models the α_i are unknown constants and in random effects models (re) models the α_i are random. OLS can be used to estimate the unknown parameters in this form where binary variables are added to the set of exogenous variables.

STATA uses a slight variation on this formulation in estimation

$\alpha_i = \alpha + v_i$ v_i where the are estir $\sum_i v_i = 0$ that ;

hence, $$Y_{it} = \alpha + X_{it}\beta + v_i + \varepsilon_{it}$$.

Consider taking the following averages of (3):

$$\bar{y}_i = \alpha + \bar{x}_i\beta + v_i + \bar{\varepsilon}_i \qquad \text{(4) (average over i)}$$
$$\bar{\bar{y}} = \alpha + \bar{\bar{x}}\beta + \bar{v} + \bar{\bar{\varepsilon}} \qquad \text{(5) (average over i \&t)}$$

Combining equations (3) and (4), and (3), (4), and (5), respectively, enables us to write

$$Y_{it} - \overline{y}_i = \left(X_{it} - \overline{x}_i\right)\beta + \left(\varepsilon_{it} - \overline{\varepsilon}_i\right) \qquad (6)$$

$$Y_{it} - \overline{y}_i + \overline{\overline{y}} = \alpha + \left(X_{it} - \overline{x}_i + \overline{\overline{x}}\right)\beta + \nu_i + \left(\varepsilon_{it} - \overline{\varepsilon}_i + \overline{\nu}\right) + \overline{\overline{\varepsilon}} \qquad (7)$$

STATA's **fixed effects** (**within**) estimation procedure, **xtreg y x's, fe**, correspond to estimating β in equation (6) or equation (7), as adding in the overall mean of y has no impact on the estimates of β. Note, from (6) or (7), that coefficients of explanatory variables which do not vary across time can't be estimated. Estimates of ν_i, $\hat{\nu}_i = u_i$ are obtained, but not reported, from (4) as $u_i = \overline{y}_i - \hat{\alpha} - \overline{x}_i\hat{\beta}$ witl $\overline{x}_i\hat{\beta}$ orrelation with being reported.

Three R^2 's are reported:

$R^2_{within} = corr^2(y_{it} - \overline{y}_i, (X_{it} - \overline{x}_i)\hat{\beta})$ R^2 , fron

$R^2_{Between} = corr^2(\overline{y}_i, \overline{x}_i\hat{\beta})$ R^2 $\overline{y}_i$, $\overline{x}_i$ from regressing on

$R^2_{Overall} = corr^2(y_{it}, X_{it}\hat{\beta})$ R^2 y_{it} , X_{it} from regressing o

regression

Least squares estimation with a dummy variable (LSDV) for the different intercepts is equivalent to running a fixed effects regression and yields estimates of the different intercepts in the fixed effects specification. This estimation is facilitated with the command

xi:reg y x's i.firm or xi: reg y x's i.firm i.time

where "firm" and "time" denote the cross sectional and time variables, respectively.

The hypothesis that there is no heterogeneity in the fixed effects or that the grouped effects are all the same $\left(\nu_i = 0, \textit{for}\ \text{all i}\right)$, can be tested using a Chow Test by comp

the pooled and LSDV regressions as follows:

$$F(m-1, mT-m-K) = \left[\frac{\left(R^2_{LSDV} - R^2_{Pooled}\right)/(m-1)}{\left(1-R^2_{LSDV}\right)/\left(mT-m-K\right)} \right]$$

where m = number of groups and T = length of time series. The results of testing this hypothesis are also reported Stata when estimating the fixed effects model.

STATA's **between effects** estimators can be obtained by estimating equation (4) using the Stata command,

xtreg y x's, be

The same R^2 reported with the fixed effects methods are reported for the between effects printouts, with the $R^2_{Between}$ corresponding to the fitted model with this estimation procedure.

In the **random effects** model the ν_i in the regression model

$$Y_{it} = \alpha + X_{it}\beta + \nu_i + \varepsilon_{it}$$

are assumed to be distributed identically and independently with mean zero and constant variance. The term $(\nu_i + \varepsilon_{it})$ can be thought of as a composite error term with

$Var(\alpha_i + \varepsilon_{i.}) = \sigma_\varepsilon^2 I_T + \sigma_u^2 i_T i_T^{'} = \Sigma$ ε $\Omega = I_m \otimes \Sigma$ and var( )=

obtain the desired estimators using the command, **xtreg y x's, re.** A maximum likelihood estimator could be used, using **xtreg y x's, mle**, and will generally give results similar to those obtained from xtreg y x's, re unless $\sum_i T_i$ is small ν_i f **the are uncorrelated with the explanatory variables, random effects estimators will be efficient**; **however, they will be inconsistent if the** ν_i **are correlated with the explanatory variables.** Random effects estimation can estimate coefficients of explanatory variables which do not change over time for individuals. The output associated with random effects estimation includes the same three R^2 's as with fixed effects (within groups) and between groups estimation.

The fixed effects estimator is consistent whether the data are generated by a fixed effects model or a random effects model; however, it is less efficient than the random effects estimator if the data generating process is a random effects model. A Hausman test can be used to test the null hypothesis that the data are generated by a random effects model (there is no correlation between the intercepts and the explanatory variables). The Stata commands for performing a Hausman test are given by *xtreg y x's, fe*

estimates store fe

xtreg y x's, re

estimates store re

hasuman fe re

In summary, the STATA commands for estimating fixed (within), between, and random effects models, respectively, are given by

xtset *panel_var* or xtset *panel_var time_var*

xtreg y x's, fe

xtreg y x's, be

xtreg y x's, re

A few general comments (some taken from Kennedy, 6th edition):

(1) Since most panel data sets have a time dimension, time series problems such as unit roots or cointegration may arise and should be tested for.

(2) Fixed effects estimation is OLS estimation when using the fixed effects model and the random effects model is actually GLS applied when using the random effects model.

(3) Fixed and random effects estimators assume that the slopes are equal across-sectional units.

(4) When there are intercept differences across individuals and time periods, the model is referred to as a two-way effects model to distinguish it from a one-way effects model where the intercepts only differ across individuals (or time). Reminder: xi: reg y x's i.firm i.time

(5) A Chow test can be used to test whether the intercepts are the same by using OLS on the pooled data and comparing the results to those obtained from fixed effects estimation. If the intercepts do not differ, OLS on the pooled data is preferred. An alternative approach is to use a Lagrange multiplier (LM) test to see if the variance of the intercept component of the composite error term is zero.

(6) Random effects is recommended when the composite error term is uncorrelated with the explanatory variables. The Hausman test can be used to explore this issue.

(7) **The estimated standard errors can be sensitive to underlying assumptions with obvious implications relative to tests of hypotheses and statistical significance.** Hence, you may want to consider correcting for clustered standard errors using the commands

xtreg y x's, fe cluster(cluster variable)

reg y x, cluster(cluster variable)

where the cluster variable might be the firm code.

Some additional Stata options:

(1) vce (variance covariance estimator) options:

vce(oim) observed information matrix

vce(opg) outer product of the gradient (OPPG) vectors

vce(robust) Huber/White sandwich estimator

vce(bootstrap [, bootstrap_options]) bootstrap estimation

vce(jackknife {,jackknife_options]) jackknife estimation

(2) The command “**xtregar y x’s, re or fe**”can be used to estimate fixed effects or random effects models when the error term is characterized by a first order autoregressive process.

(3) **xtpcse. panel-corrected standard errors**

Format: xtpcse depvar [indepvars] [if] [in] [weight] [, options]

Description: xtpcse calculates panel-corrected standard error (PCSE) estimates for linear cross-sectional time-series models where the parameters are estimated by OLS or Prais-Winsten regression. When computing the standard errors and the variance-covariance estimates, xtpcse assumes that the disturbances are, by default, heteroskedastic and contemporaneously correlated across panels.

Options	description
noconstant	suppress constant term
correlation(independent)	use independent autocorrelation structure
correlation(ar1)	use AR1 autocorrelation structure
correlation(psar1)	use panel-specific AR1 autocorrelation structure
rhotype(calc)	specify method to compute autocorrelation parameter; see Options for details; seldom used
np1	weight panel-specific autocorrelations by panel sizes
hetonly	assume panel-level heteroskedastic errors
independent	assume independent errors across panels

(4) **xtivreg**

This command is used for fitting panel-data models in which some of the right-hand side variables are endogenous.

(5) xtlogit, xtpoisson, xtprobit, xttobit are available options

(6) Numerous variations are possible, e.g., consider

$$Y_{it} = \alpha + X_{it}\beta + \nu_i + \gamma_t + \varepsilon_{it}$$

which allows for cross-sectional effects and time contrasts.

(7) xtsum [varlist] [if] [, i(varname_i)]xtsum, is a generalization of summarize, reports means and standard deviations for cross-sectional time-series (xt) data; it differs from summarize in that it decomposes the standard deviation into between and within components.

(8) areg y x's, absorb(var_name)

Stata command which runs a linear regression with many binary variables which are absorbed into the designated var_name.

(9) A special edition of the Journal of Econometrics (editited by Baltagi, Kelejian, and Prucha(140, 2007) focuses on an analysis of spatially dependent data discusses related issues of identification, estimation, and testing.

An example (also see exercise in the problem set)

Consider the data set (STATATEST.txt) (from the STATA website)

t	code	x	y	d1	d2	d3	d4
1	1	0	-5	1	0	0	0
2	1	8	23	1	0	0	0
3	1	14	44	1	0	0	0
4	2	10	29	0	1	0	0
5	2	16	26	0	1	0	0
6	3	4	17	0	0	1	0
7	3	11	17	0	0	1	0
8	3	5	31	0	0	1	0
9	4	18	50	0	0	0	1
10	4	5	26	0	0	0	1
11	4	2	17	0	0	0	1

This data set is an unbalanced panel.

Consider the output corresponding to the following commands:

reg y x d1 d2 d3

xtreg y x,fe

Note that (1) the estimated coefficient for x and standard errors are the same

xtreg y x,be

xtreg y x,re

c. Difference in differences

Consider the impact of an experiment in a particular market. In setting up the experiment there will be treatment and control groups and observations before and after the experiment. A simple model to estimate the impact of the experiment might be written as

$$y_{it} = \beta_1 + \beta_2 treat_{it} + \beta_3 after_{it} + \beta_4 (treat_{it} after_{it}) + \varepsilon_{it}$$

where *treat* = 1 if in the experimental group, 0 otherwise; *after*=1 if after the experiment and 0 if before the experiment. The following table summarizes the expected levels for different combinations of experiment/control groups and before/after, along with the corresponding marginal impacts.

	Treatment Group	Control Group	Difference
Before the experiment	$\beta_1 + \beta_2$	β_1	β_2
After the experiment	$\beta_1 + \beta_2 + \beta_3 + \beta_4$	$\beta_1 + \beta_3$	$\beta_2 + \beta_4$
Difference	$\beta_3 + \beta_4$	β_3	β_4

Thus, the coefficient of the interaction term (β_4) captures the before/after effect of the experiment and is equal to the differences of before and after experiment and treatment and control group levels. Obviously, other controls could be added to the regression model. Card and Krueger (1994, AER) use this approach to investigate the impact of minimum wage on employment in the fast food industry in New Jersey and Pennsylvania. The before/after takes account of a national recession. Another example might be the impact of Craig's list on the housing in the BYU/UVU rental market. Let CL=1 if listed on Craig's list, 0 otherwise and BYU=1 if the rental unit qualifies for BYU student rental and 0 otherwise. A possible model might be

$$\mathrm{Re}\,nt_{it} = \beta_1 + \beta_2 CL_{it} + \beta_3 BYU_{it} + \beta_4 (CL_{it} BYU_{it}) + \varepsilon_{it}$$

where β_4 denotes the impact of Craig's list on the expected rental price of BYU housing.

These formulations need to be adjusted if an endogenous variable is included as a regressor and also if the data are serially correlated. Instrumental variables provide one approach to the endogeniety problem. See Bertrand, Duflo, and Mullainathan (2004, QJE) for a discussion of the inconsistency of standard errors in the presence of serial correlation.

d. Statistical Inference

- See Chris Hansen ("GLS inference in panel and multilevel models with serial correlation and fixed effects," Journal of Econometrics, 140(2007), 670-694) for a summary of some of the issues, literature review, and suggested solutions

11. Regression Discontinuity Formulations

Regression Discontinuity (RD) were introduced by Thistlethwaite and Campbell (1960) as a way of estimating treatment effects in a non-experimental setting where a treatment is determined depending on whether a variable (forcing variable) is above a threshold or cutoff point. The RD design has become an increasingly popular method of exploring a number of empirical problems, including questions dealing with education and health issues.

For pedagogical purposes assume that all students who score above 1000 on their SAT's attend an IVY league school and all students scoring below 1000 attend Slippery Rock State. To determine the value of an IVY league degree in the market place one might be tempted to compare average salaries of graduates; however, doing so would fail to account for differences in intelligence, family background, and a host of other factors.

An approach which could account for inherent differences in the educational programs is to select individuals who scored just above and just below 1000. These individuals are likely similar across other dimensions, only differing slightly in their SAT performance. The essential features of the RD design is (1) a continuous forcing or sorting variable, (2) all persons on one side of the cutoff are assigned to one group, and (3) all persons on the other side are assigned to another group.

This model could be formulated as follows:

$$Y_i = \alpha + X_i\beta + D_i\gamma + \varepsilon_i$$

where Y denotes salary, X represents a vector of other explanatory factors, and $D_i = 1$ if the forcing variable (SAT score in this example) is greater than 1000 and 0 otherwise. The coefficient of D would represent the value of an IVY league degree.

Lee and Lemieux (2009, NBER, working paper 14723) provide an excellent users guide to Regression Discontinuity models. A couple of interesting observations emerge from the work of Lee and Lemieux: (1) "When optimizing agents do not have precise control over the forcing variable, then the variation in the treatment will be as good as randomized in a neighborhood around the discontinuity threshold." and (2) "The distribution of observed baseline covariates should not change discontinuously at the threshold."

How could a RD model be used to determine the impact of a degree from BYU upon different measures of output such as salary and church activity?

12. Exercises: Generalized Regression Model

1. Consider the classical normal linear regression model

$$Y = X\beta + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2 I)$

and the X's satisfy (A.2).

Unbiased estimators of β and σ^2 are given by

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$s^2 = SSE/N\text{-}K$$

where $SSE = (Y - \hat{Y})'(Y - \hat{Y}) = e'e$

$$= (Y - X\hat{\beta})'(Y - X\hat{\beta}) .$$

a. Demonstrate that $R^2 = 1$ if K=N.

Hint: $R^2 = 1\text{-}(SSE/SST)$, demonstrate that $Y = X\hat{\beta}$; hence SSE=0.

b. Demonstrate that e can be written as

$$e = (I\text{-}X(X'X)^{-1}X')Y.$$

The distribution of the estimated residuals e is

$$N(0,\sigma^2 (I\text{-}X(X'X)^{-1}X'))$$

because $E(e) = (I\text{-}X(X'X)^{-1}X')\ E(Y)$

$$= (I\text{-}X(X'X)^{-1}X')\ X\beta$$

$$= 0$$

$$Var(e) = (I\text{-}X(X'X)^{-1}X')\ Var(Y)(I\text{-}X(X'X)^{-1}X')$$

$$= (I-X(X'X)^{-1}X')(\sigma^2 I)(I-X(X'X)^{-1}X')$$

$$= \sigma^2(I-X(X'X)^{-1}X').$$

What can be said about the elements on the main diagonal of $\sigma^2(I-X(X'X)^{-1}X')$, are they equal? Will the off diagonal elements be zero? What implications does this result have for determining whether the unobserved random disturbances (ε) are homoskedastic or nonautocorrelated, based upon an analysis of the estimated random disturbances?

c. Demonstrate that s^2 can be expressed as

$$s^2 = Y'(I-X(X'X)^{-1}X')Y/N-K$$

$$= \varepsilon'(I-X(X'X)^{-1}X')\varepsilon/N-K.$$

d. Recall that $I-X(X'X)^{-1}X'$ is symmetric and idempotent. Demonstrate that the rank of $I-X(X'X)^{-1}X'$ is N-K.

e. Verify that

$$\frac{\varepsilon'(I-X(X'X)^{-1}X')\varepsilon}{\sigma^2} = \frac{Y'(I-X(X'X)^{-1}X')Y}{\sigma^2} = \frac{(N-K)s^2}{\sigma^2}$$

is distributed as a $\chi^2(N-K)$.

Note that this implies

$$\frac{(N-K)s^2}{\sigma^2} = \frac{(N-K)s^2 a_{ii}}{\sigma^2 a_{ii}} = \frac{(N-K)s^2_{\beta_i}}{\sigma^2_{\beta_i}} \sim \chi^2(N-K)$$

where a_{ii} denotes the i^{th} diagonal element of $(X'X)^{-1}$ and

f. Using the results in (e), demonstrate that s^2 is an unbiased estimator of σ^2. Hint: Take the expected value of the first equation in problem (e)

g. Demonstrate that $\hat{\beta}$ and s^2 are independent. This will impl $\hat{\beta}$ and $(s^2_{\hat{\beta}_i})$ are independent.

Hint: $\hat{\beta} = (X'X)^{-1}X'Y = AY$

$(N\text{-}K)\, s^2 = Y'(I - X(X'X)^{-1}X')Y = Y'BY.$

(See Statistics B.2–Hint: AB=0)

h. Indicate what the distribution of the following statistics is and provide support for your answer:

(1) $$\frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} = \frac{(\hat{\beta}_i - \beta_i)/\sigma_{\hat{\beta}_i}}{\sqrt{\left(\dfrac{(N-K)s^2_{\hat{\beta}_i}}{\sigma^2_{\hat{\beta}_i}}\right)/(N-K)}}$$ ~

(2) $$\frac{R^2/(K-1)}{(1-R^2)/(N-K)} = \frac{SSR/(K-1)}{SSE/(N-K)} = \frac{\left(\dfrac{SSR}{\sigma^2}\right)/(K-1)}{\left(\dfrac{SSE}{\sigma^2}\right)/(N-K)}$$ ~

Hint: $SSR/\sigma^2 \sim \chi^2(K\text{-}1)$. Also see Statistics Notes (2. Multivariate Distributions, c. Distribution Theory, (4) and (5))

2. In the generalized normal regression model

$$Y = X\beta + \varepsilon$$

$$\varepsilon \sim N(0, \Sigma)$$

where the X's satisfy A.2, the least squares and MLE of β are given by

$$\hat{\beta} = (X'X)^{-1}X'Y \sim N[\beta,(X'X)^{-1}X'\Sigma X(X'X)^{-1}]$$

$$\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \sim N[\beta, (X'\Sigma^{-1}X)^{-1}].$$

Demonstrate that the estimators yield identical results and hence have the same distribution if the random disturbances are independent and have constant variance, i.e., $\Sigma = \sigma^2 I$.

3. Verify the following for the classical normal linear regression model ($\Sigma = \sigma^2 I$).

(a) $$\frac{d\ell}{d\beta} = X'\varepsilon/\sigma^2$$

(b) $$E\left(\frac{d\ell}{d\beta}\frac{d\ell}{d\beta'}\right) = X'X/\sigma^2$$

(c) $$-E\left[\frac{d^2\ell}{d\beta d\beta'}\right] =$$

(d) What is the relationship between (b), (c), the Cramer-Rao matrix and the variance of $\tilde{\beta}$?

(e) Obtain an expression for the sandwich estimator of the variance. The sandwich estimator is given by

$$\left(E\frac{d^2\ell}{d\beta d\beta'}\right)^{-1}\left(E\frac{d\ell}{d\beta}\frac{d\ell}{d\beta'}\right)\left(E\frac{d^2\ell}{d\beta d\beta'}\right)^{-1}$$

4. Consider the case of heteroskedasticity,

$$Y_t = \beta_1 + \beta_2 X_{t2} + \ldots + \beta_K X_{tK} + \varepsilon_t$$

$$= (1, X_{t2}, \ldots, K_{tK})\,\beta + \varepsilon_t$$

$$= X_t\beta + \varepsilon_t$$

where $\mathrm{Var}(\varepsilon) = \begin{pmatrix} \mathrm{Var}(\varepsilon_1) & 0 & \cdots & 0 \\ 0 & \mathrm{Var}(\varepsilon_2) & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & \cdots & & \mathrm{Var}(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \cdots & & \sigma_N^2 \end{pmatrix}$

Demonstrate that an application of least squares to the transformed model

$$Y_t/\sigma_t = \beta_1(1/\sigma_t) + \beta_2(X_{t2}/\sigma_t) + \ldots + \beta_K\,(X_{tk}/\sigma_t) + \varepsilon_t/\sigma_t$$

$$= (X_t/\sigma_t)\beta + \varepsilon_t/\sigma_t$$

or

$$\begin{pmatrix} \frac{1}{\sigma_1} & 0 & & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & & \cdots & \frac{1}{\sigma_N} \end{pmatrix}\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_N} \end{pmatrix}\begin{pmatrix} 1 & X_{12} & \cdots & X_{1K} \\ 1 & X_{22} & \cdots & X_{2K} \\ & \vdots & & \vdots \\ 1 & X_{N2} & \cdots & X_{NK} \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}$$

$$+\begin{pmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & \frac{1}{\sigma_N} \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

or $\quad TY = TX\beta + T\varepsilon$

yields maximum likelihood estimators

$$\tilde{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

Hint: T'T is proportional to Σ^{-1}.

5. Verify that T'T is proportional to Σ^{-1} for the case of random disturbances of the form AR(1), hence $\hat{\beta}_T = \tilde{\beta}_{MLE}$.

Optional problem. Now, delete the first row from T to form T* and evaluate T*'T* and compare this result with Σ^{-1}. What implications do these results have with respect to the relationship between the Prais-Winsten, Cochrane Orcutt and maximum likelihood estimators?

6. Consider the returns data in S&P500.dat

a. Investigate the distributional characteristics (mean, variance, skewness, and kurtosis) of the returns data. You might consider using the command, "sum y,detail."

b. Use the reg command, reg y, to estimate the sample mean and to calculate deviations from the sample mean, with predict e, resid.

(1) Test for the presence of autocorrelation using

(a) the Durbin Watson test statistic and

(b) a t-statistic based on regressing the OLS estimated errors on their lags

(2) Reconcile your answers to (1) (a-b). You might use a non-parametric test, like the runs test, runtest e.

(3) Test for ARCH behavior using

(a) the estat command with one lag

(b) the estat command with fourteen lags

c. Taking account of possible ARCH/GARCH behavior, estimate the mean return.

7. Instrumental variables

Consider the model defined by

$$\log(wage) = \beta_1 + \beta_2 educ + \beta_3 \exp er + \beta_4 \exp er^2 + \varepsilon_t$$

In models like this it is often argued that education (educ) can be thought of as an endogenous regressor. One approach to circumventing this problem is to implement the method of instrumental variables. Two instrumental variables which have been considered are mother's education and father's education.

a. Using the mroz data set (mroz.dta), estimate this regression model using

(1) OLS

(2) Instrumental variables estimation with Z=fatheduc

(3) Instrumental variables estimation with Z={motheduc, fatheduc}

b. Compare the coefficients obtained from (1), (2), and (3) above and comment on the relative merits of each. Which estimates would you feel most comfortable with (why)?

c. Comparing the results from (1) and (3), perform a Hausman test of the assumption that education is uncorrelated with the error term.

d. An alternative to the Hausman Test is to (1) calculate the OLS residuals obtained from regressing education on experience, experience^2, mother's education, and father's education; (2) include these residuals as a potential regressor in the equation of interest (the log wage equation), and (3) perform a test of the statistical significance of the corresponding coefficient. Perform this test and compare the results with those obtained from the Hausman Test.

8. Verify that the least squares and SURE yield the same results if either

(a) $\sum = \sigma^2 I$ or

(b) $X_i = X_j$ for all i and j.

Hint: (a) and (b) are equivalent to

(a)' $\Omega = \sigma^2 I \otimes I = \sigma^2 I$

or

$$(b)' \quad X = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & & & 0 \\ \vdots & X_2 & & \vdots \\ 0 & 0 & \cdots & X_G \end{pmatrix} = I \otimes X_*$$

where $X_* = X_1 = \ldots = X_G$

Hint: Use the properties of Kronnecker products.

9. Use the data (statatext.txt) in the class notes for Generalized Regression.

It might be helpful to learn some STATA conversion procedures to create a STATA data file. Go onto Blackboard and copy the table on page 62 of the pdf to the Clipboard. Go into Excel and past the data. With the data blocked, go to Data? Text to Columns? Delimited? Space? Finish. Then the data will appear in columns. Copy it to the clipboard, open Stata, go to the Data Editor, paste in your data and close the data editor. As a former TA, Christopher Palmer would say, “Voila, you showed that text data whose boss.”

Using this data estimate and contrast the results obtained from the following commands:

reg y x d1 d2 d3

xtgls y x, i(code)

xtgls y x d1 d2 d3, i(code)

xtreg y x,fe i(code)

xtreg y x,be i(code)

xtreg y x,re i(code)

10. Another panel data problem

Consider the following data:

Year	Firm	Cost	Output	D1	D2	D3	D4	D5	D6
1955	1	3.154	214	1	0	0	0	0	0
1960	1	4.271	419	1	0	0	0	0	0
1965	1	4.584	588	1	0	0	0	0	0
1970	1	5.849	1025	1	0	0	0	0	0
1955	2	3.859	696	0	1	0	0	0	0
1960	2	5.535	811	0	1	0	0	0	0
1965	2	8.127	1640	0	1	0	0	0	0
1970	2	10.966	2506	0	1	0	0	0	0
1955	3	19.035	3202	0	0	1	0	0	0
1960	3	26.041	4802	0	0	1	0	0	0
1965	3	32.444	5821	0	0	1	0	0	0
1970	3	41.180	9275	0	0	1	0	0	0
1955	4	35.229	5668	0	0	0	1	0	0
1960	4	51.111	7612	0	0	0	1	0	0
1965	4	61.045	10206	0	0	0	1	0	0
1970	4	77.885	13702	0	0	0	1	0	0
1955	5	33.154	6000	0	0	0	0	1	0
1960	5	40.044	8222	0	0	0	0	1	0
1965	5	43.125	8484	0	0	0	0	1	0
1970	5	57.727	10004	0	0	0	0	1	0
1955	6	73.050	11796	0	0	0	0	0	1
1960	6	98.846	15551	0	0	0	0	0	1
1965	6	138.880	27218	0	0	0	0	0	1
1970	6	191.560	30958	0	0	0	0	0	1

a. Generate y =log(cost) and x=log(output) and perform *pooled OLS (POLS)* by estimating

$$y_{it} = \beta_1 + \beta_2 x + \varepsilon_{it}$$

using the Stata command **reg y x**. This formulation assumes that the coefficients are the same over time and across firms, uncorrelated for a given individual, with the ε_{it} being homosckedastic over time and across individuals.

b. Now relax the assumption of identical coefficients and allow different firms to have the same slopes, but with possibly different intercepts,

$$y_{it} = \beta_i + \beta_2 x + \varepsilon_{it} = \alpha + \beta_2 x + \alpha_i + \varepsilon_{it}$$
$$\left(where \sum_i \alpha_i = 0\right)$$
.

(1) Perform the following Stata estimations and explain the results
xtset firm
reg y x D1 D2 D3 D4 D5 or reg y x D1 D2 D3 D4 D5 D6, noconstant
xtreg y x, fe

This estimator is referred to as the *fixed effects estimator* or the *within groups estimator*. It can be obtained by estimating a regular regression with binary variables for the intercept or by regression the deviations of the dependent variable from their individual means means on deviations of the explanatory variables from their individual means.

(2) Test the hypothesis that the intercepts are the same for each firm. Use a Chow test and look for related information on the fixed effects printout.

c. Another estimator that is sometimes considered is obtained by regressing the average y for each firm on the average x's for each firm. This estimator is called the *between-groups estimator* and can be obtained in Stata by using the command

xtreg y x, be

Perform this estimation and compare the results with those obtained in b(1).

d. Consider the case where the α_i's are not constant for each firm, but may vary. Then $Var(\alpha_i + \varepsilon_i) = \sigma_\varepsilon^2 I_T + \sigma_u^2 i_T i_T' = \Sigma$ ε $\Omega = I_m \otimes \Sigma$ and var( )= assumes that the α_i's ε's , , and x's are all independent. The corresponding GLS estimator is called the random effects estimator and can be obtained using the Stata command,

xtreg y x, re

Obtain the random effects estimators of the cost function.

e. Use the Hausman test to explore the use of fixed vs. random effects estimators.

11. (Just for fun, but not required)

Consider the model defined by

$$Y = X\beta + U$$

where

$$U = \sigma \,\mathrm{sign}\,(\mu)\,[\sinh(\theta v) - F(\theta,\mu)]/\theta$$

$$V \sim N[\mu,1]$$

$$E(\sinh(\theta v)) = F(\theta,\mu)$$

$$= e^{\theta^2/2}\sinh(\mu\theta)\,.$$

Sin h(s) denotes the hyperbolic sin $[e^{s} + e^{-s}]/2$. This formulation allows for skewness and thick-tailed error distributions.

It can be shown that

$$E(U) = 0$$

$$\mathrm{Var}\,(U) = \frac{\sigma^2[e^{\theta^2} - 1][e^{\theta^2}\cosh(2\mu\theta) + 1]}{2\theta^2}$$

$$\mathrm{Skew}\,(U) = E\,(\mu^3) = -\frac{\mathrm{sign}\,(\mu)\sigma^3}{4\theta^3}\{e^{\theta^2/2}(e^{\theta^2} - 1)^{1/2}\}$$

$$\bullet\,\frac{e^{\theta^2}(e^{\theta^2} + 2)\sinh(3\mu\theta) + 3\sin(\mu\theta)}{(e^{\theta^2}\cosh(2\mu\theta) + 1)^{3/2}}$$

This information is given to help place you in the top .001% of undergraduate economics majors in the world. The previous results can be obtained, with patience, from the result

$$E\,(e^{\theta hv}) = e^{h^2\theta^2/2 + h\mu\theta}$$

where $V \sim N[\mu, \sigma^2 = 1]$. **Derive** this result. Hint: Consider the n

generating function.