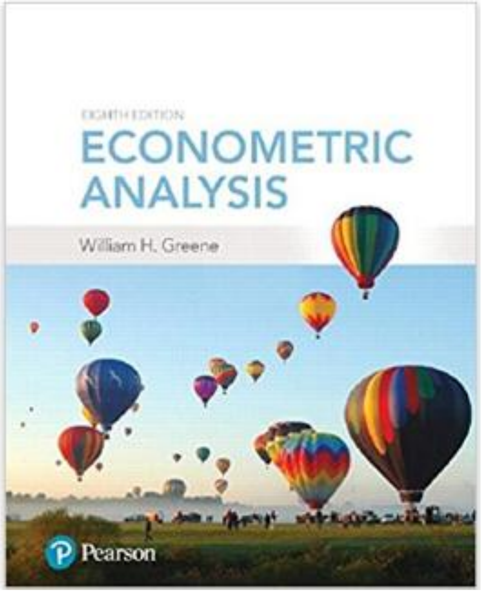
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



Econometrics I

Part 11 – Hypothesis Testing

Classical Hypothesis Testing

We are interested in using the linear regression to support or cast doubt on the validity of a theory about the real world counterpart to our statistical model. The model is used to test hypotheses about the underlying data generating process.

Types of Tests

- **Nested Models**: Restriction on the parameters of a particular model
  $y = \beta_1 + \beta_2 x + \beta_3 T + \varepsilon, \quad \beta_3 = 0$
  (The "treatment" works; $\beta_3 \neq 0$.)
- **Nonnested models**: E.g., different RHS variables
  $y_t = \beta_1 + \beta_2 x_t + \beta_3 \mathbf{x}_{t-1} + \varepsilon_t$
  $y_t = \gamma_1 + \gamma_2 x_t + \gamma_3 \mathbf{y}_{t-1} + w_t$
  (Lagged effects occur immediately or spread over time.)
- **Specification tests**:
  $\varepsilon \sim N[0,\sigma^2]$ vs. some other distribution
  (The "null" spec. is true or some other spec. is true.)

Hypothesis Testing

- Nested vs. nonnested specifications
  - $y=b_1x+e$ vs. $y=b_1x+b_2z+e$: Nested
  - y=bx+e vs. y=cz+u: Not nested
  - y=bx+e vs. logy=clogx: Not nested
  - y=bx+e; e ~ Normal vs. e ~ t[.]: Not nested
  - Fixed vs. random effects: Not nested
  - Logit vs. probit: Not nested
  - x is (not) endogenous: Maybe nested. We'll see …
- Parametric restrictions
  - Linear: $\mathbf{R}\beta\text{-}\mathbf{q} = \mathbf{0}$, **R** is JxK, J < K, full row rank
  - General: $\mathbf{r}(\beta,\mathbf{q}) = \mathbf{0}$, **r** = a vector of J functions,

    $\mathbf{R}(\beta,\mathbf{q}) = \partial\mathbf{r}(\beta,\mathbf{q})/\partial\beta'$.
  - Use $\mathbf{r}(\beta,\mathbf{q})=\mathbf{0}$ for linear and nonlinear cases

Broad Approaches

- **Bayesian: Does not reach a firm conclusion. Revises odds.**
  - Prior odds compares strength of prior beliefs in two states of the world
  - Posterior odds compares revised beliefs
  - Symmetrical treatment of competing ideas
  - Not generally practical to carry out in meaningful situations
- **Classical: All or nothing; reject the theory or do not reject it.**
  - "Null" hypothesis given prominence
  - Propose to "reject" toward favor of "alternative"
  - Asymmetric treatment of null and alternative
  - Huge gain in practical applicability

Inference in the Linear Model

Formulating hypotheses: linear restrictions as a general framework

Hypothesis Testing J linear restrictions

Analytical framework: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Hypothesis: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$,

Substantive restrictions: What is a "testable hypothesis?"

- Substantive restriction on parameters
- Reduces dimension of parameter space
- Imposition of restriction degrades estimation criterion

Testable Implications of a Theory

Investors care about nominal interest rates and expected inflation:

$$I = \beta_1 + \beta_2 r + \beta_3 dp + e$$

Restriction is $\beta_3 = -\beta_2$.

Investors care only about real interest rates.

(1) Substantive restriction on parameters

(2) Parameter space is $\Re^3$. Restricted space is a 2 dimensional subspace (not $\Re^2$).

(3) Restrictions must lead to increase in sum of squared residuals

The General Linear Hypothesis: H_0: Rβ - q = 0

A unifying departure point: Regardless of the hypothesis, least squares is unbiased.

$$E[\mathbf{b}] = \beta$$

The hypothesis makes a claim about the population

Rβ – q = 0. Then, if the hypothesis is true, E[Rb – q] = 0.

The sample statistic, Rb – q will not equal zero.
Two possibilities:
- **Rb – q is small enough to attribute to sampling variability**
- **Rb – q is too large (by some measure) to be plausibly attributed to sampling variability**

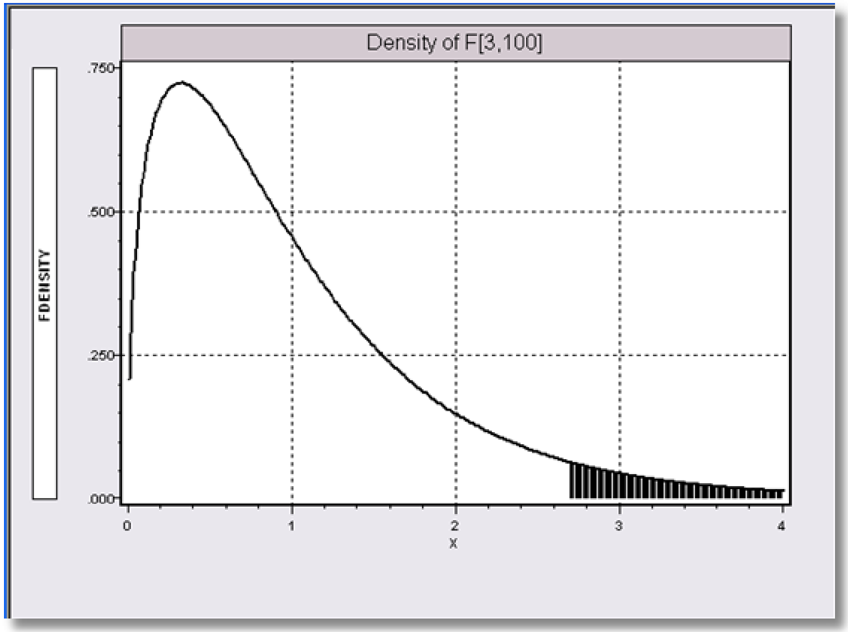
Large Rb – q is the rejection region.

Neyman – Pearson Classical Methodology

- Formulate null and alternative hypotheses
  - Hypotheses are exclusive and exhaustive
  - Null hypothesis is of particular interest
- Define “Rejection” region = sample evidence that will lead to rejection of the null hypothesis.
- Gather evidence
- Assess whether evidence falls in rejection region or not.

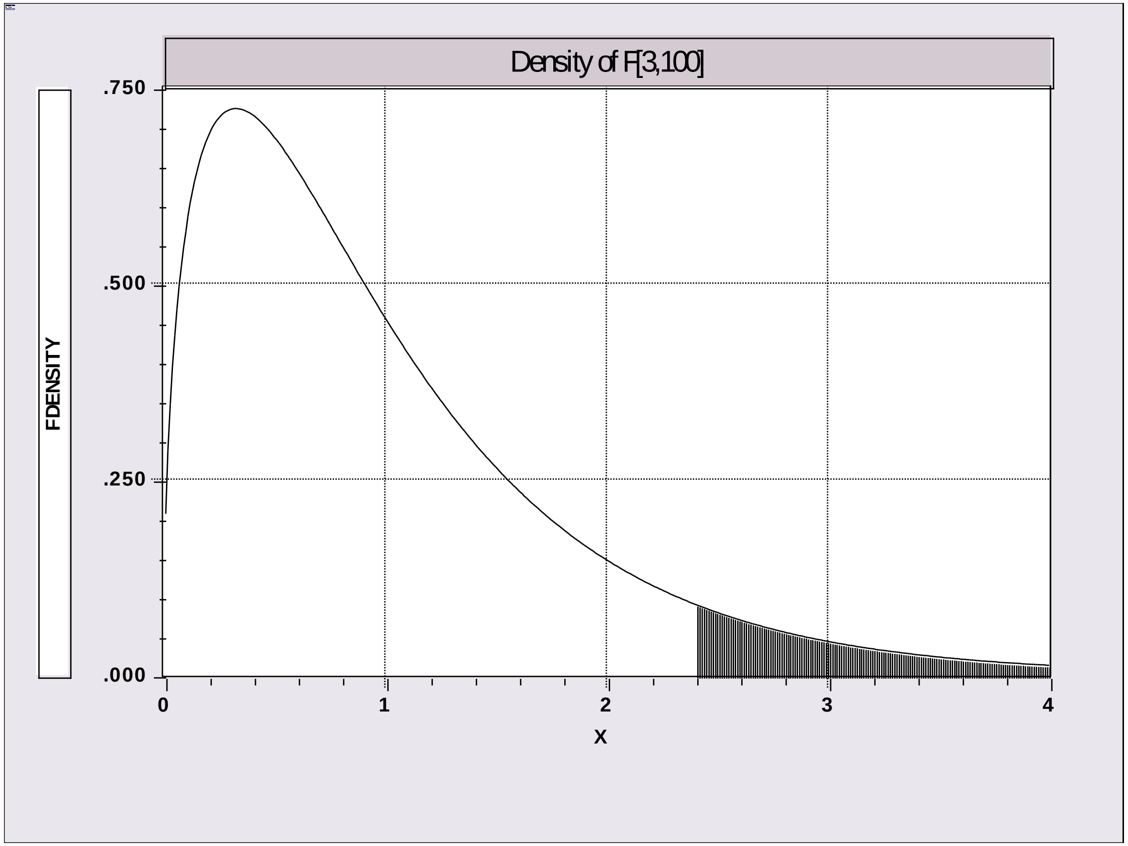
Testing Fundamentals - I

- **SIZE of a test** = Probability it will incorrectly reject a “true” null hypothesis.
- This is the probability of a Type I error.



Under the null hypothesis, F(3,100) has an F distribution with (3,100) degrees of freedom. Even if the null is true, F will be larger than the 5% critical value of 2.7 about 5% of the time.

Distribution Under the Null

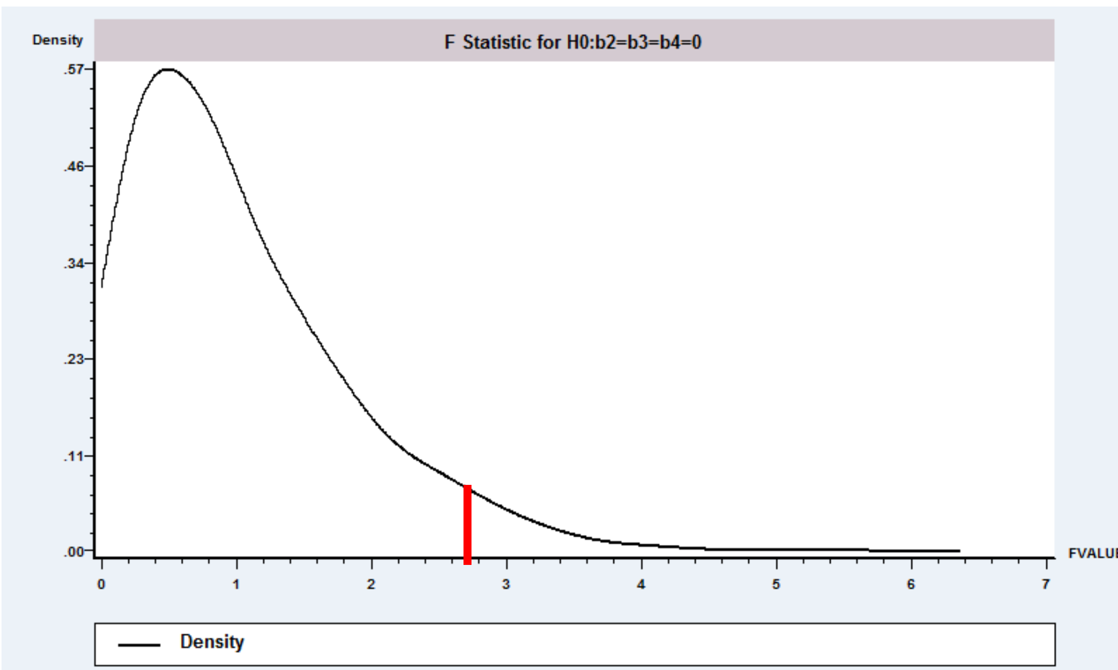


A Simulation Experiment

```
sample    ; 1 - 500 $
create    ; x1=rnn(0,1);x2=rnn(0,1);x3=rnn(0,1)$
matrix    ; fvalues=init(1000,1,0)$
proc$
create    ; fakey = rnn(-.3,1.5) $      $ Coefficients all = 0
regress   ; quietly ; lhs = fakey       ?     Compute regression
          ; rhs=one,x1,x2,x3$
calc      ; fstat = (rsqrd/3)/((1-rsqrd)/(n-4))$  Compute F
matrix    ; fvalues(i)=fstat$           ?  Save 1000 Fs
endproc
execute   ; i= 1,1000 $                 ? 1000 replications
kernel    ; rhs = fvalues
          ; title=F Statistic for H0:b2=b3=b4=0 $
quantile  ; rhs=fvalues$
```

Simulation Results

About 5% of computed F values are in the rejection region, though $\beta_1=\beta_2=\beta3=0$ is true.

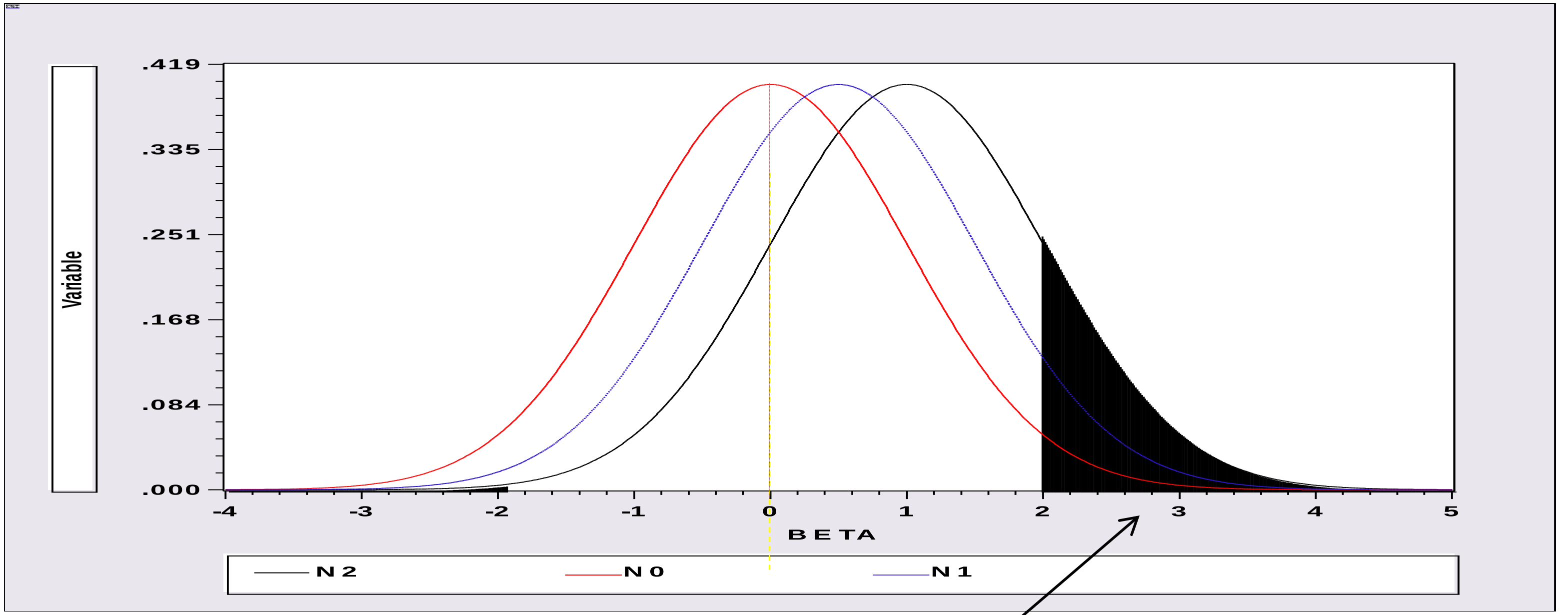


Percentiles	FVALUES:1
Sample size	1000
Min.	0.003284
01th	0.028683
*025	0.070420
05th	0.115276
10th	0.182580
20th	0.309095
25th	0.376636
30th	0.444732
40th	0.633529
Med.	0.798674
60th	0.951718
70th	1.262215
75th	1.392785
80th	1.592293
90th	2.086798
95th	2.626567
*975	2.992109
99th	3.494726
Max.	6.182722

Testing Fundamentals - II

- **POWER of a test** = the probability that it will correctly reject a “false null” hypothesis
- This is 1 – the probability of a Type II error.
- The power of a test depends on the specific alternative.

Power of a Test



Null: Mean = 0. Reject if observed mean < -1.96 or > +1.96.

Prob(Reject μ = 0 | μ = 0) = 0.05

Prob(Reject μ = 0 | μ =.5) = 0.07902

Prob(Reject μ = 0 | μ = 1) = 0.170066. Increases as the (alternative) μ rises.

3 Approaches to Defining the Rejection Region

(1) Imposing the restrictions leads to a loss of fit. Sum of squared residuals must increase.
R^2 must go down. Does it go down "a lot?" (I.e., significantly?).
R_u^2 = unrestricted model, R_r^2 = restricted model fit. Test is based on $R_u^2 – R_r$. Is this difference large?

(2) Is **Rb - q** close to **0**? Basing the test on the discrepancy vector: **m** = **Rb - q.**
Using the *Wald criterion*: $\mathbf{m}'(\text{Var}[\mathbf{m}])^{-1}\mathbf{m}$. A distance measure of how far m is from zero.

(3) Does the restricted model appear to satisfy the restrictions? Examine the residuals from the restricted model. Do the residuals appear to be random noise?

Testing Strategy

How to determine if the statistic is 'large.'

Need a 'null distribution.'

If the hypothesis is true, then the statistic will have a certain distribution. This tells you how likely certain values are, and in particular, if the hypothesis is true, then 'large values' will be unlikely.

If the observed statistic is too large, conclude that the assumed distribution must be incorrect and the hypothesis should be rejected.

Robust Tests

- The Wald test generally will (when properly constructed) be more robust to failures of the narrow model assumptions than the t or F
- Reason: Based on “robust” variance estimators and asymptotic results that hold in a wide range of circumstances.

Dear Professor Greene,

I recently submitted a paper in a journal estimating a wage equation corrected for sample selection. The referee has come back with the following comment:

"*I do not trust the standard errors reported in Table 5. Most of the coefficients are highly statistically significant (at the 1% level), while the R-squares are still relatively low. I suspect this is because the standard error are not corrected for heteroscedasicity. The author should report robust standard errors at the minimum. Standard errors allowing for occupational clustering would be even better.*"

What command does one use to calculate robust standard error in the sample selection model?.

Also, if there are 6 occupations included in the regression equation, what command should one use for standard errors allowing for occupational clustering.

In both cases, assume we are estimating an earnings function corrected for sample selection.

Robustness

- Assumptions are narrower than necessary
  - (1) Disturbances might be heteroscedastic
  - (2) Disturbances might be correlated across observations – these are panel data
  - (3) Normal distribution assumption is unnecessary
- F, LM and LR tests rely on normality, no longer valid
- Wald test relies on appropriate covariance matrix. (1) and (2) invalidate $s^2(\mathbf{X'X})^{-1}$.

Robust Inference Strategy

(1) Use a robust estimator of the asymptotic covariance matrix. (Next class)

(2) The Wald statistic based on an appropriate covariance matrix is robust to distributional assumptions – it relies on the CLT.

The Nonrobust F Statistic

An application: (Familiar) Suppose $\mathbf{b}_n$ is the least squares estimator of $\boldsymbol{\beta}$ based on a sample of n observations. No assumption of normality of the disturbances or about nonstochastic regressors is made. The standard F statistic for testing the hypothesis H0: $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$ is

$$F[J, n-K] = [(\mathbf{e}^{*\prime}\mathbf{e}^{*} - \mathbf{e}'\mathbf{e})/J] \;/\; [\mathbf{e}'\mathbf{e} / (n-K)]$$

where this is built of two sums of squared residuals. The statistic does not have an F distribution. How can we test the hypothesis?

Application - 1

Regression Model

$$\text{LogG} = \beta_1 + \beta_2\text{logY} + \beta_3\text{logPG} + \beta_4\text{logPNC} + \beta_5\text{logPUC} + \beta_6\text{logPPT} + \beta_7\text{logPN} + \beta_8\text{logPD} + \beta_9\text{logPS} + \varepsilon$$

Period = 1960 - 1995. Note that all coefficients in the model are elasticities.

Full Model by Least Squares

```
----------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LG       Mean                 =         5.39299
             Standard deviation   =          .24878
             Number of observs.   =              36
Model size   Parameters           =               9
             Degrees of freedom   =              27
Residuals    Sum of squares       =          .00855
             Standard error of e  =          .01780
Fit          R-squared            =          .99605
             Adjusted R-squared   =          .99488
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]   Mean of X
--------+-------------------------------------------------------------
Constant|   -6.95326***     1.29811       -5.356   .0000
      LY|    1.35721***      .14562        9.320   .0000      9.11093
     LPG|    -.50579***      .06200       -8.158   .0000       .67409
    LPNC|    -.01654         .19957        -.083   .9346       .44320
    LPUC|    -.12354*        .06568       -1.881   .0708       .66361
    LPPT|     .11571         .07859        1.472   .1525       .77208
     LPN|    1.10125***      .26840        4.103   .0003       .60539
     LPD|     .92018***      .27018        3.406   .0021       .43343
     LPS|   -1.09213***      .30812       -3.544   .0015       .68105
--------+-------------------------------------------------------------
```

Testing a Hypothesis Using a Confidence Interval

Given the range of plausible values

Testing the hypothesis that a coefficient equals zero or some other particular value:

Is the hypothesized value in the confidence interval?

Is the hypothesized value within the range of plausible values?

If not, reject the hypothesis.

Test About One Parameter

Is the price of public transportation really 'relevant?' $H_0 : \beta_6 = 0$.

Confidence interval: $b_6 \pm t(.95,27) \times$ Standard error

$= .11571 \pm 2.052(.07859)$

$= .11571 \pm .16127 = (-.045557, .27698)$

Contains 0.0. Do not reject hypothesis $H_0 : \beta_6 = 0$.

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	-6.95326***	1.29811	-5.356	.0000	
LY	1.35721***	.14562	9.320	.0000	9.11093
LPG	-.50579***	.06200	-8.158	.0000	.67409
LPNC	-.01654	.19957	-.083	.9346	.44320
LPUC	-.12354*	.06568	-1.881	.0708	.66361
LPPT	.11571	.07859	1.472	.1525	.77208
LPN	1.10125***	.26840	4.103	.0003	.60539
LPD	.92018***	.27018	3.406	.0021	.43343
LPS	-1.09213***	.30812	-3.544	.0015	.68105

Test a Hypothesis About a Coefficient

Confidence intervals are often displayed in results

```
----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=logG    Mean                     =        5.39299
            Standard deviation       =         .24878
----------  No. of observations      =             36  DegFreedom   Mean square
Regression  Sum of Squares           =        2.15567           6       .35928
Residual    Sum of Squares           =    .105227E-01          29       .00036
Total       Sum of Squares           =        2.16619          35       .06189
----------  Standard error of e      =         .01905  Root MSE             .01710
Fit         R-squared                =         .99514  R-bar squared        .99414
Model test  F[  6,    29]            =      990.14814  Prob F > F*          .00000
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    logG|  Coefficient       Error        t    |t|>T*        Interval
--------+-------------------------------------------------------------------
Constant|   -23.9398*      12.19453    -1.96  .0593    -48.8804     1.0008
    YEAR|     .00902         .00677     1.33  .1931     -.00483     .02287
   logPG|    -.47878***      .06162    -7.77  .0000     -.60482    -.35275
    logY|    1.27365***      .15010     8.49  .0000      .96666    1.58064
   logPN|    1.13036***      .18906     5.98  .0000      .74369    1.51703
   logPD|     .79402***      .22581     3.52  .0015      .33217    1.25586
   logPS|   -1.19347***      .20592    -5.80  .0000    -1.61462    -.77233
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

This model does not contain the micro price indices (PNC, PUC, PPT).

A Distance Measure

Testing more generally about a single parameter.

Sample estimate is b_k Sample estimated standard error is v_k.

Hypothesized value is β_k

How far is β_k from b_k? If too far, the hypothesis is inconsistent with the sample evidence.

Measure distance in standard error units

$$\mathbf{t = (b_k - \beta_k)/Estimated\ v_k.}$$

If t is "large" (larger than critical value), reject the hypothesis. The critical value is obtained in a table (computed) for the t distribution.

```
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=logG     Mean                 =         5.39299
             Standard deviation   =          .24878
----------   No. of observations  =              36  DegFreedom   Mean square
Regression   Sum of Squares       =         2.15567          6        .35928
Residual     Sum of Squares       =     .105227E-01         29        .00036
Total        Sum of Squares       =         2.16619         35        .06189
----------   Standard error of e  =          .01905  Root MSE             .01710
Fit          R-squared            =          .99514  R-bar squared        .99414
Model test   F[  6,    29]        =       990.14814  Prob F > F*          .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
    logG|  Coefficient       Error       t    |t|>T*          Interval
--------+--------------------------------------------------------------------
Constant|   -23.9398*      12.19453    -1.96  .0593    -48.8804     1.0008
    YEAR|      .00902        .00677     1.33  .1931     -.00483     .02287
   logPG|    -.47878***      .06162    -7.77  .0000     -.60482    -.35275
    logY|    1.27365***      .15010     8.49  .0000      .96666    1.58064
   logPN|    1.13036***      .18906     5.98  .0000      .74369    1.51703
   logPD|     .79402***      .22581     3.52  .0015      .33217    1.25586
   logPS|   -1.19347***      .20592    -5.80  .0000    -1.61462    -.77233
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Test Statistic Based on Fit Measures

For the fit measures, use a normalized measure of the loss of fit:

$$F[J,n-K] = \frac{\left(R_u^2 - R_r^2\right)/J}{\left(1-R_u^2\right)/(n-K)} \geq 0 \text{ since } R_u^2 \geq R_r^2$$

Often useful

$$R_u^2 = 1 - \frac{\mathbf{e}_u'\mathbf{e}_u}{S_{yy}} \text{ and } R_r^2 = 1 - \frac{\mathbf{e}_r'\mathbf{e}_r}{S_{yy}}$$

Insert these in F and it becomes

$$F[J,n-K] = \frac{\left(\mathbf{e}_r'\mathbf{e}_r - \mathbf{e}_u'\mathbf{e}_u\right)/J}{\left(\mathbf{e}_u'\mathbf{e}_u\right)/(n-K)} \geq 0 \text{ since } \mathbf{e}_r'\mathbf{e}_r \geq \mathbf{e}_u'\mathbf{e}_u$$

Hypothesis Test: Sum of Coefficients = 0?

```
----------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LG       Mean                  =       5.39299
             Standard deviation    =        .24878
             Number of observs.    =            36
Model size   Parameters            =             9
             Degrees of freedom    =            27
Residuals    Sum of squares        =        .00855  <*******
             Standard error of e   =        .01780
Fit          R-squared             =        .99605  <*******
             Adjusted R-squared    =        .99488
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]   Mean of X
--------+-------------------------------------------------------------
Constant|   -6.95326***      1.29811       -5.356   .0000
      LY|    1.35721***       .14562        9.320   .0000      9.11093
     LPG|    -.50579***       .06200       -8.158   .0000       .67409
    LPNC|    -.01654          .19957        -.083   .9346       .44320
    LPUC|    -.12354*         .06568       -1.881   .0708       .66361
    LPPT|     .11571          .07859        1.472   .1525       .77208
     LPN|    1.10125***       .26840        4.103   .0003       .60539
     LPD|     .92018***       .27018        3.406   .0021       .43343
     LPS|   -1.09213***       .30812       -3.544   .0015       .68105
--------+-------------------------------------------------------------
```

Sum = 0.9293. Significantly different from 0.0000?

Restricted Regression

```
----------------------------------------------------------------------
Linearly restricted regression
LHS=LG       Mean                 =    5.392989
             Standard deviation   =    .2487794
             Number of observs.   =          36
Model size   Parameters           =           8  <*** 9 – 1 restriction
             Degrees of freedom   =          28
Residuals    Sum of squares       =    .0112599  <*** With the restriction
Residuals    Sum of squares       =    .0085531  <*** Without the restriction
Fit          R-squared            =    .9948020
Restrictns.  F[  1,    27] (prob) =    8.5(.01)
Not using OLS or no constant.R2 & F may be < 0
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]  Mean of X
--------+-------------------------------------------------------------
Constant|   -10.1507***       .78756      -12.889   .0000
      LY|    1.71582***       .08839       19.412   .0000      9.11093
     LPG|    -.45826***       .06741       -6.798   .0000       .67409
    LPNC|     .46945***       .12439        3.774   .0008       .44320
    LPUC|    -.01566          .06122        -.256   .8000       .66361
    LPPT|     .24223***       .07391        3.277   .0029       .77208
     LPN|    1.39620***       .28022        4.983   .0000       .60539
     LPD|     .23885          .15395        1.551   .1324       .43343
     LPS|   -1.63505***       .27700       -5.903   .0000       .68105
--------+-------------------------------------------------------------
F = [(.0112599 - .0085531)/1] / [.0085531/(36 – 9)]  =  8.544691
```

Joint Hypotheses

Joint hypothesis: Income elasticity = +1, Own price elasticity = -1.
The hypothesis implies that b2 = 1 and b3 = -1.

$$\log G = \beta_1 + \log Y - \log Pg + \beta_4 \log PNC + \ldots$$

Strategy: Regress logG – logY + logPg on the other variables and compare the sums of squared residualss

With two restrictions imposed			
Residuals	Sum of squares	=	.0286877
Fit	R-squared	=	.9979006
Unrestricted			
Residuals	Sum of squares	=	.0085531
Fit	R-squared	=	.9960515

$$F = ((.0286877 - .0085531)/2) / (.0085531/(36-9)) = 31.779951$$

The critical F for 95% with (2,27) degrees of freedom is 3.354.
The hypothesis is rejected.

Basing the Test on R^2

Based on R^2s,

F = ((.9960515 - .997096)/2)/((1-.9960515)/(36-9))
  = -3.571166 (!)

What's wrong? The unrestricted model used LHS = logG. The restricted one used logG - logY + logPG. The regressions have different LHS variables.

The calculation is always safe using the sums of squared residuals. The calculation is OK if the dependent variable is the same in the two regressions.

An important relationship between t and F

$$F = \frac{\text{Chi} - \text{Squared}[J]/J}{\text{Chi} - \text{squared}[n-K]/(n-K)}$$

where the two chi-squared variables are independent.

If J = 1, i.e., testing a single restriction,

$$F = \frac{\text{Chi} - \text{Squared}[1]/1}{\text{Chi} - \text{squared}[n-K]/(n-K)}$$

$$= \frac{(N[0,1])^2}{\text{Chi} - \text{squared}[n-K]/(n-K)}$$

$$= \left\{ \frac{N[0,1]}{\sqrt{\text{Chi} - \text{squared}[n-K]/(n-K)}} \right\}^2 = \{t[1]\}^2$$

For a single restriction, F[1,n-K] is the square of the t ratio.

For one restriction, $F = t^2$

```
----------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LG       Mean                   =        5.39299
             Standard deviation     =         .24878
             Number of observs.     =             36
Residuals    Sum of squares         =         .00855
             Standard error of e    =         .01780
Fit          R-squared              =         .99605
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]   Mean of X
--------+-------------------------------------------------------------
Constant|   -6.95326***      1.29811       -5.356   .0000
      LY|    1.35721***       .14562        9.320   .0000      9.11093
     LPG|    -.50579***       .06200       -8.158   .0000       .67409
    LPNC|    -.01654          .19957        -.083   .9346       .44320
    LPUC|    -.12354*         .06568       -1.881   .0708       .66361
    LPPT|     .11571          .07859        1.472   .1525       .77208
     LPN|    1.10125***       .26840        4.103   .0003       .60539
     LPD|     .92018***       .27018        3.406   .0021       .43343
     LPS|   -1.09213***       .30812       -3.544   .0015       .68105
--------+-------------------------------------------------------------
```

Regression fit if drop? Without LPPT, R-squared = 0.99573

Compare R^2, was 0.99605,

$F(1,27) = [(.99605 - .99573)/1]/[(1-.99605)/(36-9)]$

$= 2.187$

$= 1.472^2$ (with some rounding difference)

Regression Analysis: Expenditure versus Year, GasPrice, Income, P_NewCars, ...

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	168558	18728.7	5355.77	0.000
Year	1	42	41.7	11.91	0.001
GasPrice	1	1348	1347.7	385.39	0.000
Income	1	91	90.6	25.91	0.000
P_NewCars	1	30	30.0	8.57	0.006
P_UsedCars	1	47	47.5	13.57	0.001
P_PublicTrans	1	0	0.1	0.03	0.865
P_Durables	1	188	187.6	53.65	0.000
P_Nondurables	1	1	1.3	0.37	0.544
P_Services	1	6	5.6	1.60	0.212
Error	42	147	3.5		
Total	51	168705			

F-Value = T-Value^2

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.87000	99.91%	99.89%	99.83%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1596	467	3.42	0.001	
Year	-0.840	0.243	-3.45	0.001	198.49
GasPrice	1.3404	0.0683	19.63	0.000	64.62
Income	0.004522	0.000888	5.09	0.000	354.84
P_NewCars	0.645	0.220	2.93	0.006	974.93
P_UsedCars	0.3079	0.0836	3.68	0.001	265.78
P_PublicTrans	0.0142	0.0830	0.17	0.865	481.06
P_Durables	-1.494	0.204	-7.32	0.000	820.66
P_Nondurables	0.132	0.216	0.61	0.544	1614.88
P_Services	0.174	0.137	1.27	0.212	1229.94

The Wald Statistic

Many test statistics are Wald distance measures

W = (random vector - hypothesized value)' times

[Variance of difference]$^{-1}$ times

(random vector - hypothesized value)

= Normalized distance measure

$= (\mathbf{q} - \theta)'[\text{Var}(\mathbf{q} - \theta)]^{-1}(\mathbf{q} - \theta)$

Under the null hypothesis that $E[\mathbf{q}] = \theta$, W is exactly distributed as chi-squared(J) if

(1) the distance, **q**, is normally distributed and

(2) the variance matrix is the true one, not the estimate.

Wald Test Statistics

$$W = \mathbf{m}'[\text{Est.Var}(\mathbf{m})]^{-1}\mathbf{m}$$

For a single restriction, $m = \mathbf{r}'\mathbf{b} - q$. The variance

is $\mathbf{r}'(\text{Var}[\mathbf{b}])\mathbf{r}$

The distance measure is

$(m$ / standard deviation of $m)^2$.

Example: The standard t test that bk = 0,

$\text{Wald} = [(b_k - 0)/\text{standard error}]^2$.

t^2 is the Wald statistic.

General Result for the Wald Statistic
Full Rank Quadratic Form

A crucial distributional result (exact): If the random vector **x** has a K-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, then the random variable $W = (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ has a chi-squared distribution with K degrees of freedom.

(See text, Appendix B11.6, Theorem B.11.)

Building the Wald Statistic-1

Suppose that the same normal distribution assumptions hold, but instead of the parameter matrix **Σ** we do the computation using a matrix $\mathbf{S}_n$ which has the property plim $\mathbf{S}_n$ = **Σ**. The exact chi-squared result no longer holds, but the limiting distribution is the same as if the true **Σ** were used.

Building the Wald Statistic-2

Suppose the statistic is computed not with an **x** that has an exact normal distribution, but with an $\mathbf{x}_n$ which has a **limiting normal distribution**, but whose finite sample distribution might be something else. Our earlier results for functions of random variables give us the result

$$(\mathbf{x}_n - \boldsymbol{\mu})' \mathbf{S}_n^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) \rightarrow \chi^2[K]$$

(!!!)VVIR! Note that in fact, nothing in this relies on the normal distribution. What we used is consistency of a certain estimator ($\mathbf{S}_n$) and the central limit theorem for $\mathbf{x}_n$.

General Result for Wald Distance

The Wald distance measure: If plim $\mathbf{x}_n = \boldsymbol{\mu}$, $\mathbf{x}_n$ is asymptotically normally distributed with a mean of $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, and if $\mathbf{S}_n$ is a consistent estimator of $\boldsymbol{\Sigma}$, then the Wald statistic, which is a generalized distance measure between $\mathbf{x}_n$ converges to a chi-squared variate.

$$(\mathbf{x}_n - \boldsymbol{\mu})'\mathbf{S}_n^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) \rightarrow \chi^2[K]$$

Test Statistics

We have established the asymptotic distribution of **b**. We now turn to the construction of test statistics. In particular,

$$F[J,n-K] = (1/J)(\mathbf{Rb} - \mathbf{q})'[\mathbf{R}\, s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})$$

This is the usual test statistic for testing linear hypotheses in the linear regression model, distributed exactly as F if the disturbances are normally distributed. We now obtain some general results that will let us construct test statistics in more general situations.

JF is a Wald Statistic

$F[J,n-K] = (1/J) \times (\mathbf{Rb_n} - \mathbf{q})'[\mathbf{R}\, s^2(\mathbf{X'X})^{-1}\, \mathbf{R'}]^{-1}\, (\mathbf{Rb_n} - \mathbf{q})$.

Write $\mathbf{m} = (\mathbf{Rb_n} - \mathbf{q})$. Under the hypothesis, plim $\mathbf{m}=\mathbf{0}$.

$\sqrt{n}\, \mathbf{m} \rightarrow N[0, \mathbf{R}(\sigma^2/n)\mathbf{Q}^{-1}\mathbf{R}']$

Estimate the variance with $\mathbf{R}(s^2/n)(\mathbf{X'X}/n)^{-1}\mathbf{R}']$

Then, $(\sqrt{n}\, \mathbf{m})'\, [\text{Est.Var}(\sqrt{n}\, \mathbf{m})]^{-1}\, (\sqrt{n}\, \mathbf{m})$

fits exactly into the apparatus developed earlier. If plim $\mathbf{b}_n = \boldsymbol{\beta}$, plim $s^2 = \sigma^2$, and the other asymptotic results we developed for least squares hold, then

$JF[J,n-K] \rightarrow \chi^2[J]$.

The Wald Statistic is Robust

Estimate the variance with **R V R**' where **V** is any appropriate (conventional, heteroscedasticity robust, cluster corrected, etc.)

Then, $(\sqrt{n}\,\mathbf{m})'\,[\text{Est.Var}(\sqrt{n}\,\mathbf{m})]^{-1}\,(\sqrt{n}\,\mathbf{m})$

fits exactly into the apparatus developed earlier. If plim $\mathbf{b}_n = \boldsymbol{\beta}$, and the other asymptotic results we developed for least squares hold, then, specifically for the linear regression model,

$$JF[J,n-K] \rightarrow \chi^2[J].$$

(Use JF and critical values for $\chi^2[J]$ for tests.

Hypothesis Test: Sum of Coefficients

Do the three aggregate price elasticities sum to zero?

$H_0 : \beta_7 + \beta_8 + \beta_9 = 0$

$\mathbf{R} = [0, 0, 0, 0, 0, 0, 1, 1, 1], \;\; \mathbf{q} = [0]$

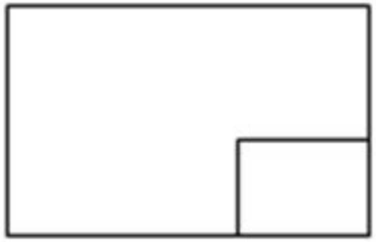
```
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]   Mean of X
--------+-------------------------------------------------------------
 LPN|     1.10125***       .26840         4.103    .0003       .60539
 LPD|      .92018***       .27018         3.406    .0021       .43343
 LPS|    -1.09213***       .30812        -3.544    .0015       .68105
```

	1	2	3	4	5	6	7	8	9
1	1.6851	-0.189024	-0.0256198	-0.218091	-0.0240267	-0.0295907	-0.0261772	0.197857	0.176068
2	-0.189024	0.0212045	0.00290895	0.0243971	0.00269963	0.0032894	0.00280174	-0.0222154	-0.0195876
3	-0.0256198	0.00290895	0.00384368	-0.000682307	-0.000413822	-0.00176052	-0.0114883	-0.0044953	0.0108144
4	-0.218091	0.0243971	-0.000682307	0.0398293	0.00350897	0.00824835	0.0236143	-0.0311143	-0.0453555
5	-0.0240267	0.00269963	-0.000413822	0.00350897	0.00431411	0.001419	0.00979376	-0.0118214	-0.00970482
6	-0.0295907	0.0032894	-0.00176052	0.00824835	0.001419	0.00617673	0.0134911	-0.00740557	-0.0198458
7	-0.0261772	0.00280174	-0.0114883	0.0236143	0.00979376	0.0134911	0.0720371	-0.0335608	-0.0705545
8	0.197857	-0.0222154	-0.0044953	-0.0311143	-0.0118214	-0.00740557	-0.0335608	0.0729982	0.0346625
9	0.176068	-0.0195876	0.0108144	-0.0453555	-0.00970482	-0.0198458	-0.0705545	0.0346625	0.0949391

Wald Test

m = b7 + b8 + b9 = 0.9293.

Var[m] = R × Var[b] × R′ = [0 0 0 0 0 0 1 1 1]



$$\begin{pmatrix}0\\0\\0\\0\\0\\0\\1\\1\\1\end{pmatrix}$$

$$\sum_{i=1}^{9}\sum_{j=1}^{9} R_i R_j Cov(b_i, b_j) \quad = \quad 0.10107$$

$m' \ [Var(m)]^{-1} \ m \quad = \quad 8.5446$ (See slide 31)

The critical chi squared with 1 degree of freedom is 3.84, so the hypothesis is rejected.

Wald Statistic for 2 Restrictions

Joint hypothesis:
b(LY) = 1
b(LPG) = -1

Variable	Coefficient	Standard Error
Constant	-6.95326***	1.29811
LY	1.35721***	.14562
LPG	-.50579***	.06200
LPNC	-.01654	.19957
LPUC	-.12354*	.06568
LPPT	.11571	.07859
LPN	1.10125***	.26840
LPD	.92018***	.27018
LPS	-1.09213***	.30812

	1	2	3	
1	1.6851	-0.189024	-0.0256198	
2	-0.189024	0.0212045	0.00290895	
3	-0.0256198	0.00290895	0.00384368	-0
4	-0.218091	0.0243971	-0.000682307	
5	-0.0240267	0.00269963	-0.000413822	

```
Matrix ; R = [0,1,0,0,0,0,0,0,0 /
              0,0,1,0,0,0,0,0,0]$
Matrix ; q = [1/-1]$
Matrix ; list ; m = R*b - q $
Matrix m        has  2 rows and  1 columns.
              1
       +-------------+
      1|     .35721
      2|     .49421
       +-------------+
Matrix ; list ; vm = R*varb*R' $
Matrix VM       has  2 rows and  2 columns.
              1             2
       +-------------+-------------+
      1|     .02120        .00291
      2|     .00291        .00384
       +-------------+-------------+
Matrix ; list ; w = m'<vm>m $
Matrix W        has  1 rows and  1 columns.
              1
       +-------------+
      1|   63.55962
       +-------------+
```

Example: Panel Data on Spanish Dairy Farms

N = 247 farms, T = 6 years (1993-1998)

	Units	Mean	Std. Dev.	Minimum	Maximum
Output Milk	**Milk production (liters)**	131,107	92,584	14,410	727,281
Input Cows	**# of milking cows**	22.12	11.27	4.5	82.3
Input Labor	**# man-equivalent units**	1.67	0.55	1.0	4.0
Input Land	**Hectares of land devoted to pasture and crops.**	12.99	6.17	2.0	45.1
Input Feed	**Total amount of feedstuffs fed to dairy cows (Kg)**	57,941	47,981	3,924.14	376,732

Application

- y = log output
- **x** = Cobb douglas production: **x** = $1, x_1, x_2, x_3, x_4$
  = constant and logs of 4 inputs (5 terms)
- **z** = Translog terms, x_1^2, x_2^2, etc. and all cross products,
  x_1x_2, x_1x_3, x_1x_4, x_2x_3, etc. (10 terms)
- **w** = (**x**,**z**) (all 15 terms)
- Null hypothesis is Cobb Douglas, alternative is translog = Cobb-Douglas plus second order terms.

Translog Regression Model

```
--------------------------------------------------------------------------------
Ordinary      least squares regression ............
LHS=YIT       Mean                      =      11.57749
              Standard deviation        =        .64344
----------    No. of observations       =          1482  DegFreedom   Mean square
Regression    Sum of Squares            =       584.868          14     41.77630
Residual      Sum of Squares            =       28.2836        1467       .01928
Total         Sum of Squares            =       613.152        1481       .41401
----------    Standard error of e       =        .13885  Root MSE         .13815
Fit           R-squared                 =        .95387  R-bar squared    .95343
Model test    F[ 14,   1467]            =    2166.83407  Prob F > F*      .00000
--------+-----------------------------------------------------------------------
```

YIT	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	11.5689***	.00727	1591.61	.0000	11.5546	11.5831
X1	.60693***	.02186	27.76	.0000	.56408	.64978
X2	.01352	.01169	1.16	.2476	-.00940	.03643
X3	.02385*	.01356	1.76	.0786	-.00273	.05042
X4	.45379***	.01199	37.84	.0000	.43029	.47730
X11	.47329***	.14310	3.31	.0009	.19282	.75376
X22	-.08046	.04930	-1.63	.1026	-.17708	.01615
X33	-.04840	.09251	-.52	.6008	-.22972	.13292
X44	.17969***	.04556	3.94	.0001	.09040	.26897
X12	-.08380	.06167	-1.36	.1742	-.20467	.03707
X13	.18430**	.07248	2.54	.0110	.04225	.32635
X14	-.28574***	.07560	-3.78	.0002	-.43391	-.13757
X23	-.00816	.04326	-.19	.8505	-.09295	.07664
X24	.05222*	.03096	1.69	.0916	-.00846	.11290
X34	-.05821	.04041	-1.44	.1497	-.13741	.02099

β_x (X1–X4)

$H_0: \beta_z = 0$ (X11–X34)

Wald Tests

- **r(b,q)= close to zero?**
- **Wald distance function:**
- $\mathbf{r}(\mathbf{b},\mathbf{q})'\{\text{Var}[\mathbf{r}(\mathbf{b},\mathbf{q})]\}^{-1}\mathbf{r}(\mathbf{b},\mathbf{q}) \rightarrow \chi^2[J]$
- Use the delta method to estimate Var[**r**(**b**,**q**)]
  - $\text{Est.Asy.Var}[\mathbf{b}]=s^2(\mathbf{X}'\mathbf{X})^{-1}$
  - $\text{Est.Asy.Var}[\mathbf{r}(\mathbf{b},\mathbf{q})]= \mathbf{R}(\mathbf{b},\mathbf{q})\{s^2(\mathbf{X}'\mathbf{X})^{-1}\}\mathbf{R}'(\mathbf{b},\mathbf{q})$
- The standard F test is a Wald test; $JF = \chi^2[J]$.

$$\text{Wald}=(\mathbf{b}_z - \mathbf{0})'\{\text{Var}[\mathbf{b}_z - \mathbf{0}]\}^{-1}(\mathbf{b}_z - \mathbf{0}) = 42.122$$

```
Coefficient
-------------
  11.5689***
    .60693***
    .01352
    .02385*
    .45379***
    .47329***
  -.08046
  -.04840
    .17969***
  -.08380
    .18430**
  -.28574***
  -.00816
    .05222*
  -.05821
```

Close to 0?

W=J*F

```
|-> name;quadtrms=x11,x22,x33,x44,x12,x13,x14,x23,x2
|-> regr;lhs=yit;rhs=one,x1,x2,x3,x4,x11,x22,x33,x44
    ;test:quadtrms$

----------------------------------------------------------
Ordinary      least squares regression ............
LHS=YIT       Mean                    =      11.57749
              Standard deviation      =        .64344
----------    No. of observations     =          1482
Regression    Sum of Squares          =       584.868
Residual      Sum of Squares          =       28.2836
Total         Sum of Squares          =       613.152
----------    Standard error of e     =        .13885
Fit           R-squared               =        .95387
Model test    F[ 14,  1467]           =    2166.83407
Diagnostic    Log likelihood          =     830.65291
              Restricted (b=0)        =   -1448.90034
              Chi squared [ 14]       =    4559.12249
Wald Test:    Chi-squared [ 10]       =        42.122
F Test:       F ratio[10, 1467]       =         4.212
----------------------------------------------------------
```

	1
1	11.5689
2	0.606931
3	0.0135194
4	0.0238475
5	0.453793
6	0.473292
7	-0.080464
8	-0.0484005
9	0.179687
10	-0.0837999
11	0.184303
12	-0.28574
13	-0.00815564
14	0.0522221
15	-0.0582137

	1	2	3	4	5	6	7	8
1	5.28332e-005	6.96257e-006	1.04789e-005	-1.94441e-005	-9.66005e-006	-0.000323157	-0.000137561	-0.000428489
2	6.96257e-006	0.000478015	-0.000136173	-6.55597e-005	-0.000215724	0.000129724	8.43757e-005	-0.000134389
3	1.04789e-005	-0.000136173	0.000136698	7.10173e-006	2.8566e-005	-0.000176969	-5.66131e-005	-8.05683e-005
4	-1.94441e-005	-6.55597e-005	7.10173e-006	0.000183839	-7.47345e-006	0.000130347	-2.61419e-006	0.000358916
5	-9.66005e-006	-0.000215724	2.8566e-005	-7.47345e-006	0.000143842	0.000174599	6.52814e-006	4.78466e-005
6	-0.000323157	0.000129724	-0.000176969	0.000130347	0.000174599	0.0204774	0.00112112	0.000849242
7	-0.000137561	8.43757e-005	-5.66131e-005	-2.61419e-006	6.52814e-006	0.00112112	0.00243	0.000415515
8	-0.000428489	-0.000134389	-8.05683e-005	0.000358916	4.78466e-005	0.000849242	0.000415515	0.00855839
9	-0.000111405	-0.000211085	2.29923e-006	5.22235e-005	0.000176518	0.00432806	-0.00014612	0.000347859
10	9.94147e-005	-0.000202012	6.98727e-005	5.52519e-006	3.07544e-005	-0.00525622	-0.00181973	-0.000236149
11	4.33792e-005	0.000148986	-5.99719e-006	-5.55825e-005	-7.48148e-005	-0.00228512	-0.000110673	-0.00118828
12	0.000142739	0.000190158	1.27587e-005	-8.44398e-005	-0.000211884	-0.00959475	0.000117885	-0.000295217
13	3.10224e-005	5.367e-006	-1.81797e-006	-6.08703e-005	5.81279e-006	0.00033393	0.000176107	-0.000378138
14	-7.35643e-006	1.21626e-005	2.19949e-005	1.36967e-005	4.85042e-006	0.00258622	8.50701e-005	4.69664e-005
15	2.50147e-005	-7.76451e-005	1.46691e-005	6.63415e-006	3.92685e-005	0.0010118	-4.33862e-005	-0.000431011

	9	10	11	12	13	14	15
1	-0.000111405	9.94147e-005	4.33792e-005	0.000142739	3.10224e-005	-7.35643e-006	2.50147e-005
2	-0.000211085	-0.000202012	0.000148986	0.000190158	5.367e-006	1.21626e-005	-7.76451e-005
3	2.29923e-006	6.98727e-005	-5.99719e-006	1.27587e-005	-1.81797e-006	2.19949e-005	1.46691e-005
4	5.22235e-005	5.52519e-006	-5.55825e-005	-8.44398e-005	-6.08703e-005	1.36967e-005	6.63415e-006
5	0.000176518	3.07544e-005	-7.48148e-005	-0.000211884	5.81279e-006	4.85042e-006	3.92685e-005
6	0.00432806	-0.00525622	-0.00228512	-0.00959475	0.00033393	0.00258622	0.0010118
7	-0.00014612	-0.00181973	-0.000110673	0.000117885	0.000176107	8.50701e-005	-4.33862e-005
8	0.000347859	-0.000236149	-0.00118828	-0.000295217	-0.000378138	4.69664e-005	-0.000431011
9	0.00207531	-0.000489744	0.000508287	-0.0030555	-0.00017862	0.000461235	-0.000360587
10	-0.000489744	0.00380299	0.000242724	0.00164627	-0.000516602	-0.00132111	2.32476e-005
11	0.000508287	0.000242724	0.00525273	-1.40947e-005	-0.00148866	0.000206776	-0.00230405
12	-0.0030555	0.00164627	-1.40947e-005	0.00571481	0.000318155	-0.00130629	-0.000119702
13	-0.00017862	-0.000516602	-0.00148866	0.000318155	0.00187157	-0.000215743	0.000297747
14	0.000461235	-0.00132111	0.000206776	-0.00130629	-0.000215743	0.000958453	-2.92597e-005
15	-0.000360587	2.32476e-005	-0.00230405	-0.000119702	0.000297747	-2.92597e-005	0.00163286

Score or LM Test: General

- Maximum Likelihood (ML) Estimation, Adapted to LS.
- A hypothesis test
  - H_0: Restrictions on parameters are true
  - H_1: Restrictions on parameters are not true
- Basis for the test: $\mathbf{b}_0$ = parameter estimate under H_0 (i.e., restricted), $\mathbf{b}_1$ = unrestricted
- Derivative results: For the likelihood function under H_1,
  - $(\partial \log L_1/\partial \boldsymbol{\beta} \mid \boldsymbol{\beta}=\mathbf{b}_1) = \mathbf{0}$ (derivatives = 0 exactly, by definition)
  - $(\partial \log L_1/\partial \boldsymbol{\beta} \mid \boldsymbol{\beta}=\mathbf{b}_0) \neq \mathbf{0}$. Is it close? If so, the restrictions look reasonable

Restricted regression and derivatives for the LM Test

```
----------------------------------------------------------------------------
Restricted   least squares regression ............
LHS=YIT      Mean                   =     11.57749
             Standard deviation     =       .64344
----------   No. of observations    =         1482  DegFreedom   Mean square
Regression   Sum of Squares         =      584.056           4     146.01403
Residual     Sum of Squares         =      29.0957        1477        .01970
Total        Sum of Squares         =      613.152        1481        .41401
----------   Standard error of e    =       .14035  Root MSE          .14012
Fit          R-squared              =       .95255  R-bar squared     .95242
Model test   F[  4,  1477]          =   7412.18529  Prob F > F*       .00000
Restrictions F[ 10,  1467]          =      4.21224  Prob F > F*       .00001
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
     YIT|  Coefficient     Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|    11.5775***    .00365  3175.52  .0000     11.5703    11.5846
      X1|     .59518***    .01958    30.39  .0000      .55679     .63356
      X2|     .02305**     .01122     2.05  .0400      .00105     .04505
      X3|     .02319*      .01303     1.78  .0751    -.00235     .04873
      X4|     .45176***    .01078    41.89  .0000      .43062     .47290
     X11|      0.0      .....(Fixed Parameter).....
     X22|      0.0      .....(Fixed Parameter).....
     X33|      0.0      .....(Fixed Parameter).....
     X44|      0.0      .....(Fixed Parameter).....
     X12|      0.0      .....(Fixed Parameter).....
     X13|      0.0      .....(Fixed Parameter).....
     X14|      0.0      .....(Fixed Parameter).....
     X23|      0.0      .....(Fixed Parameter).....
     X24|      0.0      .....(Fixed Parameter).....
     X34|      0.0      .....(Fixed Parameter).....
--------+-------------------------------------------------------------------
```

Derivatives are

$$\mathbf{g} = \begin{bmatrix} \mathbf{X}'\mathbf{e}/s^2 \\ \mathbf{Z}'\mathbf{e}/s^2 \end{bmatrix}$$

```
 +-------------
 1|  0.0000000
 2|  0.0000000
 3|  0.0000000
 4|  0.0000000
 5|  0.0000000
 6|  110.38257
 7|  -54.32128
 8|   25.48239
 9|  226.05471
10|   37.93753
11|  177.16378
12|  258.29646
13|   87.61102
14|   42.35517
15|  205.07842
```

Are the residuals from regression of y on X alone uncorrelated with Z?

Computing the LM Statistic

Testing $\beta_z = 0$ in $y=X\beta_x+Z\beta_z+\boldsymbol{\varepsilon}$

Statistic computed from regression of y on X alone

1. **Compute Restricted Regression (y on X alone) and compute residuals, e_0**

2. **Regress e_0 on (X,Z). LM = NR^2 in this regression. (Regress e_0 on the RHS of the unrestricted regression.**

Application of the Score Test

Linear Model: $\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} = \mathbf{W}\theta + \boldsymbol{\varepsilon}$

- Test H_0: $\boldsymbol{\delta} = \mathbf{0}$
- Restricted estimator is [**b**',**0**']'

NAMELIST ; X = a list… ; Z = a list … ; W = X,Z $

REGRESS ; Lhs = y ; Rhs = X ; Res = e $

CALC ; List ; LM = N * Rsq(W,e) $

Regression Specification Tests

```
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=YIT     Mean                 =        11.57749
            Standard deviation   =          .64344
----------  No. of observations  =            1482  DegFreedom    Mean square
Regression  Sum of Squares       =         584.868          14       41.77630
Residual    Sum of Squares       =         28.2836        1467         .01928
Total       Sum of Squares       =         613.152        1481         .41401
----------  Standard error of e  =          .13885  Root MSE          .13815
Fit         R-squared            =          .95387  R-bar squared     .95343
Model test  F[ 14,  1467]        =      2166.83407  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
Constant|   11.5689***      .00727  1591.61  .0000     11.5546   11.5831
      X1|    .60693***      .02186    27.76  .0000      .56408    .64978
      X2|    .01352         .01169     1.16  .2476    -.00940    .03643
      X3|    .02385*        .01356     1.76  .0786    -.00273    .05042
      X4|    .45379***      .01199    37.84  .0000      .43029    .47730
     X11|    .47329***      .14310     3.31  .0009      .19282    .75376
     X22|   -.08046         .04930    -1.63  .1026    -.17708    .01615
     X33|   -.04840         .09251     -.52  .6008    -.22972    .13292
     X44|    .17969***      .04556     3.94  .0001      .09040    .26897
     X12|   -.08380         .06167    -1.36  .1742    -.20467    .03707
     X13|    .18430**       .07248     2.54  .0110      .04225    .32635
     X14|   -.28574***      .07560    -3.78  .0002    -.43391   -.13757
     X23|   -.00816         .04326     -.19  .8505    -.09295    .07664
     X24|    .05222*        .03096     1.69  .0916    -.00846    .11290
     X34|   -.05821         .04041    -1.44  .1497    -.13741    .02099
--------+--------------------------------------------------------------------
```

```
------------------------------------------------------------------------------
Restricted  least squares regression ............
LHS=YIT     Mean                 =        11.57749
            Standard deviation   =          .64344
----------  No. of observations  =            1482  DegFreedom    Mean square
Regression  Sum of Squares       =         584.056           4      146.01403
Residual    Sum of Squares       =         29.0957        1477         .01970
Total       Sum of Squares       =         613.152        1481         .41401
----------  Standard error of e  =          .14035  Root MSE          .14012
Fit         R-squared            =          .95255  R-bar squared     .95242
Model test  F[  4,  1477]        =      7412.18529  Prob F > F*       .00000
Restrictions F[ 10,  1467]       =         4.21224  Prob F > F*       .00001
--------+---------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
     YIT|  Coefficient       Error        z    |z|>Z*        Interval
--------+---------------------------------------------------------------------
Constant|   11.5775***      .00365  3175.52  .0000     11.5703   11.5846
      X1|    .59518***      .01958    30.39  .0000      .55679    .63356
      X2|    .02305**       .01122     2.05  .0400      .00105    .04505
      X3|    .02319*        .01303     1.78  .0751    -.00235    .04873
      X4|    .45176***      .01078    41.89  .0000      .43062    .47290
     X11|       0.0       .....(Fixed Parameter).....
     X22|       0.0       .....(Fixed Parameter).....
     X33|       0.0       .....(Fixed Parameter).....
     X44|       0.0       .....(Fixed Parameter).....
     X12|       0.0       .....(Fixed Parameter).....
     X13|       0.0       .....(Fixed Parameter).....
     X14|       0.0       .....(Fixed Parameter).....
     X23|       0.0       .....(Fixed Parameter).....
     X24|       0.0       .....(Fixed Parameter).....
     X34|       0.0       .....(Fixed Parameter).....
--------+---------------------------------------------------------------------
```

LM		=	41.365
Wald Test:	Chi-squared [10]	=	42.122
F Test:	F ratio[10, 1467]	=	4.212

The Spanish dairy data are a 6 period panel. We robustify our test by using the cluster corrected covariance matrix. We were misled by the conventional covariance matrix! Use WALD with a robust covariance matrix for the test.

```
----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=YIT      Mean                 =        11.57749
             Standard deviation   =          .64344
----------   No. of observations  =            1482  DegFreedom   Mean square
Regression   Sum of Squares       =         584.868          14      41.77630
Residual     Sum of Squares       =         28.2836        1467        .01928
Total        Sum of Squares       =         613.152        1481        .41401
----------   Standard error of e  =          .13885  Root MSE          .13815
Fit          R-squared            =          .95387  R-bar squared     .95343
Model test   F[ 14,  1467]        =      2166.83407  Prob F > F*       .00000
```

YIT	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	11.5689***	.00727	1591.61	.0000	11.5546	11.5831
X1	.60693***	.02186	27.76	.0000	.56408	.64978
X2	.01352	.01169	1.16	.2476	-.00940	.03643
X3	.02385*	.01356	1.76	.0786	-.00273	.05042
X4	.45379***	.01199	37.84	.0000	.43029	.47730
X11	.47329***	.14310	3.31	.0009	.19282	.75376
X22	-.08046	.04930	-1.63	.1026	-.17708	.01615
X33	-.04840	.09251	-.52	.6008	-.22972	.13292
X44	.17969***	.04556	3.94	.0001	.09040	.26897
X12	-.08380	.06167	-1.36	.1742	-.20467	.03707
X13	.18430**	.07248	2.54	.0110	.04225	.32635
X14	-.28574***	.07560	-3.78	.0002	-.43391	-.13757
X23	-.00816	.04326	-.19	.8505	-.09295	.07664
X24	.05222*	.03096	1.69	.0916	-.00846	.11290
X34	-.05821	.04041	-1.44	.1497	-.13741	.02099

Wald test based on conventional standard errors: Chi-squared [10] = 42.122, P = 0.00001

```
----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=YIT      Mean                 =        11.57749
             Standard deviation   =          .64344
----------   No. of observations  =            1482  DegFreedom   Mean square
Regression   Sum of Squares       =         584.868          14      41.77630
Residual     Sum of Squares       =         28.2836        1467        .01928
Total        Sum of Squares       =         613.152        1481        .41401
----------   Standard error of e  =          .13885  Root MSE          .13815
Fit          R-squared            =          .95387  R-bar squared     .95343
Model test   F[ 14,  1467]        =      2166.83407  Prob F > F*       .00000
Wald Test:   Chi-squared [ 10]    =          10.365  Prob C2 > C2* =   .40903
F Test:      F ratio[10, 1467]    =           1.037  Prob F  > F*  =   .40974
```

YIT	Coefficient	Clustered Std.Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	11.5689***	.01432	807.62	.0000	11.5408	11.5969
X1	.60693***	.03718	16.33	.0000	.53407	.67979
X2	.01352	.01969	.69	.4924	-.02508	.05211
X3	.02385	.02228	1.07	.2844	-.01982	.06751
X4	.45379***	.02148	21.13	.0000	.41169	.49589
X11	.47329	.29243	1.62	.1056	-.09985	1.04644
X22	-.08046	.08137	-.99	.3227	-.23994	.07901
X33	-.04840	.16114	-.30	.7639	-.36423	.26743
X44	.17969*	.09343	1.92	.0544	-.00343	.36280
X12	-.08380	.11935	-.70	.4826	-.31772	.15012
X13	.18430	.11787	1.56	.1179	-.04672	.41533
X14	-.28574*	.15386	-1.86	.0633	-.58730	.01582
X23	-.00816	.06074	-.13	.8932	-.12721	.11090
X24	.05222	.06327	.83	.4091	-.07178	.17623
X34	-.05821	.06472	-.90	.3684	-.18506	.06863

Wald statistic based on robust covariance matrix = 10.365. P = 0.409!!

Structural Change Test

Application: Wald Tests

Year,	G ,	Pg,	Y ,	Pnc ,	Puc ,	Ppt ,	Pd ,	Pn ,	Ps $
1960	129.7	.925	6036	1.045	.836	.810	.444	.331	.302
1961	131.3	.914	6113	1.045	.869	.846	.448	.335	.307
1962	137.1	.919	6271	1.041	.948	.874	.457	.338	.314
1963	141.6	.918	6378	1.035	.960	.885	.463	.343	.320
1964	148.8	.914	6727	1.032	1.001	.901	.470	.347	.325
1965	155.9	.949	7027	1.009	.994	.919	.471	.353	.332
1966	164.9	.970	7280	.991	.970	.952	.475	.366	.342
1967	171.0	1.000	7513	1.000	1.000	1.000	.483	.375	.353
1968	183.4	1.014	7728	1.028	1.028	1.046	.501	.390	.368
1969	195.8	1.047	7891	1.044	1.031	1.127	.514	.409	.386
1970	207.4	1.056	8134	1.076	1.043	1.285	.527	.427	.407
1971	218.3	1.063	8322	1.120	1.102	1.377	.547	.442	.431
1972	226.8	1.076	8562	1.110	1.105	1.434	.555	.458	.451
1973	237.9	1.181	9042	1.111	1.176	1.448	.566	.497	.474
1974	225.8	1.599	8867	1.175	1.226	1.480	.604	.572	.513
1975	232.4	1.708	8944	1.276	1.464	1.586	.659	.615	.556
1976	241.7	1.779	9175	1.357	1.679	1.742	.695	.638	.598
1977	249.2	1.882	9381	1.429	1.828	1.824	.727	.671	.648
1978	261.3	1.963	9735	1.538	1.865	1.878	.769	.719	.698
1979	248.9	2.656	9829	1.660	2.010	2.003	.821	.800	.756
1980	226.8	3.691	9722	1.793	2.081	2.516	.892	.894	.839
1981	225.6	4.109	9769	1.902	2.569	3.120	.957	.969	.926
1982	228.8	3.894	9725	1.976	2.964	3.460	1.000	1.000	1.000
1983	239.6	3.764	9930	2.026	3.297	3.626	1.041	1.021	1.062
1984	244.7	3.707	10421	2.085	3.757	3.852	1.038	1.050	1.117
1985	245.8	3.738	10563	2.152	3.797	4.028	1.045	1.075	1.173
1986	269.4	2.921	10780	2.240	3.632	4.264	1.053	1.069	1.224

Regression Model

Based on the gasoline data: The regression equation is

$$g = \beta_1 + \beta_2 y + \beta_3 pg + \beta_4 pnc + \beta_5 puc + \beta_6 ppt + \beta_7 pd + \beta_8 pn + \beta_9 ps + \beta_{10} t + \varepsilon$$

All variables are logs of the raw variables, so that coefficients are elasticities. The new variable, t, is a time trend, 0,1,…,26, so that β_{10} is the autonomous yearly proportional growth in G.

Structural Change

Time series regression,

$$\begin{aligned}\text{LogG} = \;& \beta_1 + \beta_2\text{logY} + \beta_3\text{logPG} \\ & + \beta_4\text{logPNC} + \beta_5\text{logPUC} + \beta_6\text{logPPT} \\ & + \beta_7\text{logPN} + \beta_8\text{logPD} + \beta_9\text{logPS} + \varepsilon\end{aligned}$$

A significant event occurs in October 1973. We will be interested to know if the model 1960 to 1973 is the same as from 1974 to 1995.

Data Setup

```
Create;
    G=log(G);
    Pg=log(PG);
    y=log(y);
    pnc=log(pnc);
    puc=log(puc);
    ppt=log(ppt);
    pd=log(pd);
    pn=log(pn);
    ps=log(ps);
    t=year-1960$
Namelist;X=one,y,pg,pnc,puc,ppt,pd,pn,ps,t$
Regress;lhs=g;rhs=X$
```

Least Squares Results

```
----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=G       Mean                  =      5.30862
            Standard deviation    =       .23135
----------  No. of observations   =           27  DegFreedom   Mean square
Regression  Sum of Squares        =      1.38783           9        .15420
Residual    Sum of Squares        =   .377694E-02         17        .00022
Residual    Sum of Squares        =   .412580E-04 (Using years 1960 - 1973)
Residual    Sum of Squares        =   .134382E-03 (Using years 1974 - 1986)
Total       Sum of Squares        =      1.39160          26        .05352
----------  Standard error of e   =       .01491  Root MSE          .01183
Fit         R-squared             =       .99729  R-bar squared     .99585
Model test  F[  9,    17]         =    694.06710  Prob F > F*       .00000
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
       G|  Coefficient     Error       t    |t|>T*          Interval
--------+-------------------------------------------------------------------
Constant|   -5.97984**    2.50176   -2.39  .0287   -11.25810    -.70158
       Y|    1.39438***    .27825    5.01  .0001      .80734   1.98143
      PG|    -.58144***    .06111   -9.51  .0000     -.71038    -.45250
     PNC|    -.29477       .25798   -1.14  .2690     -.83906     .24952
     PUC|    -.20154**     .07416   -2.72  .0146     -.35799    -.04508
     PPT|     .08051       .08707     .92  .3681     -.10319     .26420
      PD|    1.50607***    .29746    5.06  .0001      .87849   2.13364
      PN|     .99947***    .27033    3.70  .0018      .42913   1.56982
      PS|    -.81789*      .46198   -1.77  .0946    -1.79259     .15680
       T|    -.01251       .01264    -.99  .3359     -.03917     .01415
--------+-------------------------------------------------------------------
```

Covariance Matrix

Matrix - VARB

[10, 10] Cell:

	1	2	3	4	5	6	7	8	9	10
1	6.25882	-0.685584	0.0159666	-0.252511	-0.0992025	-0.121959	0.0767857	-0.210285	0.41674	0.0204969
2	-0.685584	0.0774203	-0.00186804	0.016999	0.00926198	0.0115885	0.000248256	0.0170407	-0.0291785	-0.00269606
3	0.0159666	-0.00186804	0.00373485	-0.00287659	-0.00105386	-0.00248163	-0.00607819	-0.0112643	0.0145609	0.000101201
4	-0.252511	0.016999	-0.00287659	0.0665533	0.00947888	0.0132049	-0.0406975	0.0418232	-0.0988791	0.00126402
5	-0.0992025	0.00926198	-0.00105386	0.00947888	0.00549911	0.00358764	-0.00915534	0.0135477	-0.0226984	-6.24541e-005
6	-0.121959	0.0115885	-0.00248163	0.0132049	0.00358764	0.00758068	-0.00443961	0.0175285	-0.0319759	-0.000146502
7	0.0767857	0.000248256	-0.00607819	-0.0406975	-0.00915534	-0.00443961	0.0884802	-0.0267256	0.0314479	-0.00121354
8	-0.210285	0.0170407	-0.0112643	0.0418232	0.0135477	0.0175285	-0.0267256	0.0730773	-0.103791	0.000193505
9	0.41674	-0.0291785	0.0145609	-0.0988791	-0.0226984	-0.0319759	0.0314479	-0.103791	0.213425	-0.00168906
10	0.0204969	-0.00269606	0.000101201	0.00126402	-6.24541e-005	-0.000146502	-0.00121354	0.000193505	-0.00168906	0.000159658

Chow Test

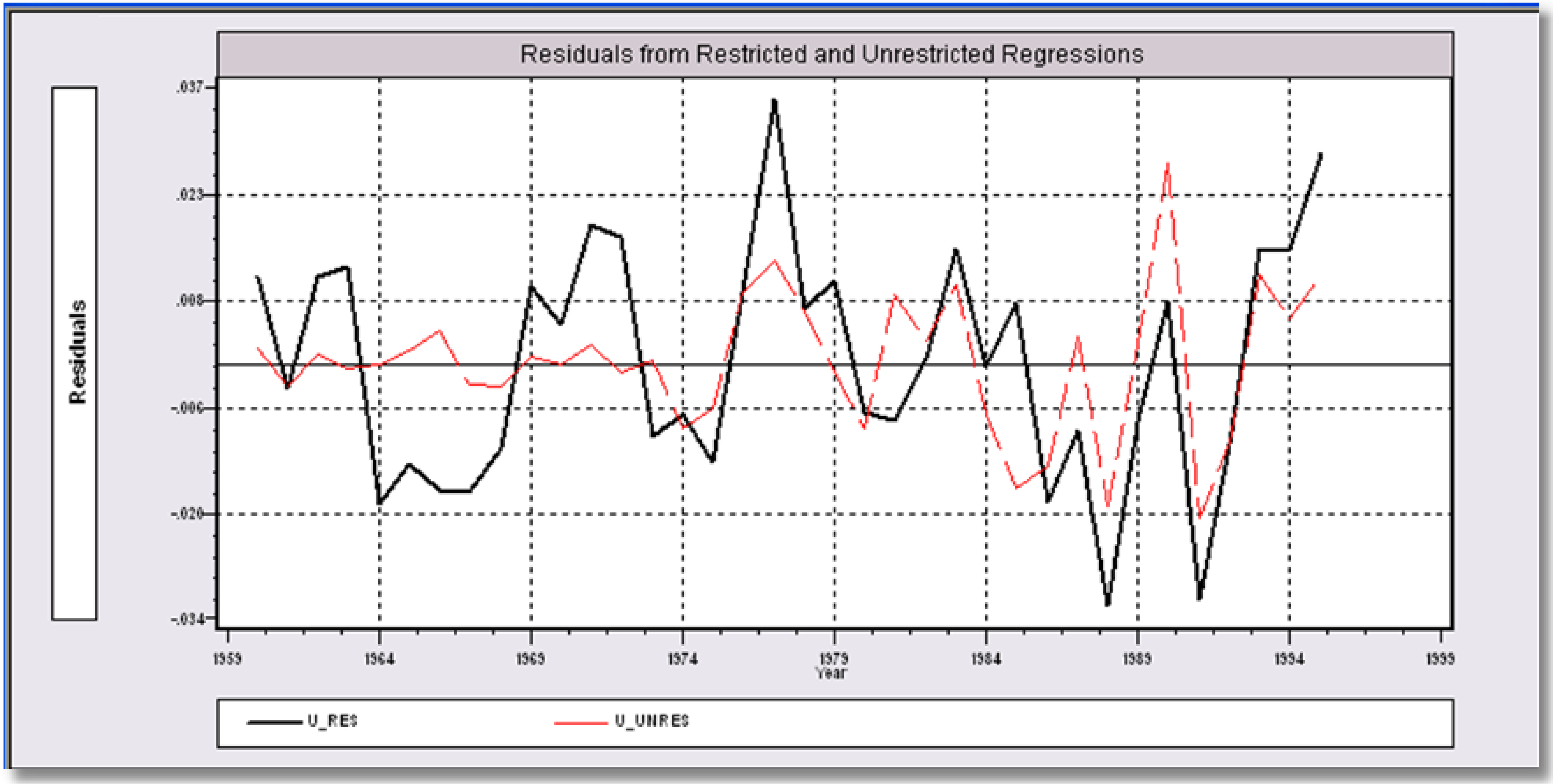
Structural change test – The CHOW test. Is the regression model the same in the two subperiods, before and after 1973. Use 3 regressions to find out.

Pooled sum of squares = 0.0037769400 = ss01
1960-1973 = 0.0000412580 = ss0
1974-1986 = 0.0001343820 = ss1

F[10, 27 – 20] = [(ss01 – (ss0 + ss1))/K] / [(ss0 + ss1)/(n0 + n1 – 2K]
= 14.353

The critical value is 3.637. The hypothesis of no structural change is rejected.

Residuals Show the Implication of the Restriction of Equal Coefficients. Loss of fit in the first period.



Algebra for the Chow Test

Unrestricted regression is

$$\begin{pmatrix} \mathbf{y}_{1960\text{-}1973} \\ \mathbf{y}_{1974\text{-}1995} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1960\text{-}1973} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{1974\text{-}1995} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1960\text{-}1973} \\ \varepsilon_{1974\text{-}1995} \end{pmatrix}$$

Restricted regression is

$$\begin{pmatrix} \mathbf{y}_{1960\text{-}1973} \\ \mathbf{y}_{1974\text{-}1995} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1960\text{-}1973} \\ \mathbf{X}_{1974\text{-}1995} \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_{1960\text{-}1973} \\ \varepsilon_{1974\text{-}1995} \end{pmatrix}$$

In the unrestricted model, $\mathbf{R} = [\mathbf{I},-\mathbf{I}]$, $\mathbf{q}=\mathbf{0}$.

$$\mathbf{Rb} - \mathbf{q} = \mathbf{b}_1 - \mathbf{b}_2;$$

$$\mathbf{R}[\text{Var}(\mathbf{b}_1, \mathbf{b}_2)]\mathbf{R}' = \text{Var}[\mathbf{b}_1] + \text{Var}[\mathbf{b}_2] \text{ (no covariance)}$$

Structural Change Test

Alternative Formulation

Unrestricted regression is

$$\begin{pmatrix} \mathbf{y}_{1960\text{-}1973} \\ \mathbf{y}_{1974\text{-}1995} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1960\text{-}1973} & \mathbf{0} \\ \mathbf{X}_{1974\text{-}1995} & \mathbf{X}_{1974\text{-}1995} \end{pmatrix} \begin{pmatrix} \beta \\ \delta \end{pmatrix} + \begin{pmatrix} \varepsilon_{1960\text{-}1973} \\ \varepsilon_{1974\text{-}1995} \end{pmatrix}$$

Restricted regression is

$$\begin{pmatrix} \mathbf{y}_{1960\text{-}1973} \\ \mathbf{y}_{1974\text{-}1995} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{1960\text{-}1973} & \mathbf{0} \\ \mathbf{X}_{1974\text{-}1995} & \mathbf{X}_{1974\text{-}1995} \end{pmatrix} \begin{pmatrix} \beta \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1960\text{-}1973} \\ \varepsilon_{1974\text{-}1995} \end{pmatrix}$$

In the unrestricted model, $\mathbf{R} = [\mathbf{0},\mathbf{I}]$, $\mathbf{q}=\mathbf{0}$.

$$\mathbf{Rb} - \mathbf{q} = \mathbf{d}; \quad \mathbf{R}[\text{Var}(\mathbf{b}_1, \mathbf{b}_2)]\mathbf{R}' = \text{Var}[\mathbf{d}]$$

$$\text{Wald} = \mathbf{d}' \left\{ \text{Var}[\mathbf{d}] \right\}^{-1} \mathbf{d}$$

Application – Health and Income

German Health Care Usage Data, 7,293 Individuals, Varying Numbers of Periods
Variables in the file are

Data downloaded from Journal of Applied Econometrics Archive. This is an unbalanced panel with 7,293 individuals. There are altogether 27,326 observations. The number of observations ranges from 1 to 7 per family. (Frequencies are: 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987). The dependent variable of interest is

DOCVIS = number of visits to the doctor in the observation period

HHNINC = household nominal monthly net income in German marks / 10000.
(4 observations with income=0 were dropped)
HHKIDS = children under age 16 in the household = 1; otherwise = 0
EDUC = years of schooling
AGE = age in years
MARRIED=marital status
WHITEC = 1 if has “white collar” job

Men

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=HHNINC   Mean                 =    .3590541    |
|              Standard deviation   =    .1735639    |
|              Number of observs.   =       14243    |
| Model size   Parameters           =           5    |
|              Degrees of freedom   =       14238    |
| Residuals    Sum of squares       =    379.8470    |
|              Standard error of e  =    .1633352    |
| Fit          R-squared            =    .1146423    |
|              Adjusted R-squared   =    .1143936    |
+----------------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	.04169***	.00894	4.662	.0000	
AGE	.00086***	.00013	6.654	.0000	42.6528
EDUC	.02044***	.00058	35.528	.0000	11.7287
MARRIED	.03825***	.00341	11.203	.0000	.76515
WHITEC	.03969***	.00305	13.002	.0000	.29994

Women

Ordinary	least squares regression		
LHS=HHNINC	Mean	=	.3444951
	Standard deviation	=	.1801790
	Number of observs.	=	13083
Model size	Parameters	=	5
	Degrees of freedom	=	13078
Residuals	Sum of squares	=	363.8789
	Standard error of e	=	.1668045
Fit	R-squared	=	.1432098
	Adjusted R-squared	=	.1429477

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	.01191	.01158	1.029	.3036	
AGE	.00026*	.00014	1.875	.0608	44.4760
EDUC	.01941***	.00072	26.803	.0000	10.8764
MARRIED	.12081***	.00343	35.227	.0000	.75151
WHITEC	.06445***	.00334	19.310	.0000	.29924

All

```
+----------------------------------------------------+
| Ordinary    least squares regression                |
| LHS=HHNINC   Mean                 =   .3520836      |
|              Standard deviation   =   .1769083      |
|              Number of observs.   =      27326      |
| Model size   Parameters           =          5      |
|              Degrees of freedom   =      27321      |
| Residuals    Sum of squares       =   752.4767      | All
| Residuals    Sum of squares       =   379.8470      | Men
| Residuals    Sum of squares       =   363.8789      | Women
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+--------------+----------------+--------+--------+----------+
|Constant|     .04186***      .00704         5.949   .0000            |
|AGE     |     .00030***    .919581D-04      3.209   .0013     43.5257|
|EDUC    |     .01967***      .00045        44.180   .0000     11.3206|
|MARRIED |     .07947***      .00239        33.192   .0000      .75862|
|WHITEC  |     .04819***      .00225        21.465   .0000      .29960|
+--------+------------------------------------------------------------+
```

F Statistic for Chow Test

```
--> Calc     ; k = col(x)
             ; List; dfd = (tm + tf - 2*k)
             ; Chowtest = ((sall - sm - sf)/k) /
                          ((sm+sf)/dfd)
             ; FCrit = Ftb(.95,k,dfd)  $

DFD       =  27316.000000
CHOWTEST  =     64.281630
FCRIT     =      2.214100
```

Use Dummy Variables (and Base the Test on a Robust Covariance Matrix)

```
sample   ; all$
namelist ; x=one,age,educ,married,whitec$
regress  ; lhs=hhninc;rhs=x$
namelist ; xf=female,female*age,female*educ,
              female*married,female*whitec$
regress  ; lhs=hhninc;rhs=x,xf ; test:xf=0$
```

```
----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HHNINC   Mean                 =         .35208
             Standard deviation   =         .17691
----------   No. of observations  =          27326  DegFreedom   Mean square
Regression   Sum of Squares       =        111.452          9     12.38354
Residual     Sum of Squares       =        743.726      27316       .02723
Total        Sum of Squares       =        855.178      27325       .03130
----------   Standard error of e  =         .16501  Root MSE          .16498
Fit          R-squared            =         .13033  R-bar squared     .13004
Model test   F[  9, 27316]        =      454.82993  Prob F > F*       .00000
Wald Test:   Chi-squared [  5]    =      321.40815  Prob C2 > C2* =   .00000
F Test:      F ratio[ 5,27316]    =       64.28163  Prob F > F*   =   .00000
--------+-------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  HHNINC|  Coefficient     Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------------
Constant|      .04169***     .00903    4.62  .0000      .02399     .05940
     AGE|      .00086***     .00013    6.59  .0000      .00060     .00111
    EDUC|      .02044***     .00058   35.17  .0000      .01930     .02158
 MARRIED|      .03825***     .00345   11.09  .0000      .03149     .04501
  WHITEC|      .03969***     .00308   12.87  .0000      .03365     .04574
  FEMALE|     -.02978**      .01459   -2.04  .0412     -.05837    -.00119
        |Interaction FEMALE*AGE
_ntrct01|     -.00060***     .00019   -3.21  .0013     -.00097    -.00023
        |Interaction FEMALE*EDUC
_ntrct02|     -.00102        .00092   -1.11  .2678     -.00283     .00079
        |Interaction FEMALE*MARRIED
_ntrct03|      .08256***     .00484   17.07  .0000      .07308     .09205
        |Interaction FEMALE*WHITEC
_ntrct04|      .02476***     .00452    5.48  .0000      .01591     .03362
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

Specification Test: Normality of ε

- Specification test for distribution
- Standard tests:
  - Kolmogorov-Smirnov: compare empirical cdf of X to normal with same mean and variance
  - Bowman-Shenton: Compare third and fourth moments of X to normal, 0 (no skewness) and $3\sigma^4$ (meso kurtosis)
- Bera-Jarque – adapted Bowman/Shenton to linear regression residuals

Testing for Normality

Normality Test for Random Variable e

$$s = \sqrt{\frac{\sum_{i=1}^{n}(e_i - \bar{e})^2}{N}}, \quad m_j = \frac{\sum_{i=1}^{n}(e_i - \bar{e})^j}{N},$$

$\bar{e} = 0$ for regression residuals

$$\text{Chi-squared}[2] = \frac{(m_3 / s^3)^2}{6} + \frac{[(m_4 / s^4) - 3]^2}{20}$$

The Stochastic Frontier Model

$$y_i = f(\mathbf{x}_i) TE_i e^{v_i}$$

$$\ln y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + v_i - u_i$$

$$= \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i.$$

$u_i > 0$, usually assumed to be $|N[0,\sigma]|$
v_i may take any value.

A symmetric distribution, such as the normal distribution, is usually assumed for v_i.

Closed Skew Normal Distribution

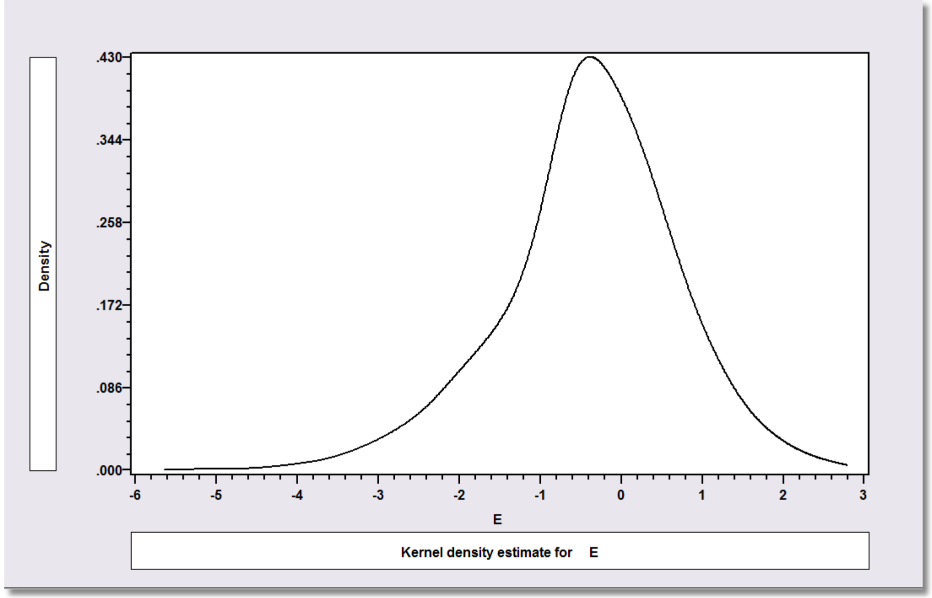
$$v \sim N[0,\sigma_v^2], \quad U \sim N[0,\sigma_u^2]$$

$$u = |U| \quad \text{(absolute value)}$$

$$\varepsilon = v - u$$

$$\text{Let } \sigma = \sqrt{\sigma_v^2 + \sigma_u^2} \qquad \lambda = \frac{\sigma_u}{\sigma_v}$$

$$f(\varepsilon) = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(\frac{-\lambda\varepsilon}{\sigma}\right)$$



$$\sigma_v = \sigma_u = 1$$

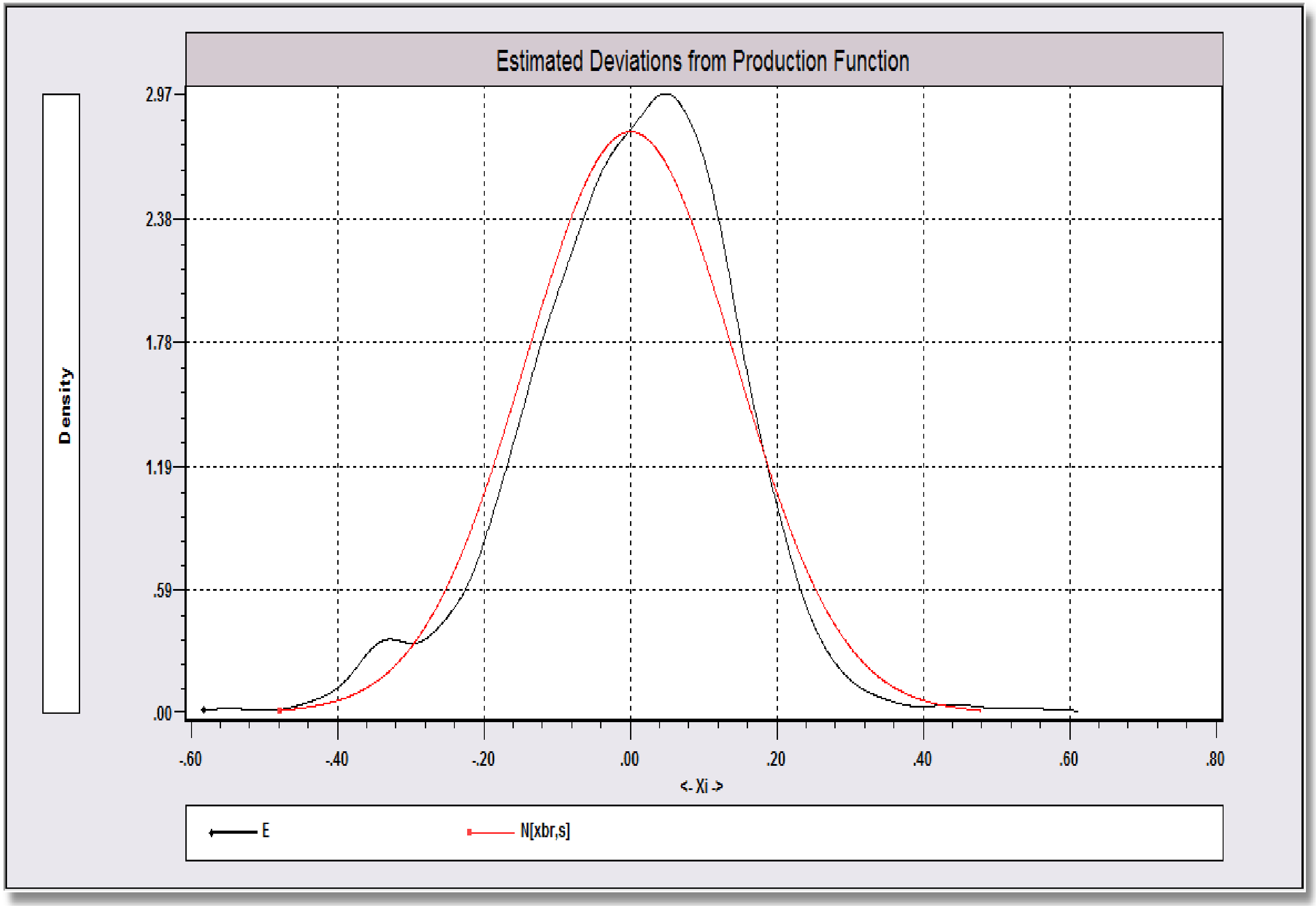
Application to Spanish Dairy Farms

N = 247 farms, T = 6 years (1993-1998)

Input	Units	Mean	Std. Dev.	Minimum	Maximum
Milk	Milk production (liters)	131,108	92,539	14,110	727,281
Cows	# of milking cows	2.12	11.27	4.5	82.3
Labor	# man-equivalent units	1.67	0.55	1.0	4.0
Land	Hectares of land devoted to pasture and crops.	12.99	6.17	2.0	45.1
Feed	Total amount of feedstuffs fed to dairy cows (tons)	57,941	47,981	3,924.14	376,732

Stochastic Frontier Model

```
-----------------------------------------------------------------------------
Ordinary       least squares regression ............
LHS=YIT        Mean                  =        11.57749
               Standard deviation    =          .64344
----------     No. of observations   =            1482  DegFreedom   Mean square
Regression     Sum of Squares        =         584.056           4    146.01403
Residual       Sum of Squares        =         29.0957        1477       .01970
Total          Sum of Squares        =         613.152        1481       .41401
----------     Standard error of e   =          .14035  Root MSE          .14012
Fit            R-squared             =          .95255  R-bar squared     .95242
Model test     F[  4,  1477]         =      7412.18529  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |                Standard               Prob.      95% Confidence
     YIT|  Coefficient     Error         z     |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    11.5775***     .00365   3175.52  .0000     11.5703    11.5846
      X1|     .59518***     .01958     30.39  .0000      .55679     .63356
      X2|     .02305**      .01122      2.05  .0400      .00105     .04505
      X3|     .02319*       .01303      1.78  .0751     -.00235     .04873
      X4|     .45176***     .01078     41.89  .0000      .43062     .47290
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```


Estimated Deviations from Production Function
Density
2.97
2.38
1.78
1.19
.59
.00
-.60
-.40
-.20
.00
.20
.40
.60
.80
<- Xi ->
E
N[xbr,s]

```
---------------------------------------
Kernel Density Estimator for EIT
Kernel Function      =          Normal
Observations         =            1482
Points plotted       =            1482
Bandwidth            =         .029289
Statistics for abscissa values----
Mean                 =         .000000
Standard Deviation =           .140164
Skewness             =        -.296061
Kurtosis-3 (excess)=           .802483
Chi2 normality test=          1.801944
Minimum              =        -.554019
Maximum              =         .583301
Results matrix       =          KERNEL
---------------------------------------
|-> dsta;rhs=eit;normal$

Descriptive Statistics for   1 variables
--------+--------------------------------------------------------------------
Variable|        Mean        Std.Dev.       Minimum       Maximum     Cases Missing
--------+--------------------------------------------------------------------
     EIT|          0.0        .140164      -.554019       .583301      1482       0
        | Skewness    -.30  Kurtosis       3.80    Chisq=    61.42    Prob= .0000
--------+--------------------------------------------------------------------
```

Appendix
Miscellaneous Results

Nonlinear Restrictions

I am interested in testing the hypothesis that certain ratios of elasticities are equal. In particular,

$\phi_1 = \beta_4/\beta_5 \; - \; \beta_7/\beta_8 = 0$

$\phi_2 = \beta_4/\beta_5 \; - \; \beta_9/\beta_8 = 0$

Setting Up the Wald Statistic

To do the Wald test, I first need to estimate the asymptotic covariance matrix for the sample estimates of ϕ_1 and ϕ_2. After estimating the regression by least squares, the estimates are

$$f_1 = b_4/b_5 - b_7/b_8$$
$$f_2 = b_4/b_5 - b_9/b_8.$$

Then, using the delta method, I will estimate the asymptotic variances of f_1 and f_2 and the asymptotic covariance of f_1 and f_2. For this, write $f_1 = f_1(\mathbf{b})$, that is a function of the entire 10×1 coefficient vector. Then, I compute the 1×10 derivative vectors, $\mathbf{d}_1 = \partial f_1(\mathbf{b})/\partial\mathbf{b}'$ and $\mathbf{d_2} = \partial f_2(\mathbf{b})/\partial\mathbf{b}'$ These vectors are

	1	2	3	4	5	6	7	8	9	10
$\mathbf{d}_1$ =	0,	0,	0,	$1/b_5$,	$-b_4/b_5^2$,	0,	$-1/b_8$,	b_7/b_8^2,	0,	0
$\mathbf{d}_2$ =	0,	0,	0,	$1/b_5$,	$-b_4/b_5^2$,	0,	0,	b_9/b_8^2,	$-1/b_8$,	0

Wald Statistics

Then, **D** = the 2×10 matrix with first row $\mathbf{d}_1$ and second row $\mathbf{d}_2$. The estimator of the asymptotic covariance matrix of $[f_1,f_2]'$ (a 2×1 column vector) is $\mathbf{V} = \mathbf{D} \times s^2 (\mathbf{X}'\mathbf{X})^{-1} \times \mathbf{D}'$. Finally, the Wald test of the hypothesis that $\phi = 0$ is carried out by using the chi-squared statistic $W = (\mathbf{f}-\mathbf{0})'\mathbf{V}^{-1}(\mathbf{f}-\mathbf{0})$. This is a chi-squared statistic with 2 degrees of freedom. The critical value from the chi-squared table is 5.99, so if my sample chi-squared statistic is greater than 5.99, I reject the hypothesis.

Wald Test

In the example below, to make this a little simpler, I computed the 10 variable regression, then extracted the 5×1 subvector of the coefficient vector **c** = (b_4,b_5,b_7,b_8,b_9) and its associated part of the 10×10 covariance matrix. Then, I manipulated this smaller set of values.

Application of the Wald Statistic

```
?  Extract subvector and submatrix for the test
matrix;list ; c  =b(4:9)]$
matrix;list ; vc=varb(4:9,4:9)
?  Compute derivatives
calc  ;list
; g11=1/c(2); g12=-c(1)*g11*g11; g13=-1/c(4) ; g14=c(3)*g13*g13 ;  g15=0
; g21=   g11 ; g22=g12                ; g23=0         ; g24=c(5)/c(4)^2      ; g25=-1/c(4)$
?  Move derivatives to matrix
matrix;list; dfdc=[g11,g12,g13,g14,g15 / g21,g22,g23,g24,g25]$
?  Compute functions, then move to matrix and compute Wald statistic
calc;list ; f1=c(1)/c(2) - c(3)/c(4)
             ; f2=c(1)/c(2) - c(5)/c(4) $
matrix ; list; f = [f1/f2]$
matrix ; list; vf=dfdc * vc * dfdc' $
matrix ; list ; wald = f' * <vf> * f$
```

Computations

```
Matrix C        is     5 rows by     1 columns.
                1
     1   -0.2948  -0.2015   1.506   0.9995  -0.8179
Matrix VC       is     5 rows by     5 columns.
                1           2           3           4           5
     1    0.6655E-01  0.9479E-02 -0.4070E-01  0.4182E-01 -0.9888E-01
     2    0.9479E-02  0.5499E-02 -0.9155E-02  0.1355E-01 -0.2270E-01
     3   -0.4070E-01 -0.9155E-02  0.8848E-01 -0.2673E-01  0.3145E-01
     4    0.4182E-01  0.1355E-01 -0.2673E-01  0.7308E-01 -0.1038
     5   -0.9888E-01 -0.2270E-01  0.3145E-01 -0.1038      0.2134
 G11 =  -4.96184        G12 =  7.25755      G13= -1.00054    G14      =   1.50770   G15      =  0.000000
 G21 =  -4.96184        G22 =  7.25755      G23 = 0          G24      = -0.818753  G25      =  -1.00054
DFDC=[G11,G12,G13,G14,G15/G21,G22,G23,G24,G25]
Matrix DFDC     is     2 rows by     5 columns.
                1           2           3           4           5
     1    -4.962       7.258      -1.001       1.508      0.0000
     2    -4.962       7.258      0.0000     -0.8188      -1.001
F1= -0.442126E-01
F2=  2.28098
F=[F1/F2]
VF=DFDC*VC*DFDC'
Matrix VF       is     2 rows by     2 columns.
                1           2
     1    0.9804      0.7846
     2    0.7846      0.8648
WALD   Matrix Result   is     1 rows by     1 columns.
                1
     1     22.65
```

Noninvariance of the Wald Test

I also did a second test (using the built-in procedure) to illustrate a problem with Wald tests. Note that the hypothesis can be written a bit differently. An equivalent way to write them

$$\gamma_1 = \beta_5\beta_7 - \beta_4\beta_8 = 0$$
$$\gamma_2 = \beta_4\beta_8 - \beta_5\beta_9 = 0$$

In a small sample, one can get a different answer depending on how they write the hypothesis.

```
+-----------------------------------------------+
¦ WALD procedure. Estimates and standard errors ¦
¦ Wald Statistic            =      10.68662     ¦   USING PRODUCTS
¦ Prob. from Chi-squared[ 2] =      0.00478     ¦
+-----------------------------------------------+
Variable   Coefficient   Standard Error  z=b/s.e.  P[¦Z¦=z]
----------------------------------------------------------
Fncn( 1)  -0.8905728E-02   0.20022         -0.044   0.96452
Fncn( 2)   0.4594581       0.18578          2.473   0.01339
```

Unlike likelihood ratio tests and Lagrange multiplier tests, the Wald test is not invariant to such transformations.

Nonnested Regression Models

- Davidson and MacKinnon: If model A is correct, then predictions from model B will not add to the fit of model A to the data.

- Vuong: If model A is correct, then the likelihood function will generally favor model A and not model B

Davidson and MacKinnon Strategy

- Obtain predictions from model A = AFit
- Obtain predictions from model B = Bfit
- If A is correct, in the combined model (A,Bfit), Bfit should not be significant.
- If B is correct, in the combined model (B,Afit), Afit should not be significant.
- (Unfortunately), all four combinations of significance and not are possible.

Application

Model A

$$\text{LogG}(t) = \beta_1 + \beta_2 \text{logY}(t) + \beta_3 \text{logPG}(t) + \beta_4 \text{logPNC}(t) + \beta_5 \text{logPUC}(t) + \beta_6 \text{logPPT}(t) + \beta_7 \text{logG}(t-1) + \varepsilon$$

Model B

$$\text{LogG}(t) = \alpha_1 + \alpha_2 \text{logY}(t) + \alpha_3 \text{logPG}(t) + \alpha_4 \text{logPNC}(t) + \alpha_5 \text{logPUC}(t) + \alpha_6 \text{logPPT}(t) + \alpha_7 \text{logY}(t-1) + w$$

B does not add to Model A

```
----------------------------------------------------------------------------
Ordinary      least squares regression ............
LHS=LG        Mean                 =        4.61092
              Standard deviation   =         .14303
----------    No. of observations  =             35  DegFreedom   Mean square
Regression    Sum of Squares       =        .681767           7        .09740
Residual      Sum of Squares       =     .138314E-01         27        .00051
Total         Sum of Squares       =        .695598          34        .02046
----------    Standard error of e  =         .02263  Root MSE              .01988
Fit           R-squared            =         .98012  R-bar squared         .97496
Model test    F[  7,    27]        =      190.12276  Prob F > F*           .00000
Model was estimated on Jul 24, 2012 at 07:04:54 PM
--------+-------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
      LG|  Coefficient     Error       t     |t|>T*          Interval
--------+-------------------------------------------------------------------
Constant|   -12.7218     9.72682    -1.31   .2019     -31.7860     6.3424
      LY|    2.15927     1.62750     1.33   .1957     -1.03057    5.34911
     LPG|    -.18067**    .08418    -2.15   .0410      -.34566    -.01567
    LPNC|    -.34024**    .13119    -2.59   .0152      -.59736    -.08312
    LPUC|     .13004      .08385     1.55   .1326      -.03430     .29439
    LPPT|    -.16018      .16024    -1.00   .3264      -.47424     .15388
    LAGG|     .74018***   .13752     5.38   .0000       .47064    1.00972
    BFIT|   -1.18185     1.10554    -1.07   .2945     -3.34867     .98498
--------+-------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

A Does Add to Model B

```
--------------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LG       Mean                      =       4.61092
             Standard deviation        =        .14303
----------   No. of observations       =            35   DegFreedom   Mean square
Regression   Sum of Squares            =       .681767            7        .09740
Residual     Sum of Squares            =   .138314E-01           27        .00051
Total        Sum of Squares            =       .695598           34        .02046
----------   Standard error of e       =        .02263   Root MSE          .01988
Fit          R-squared                 =        .98012   R-bar squared     .97496
Model test   F[  7,    27]             =     190.12276   Prob F > F*       .00000
Model was estimated on Jul 24, 2012 at 07:05:07 PM
--------+-----------------------------------------------------------------------
        |               Standard            Prob.      95% Confidence
      LG|  Coefficient    Error       t     |t|>T*        Interval
--------+-----------------------------------------------------------------------
Constant|     .82775    1.98775      .42   .6804    -3.06817   4.72367
      LY|     .25746     .38092      .68   .5049     -.48912   1.00404
     LPG|     .02190     .03341      .66   .5178     -.04359    .08739
    LPNC|     .04991     .12108      .41   .6835     -.18740    .28722
    LPUC|    -.05097     .07715     -.66   .5144     -.20219    .10025
    LPPT|     .02833     .06749      .42   .6780     -.10396    .16062
    LAGY|    -.39154     .36626    -1.07   .2945    -1.10939    .32631
    AFIT|    1.07855***  .20039     5.38   .0000      .68579   1.47131
--------+-----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
--------------------------------------------------------------------------------
```

Voung

- Log density for an observation is
  $L_i = -.5*[\log(2\pi) + \log(s^2) + e_i^2/s^2]$
- Compute Li(A) and Li(B) for each observation
- Compute $D_i = L_i(A) – L_i(B)$
- Test hypothesis that mean of D_i equals zero using familiar “z” test.
- Test statistic > +2 favors model A, < -2 favors model B, in between is inconclusive.

```
namelist ; x = one,ly,lpg,lpnc,lpuc,lppt$
regress  ; lhs    = lg ; rhs = x $
create   ; lagy   = ly[-1] $
create   ; lagg   = lg[-1] $
namelist ; modelA = x,lagg $
namelist ; modelB = x,lagy $
sample   ; 2 - 36 $
regress  ; lhs = lg ; rhs = modela ; keep = afit
         ; res = ea $
calc     ; sa2=sumsqdev/n$
regress  ; lhs = lg ; rhs = modelb ; keep = bfit
         ; res = eb $
calc     ; sb2=sumsqdev/n$
create ; la = -.5*(log(2*pi)+log(sa2)+ea*ea/sa2) $
create ; lb = -.5*(log(2*pi)+log(sb2)+eb*eb/sb2) $
create ; v = la-lb $
calc   ; list ; vuong = sqr(n)*xbr(v)/sdv(v)$
```

```
|-> create ; la = -.5*(log(2*pi)+log(sa2)+ea*ea/sa2) $
|-> create ; lb = -.5*(log(2*pi)+log(sb2)+eb*eb/sb2) $
|-> create ; v = la-lb $
|-> calc   ; list ; vuong = sqr(n)*xbr(v)/sdv(v)$
[CALC] VUONG   =        2.6745922
```

Oaxaca Decomposition

Two groups, two regression models: (Two time periods, men vs. women, two countries, etc.)

$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$ and $\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2$

Consider mean values,

$$y_1^* = E[y_1|\text{mean } x_1] = x_1^{*\prime}\boldsymbol{\beta}_1$$
$$y_2^* = E[y_2|\text{mean } x_2] = x_2^{*\prime}\boldsymbol{\beta}_2$$

Now, explain why y_1^* is different from y_2^*. (I.e., departing from y_2, why is y_1 different?) (Could reverse the roles of 1 and 2.)

$$y_1^* - y_2^* = \mathbf{x}_1^{*\prime}\boldsymbol{\beta}_1 - \mathbf{x}_2^{*\prime}\boldsymbol{\beta}_2$$
$$= \mathbf{x}_1^{*\prime}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) + (\mathbf{x}_1^* - \mathbf{x}_2^*)'\boldsymbol{\beta}_2$$

(change in model) (change in conditions)

The Oaxaca Decomposition

Two groups (e.g., men=1, women=2)

Regression predictions:

$\hat{y}_1 = \bar{\mathbf{x}}_1'\mathbf{b}_1,\ \ \hat{y}_2 = \bar{\mathbf{x}}_2'\mathbf{b}_2$ (e.g., wage equations)

Explain $\hat{y}_1 - \hat{y}_2$.

$$\hat{y}_1 - \hat{y}_2 = \bar{\mathbf{x}}_1'(\mathbf{b}_1 - \mathbf{b}_2) \quad + \quad (\bar{\mathbf{x}}_1' - \bar{\mathbf{x}}_2')\mathbf{b}_2$$

discrimination + qualifications

$$\text{Var}[\bar{\mathbf{x}}_1'(\mathbf{b}_1 - \mathbf{b}_2)] = \bar{\mathbf{x}}_1'\{\sigma_1^2(\mathbf{X}_1'\mathbf{X}_1)^{-1} + \sigma_2^2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\}\bar{\mathbf{x}}_1$$

$$\text{Wald: } W = (\bar{\mathbf{x}}_1'(\mathbf{b}_1 - \mathbf{b}_2))^2 / [\bar{\mathbf{x}}_1'\{\sigma_1^2(\mathbf{X}_1'\mathbf{X}_1)^{-1} + \sigma_2^2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\}\bar{\mathbf{x}}_1]$$

What is the hypothesis?

Application - Income

German Health Care Usage Data, 7,293 Individuals, Varying Numbers of Periods

Variables in the file are

Data downloaded from Journal of Applied Econometrics Archive. This is an unbalanced panel with 7,293 individuals. They can be used for regression, count models, binary choice, ordered choice, and bivariate binary choice. This is a large data set. There are altogether 27,326 observations. The number of observations ranges from 1 to 7. (Frequencies are: 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987).

HHNINC = household nominal monthly net income in German marks / 10000.
(4 observations with income=0 were dropped)

HHKIDS = children under age 16 in the household = 1; otherwise = 0

EDUC = years of schooling

AGE = age in years

MARRIED = 1 if married, 0 if not

FEMALE = 1 if female, 0 if male

Regression: Female=0 (Men)

```
----------------------------------------------------------------------
Subsample analyzed for this command is FEMALE   =       0
----------------------------------------------------------------------
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HHNINC   Mean                   =       .35905
             Standard deviation     =       .17356
             No. of observations    =        14243  Degrees of freedom
Regression   Sum of Squares         =      45.0724            4
Residual     Sum of Squares         =      383.960        14238
Total        Sum of Squares         =      429.032        14242
             Standard error of e    =       .16422
Fit          R-squared              =       .10506  R-bar squared =   .10480
Model test   F[  4, 14238]          =    417.84403  Prob F > F*   =   .00000
Diagnostic   Log likelihood         =   5523.47748  Akaike I.C.   = -3.61278
             Restricted (b=0)       =   4733.03242  Bayes  I.C.   = -3.61013
             Chi squared [  4]      =   1580.89012  Prob C2 > C2* =   .00000
Model was estimated on Aug 25, 2011 at 10:00:15 AM
--------+--------------------------------------------------------------------
        |                Standard             Prob.      95% Confidence
  HHNINC|  Coefficient     Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|     .04679***    .00931      5.03  .0000      .02855     .06504
     AGE|     .00056***    .00014      3.94  .0001      .00028     .00084
    EDUC|     .02190***    .00057     38.61  .0000      .02079     .02301
 MARRIED|     .04800***    .00387     12.39  .0000      .04040     .05559
  HHKIDS|    -.01255***    .00327     -3.84  .0001     -.01896    -.00614
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Regression Female=1 (Women)

```
-----------------------------------------------------------------
Subsample analyzed for this command is FEMALE    =       1
-----------------------------------------------------------------
-----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HHNINC   Mean                    =          .34450
             Standard deviation      =          .18018
             No. of observations     =           13083  Degrees of freedom
Regression   Sum of Squares          =         51.3061          4
Residual     Sum of Squares          =         373.394      13078
Total        Sum of Squares          =         424.700      13082
             Standard error of e     =          .16897
Fit          R-squared               =          .12081  R-bar squared =   .12054
Model test   F[  4, 13078]           =       449.24453  Prob F > F*   =   .00000
Diagnostic   Log likelihood          =      4700.44519  Akaike I.C.   = -3.55567
             Restricted (b=0)        =      3858.23327  Bayes  I.C.   = -3.55281
             Chi squared [  4]       =      1684.42384  Prob C2 > C2* =   .00000
Model was estimated on Aug 25, 2011 at 10:00:15 AM
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  HHNINC|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    .06737***    .01241     5.43  .0000      .04304     .09170
     AGE|   -.00070***    .00015    -4.52  .0000     -.00100    -.00039
    EDUC|    .02087***    .00073    28.62  .0000      .01944     .02230
 MARRIED|    .11790***    .00357    33.06  .0000      .11091     .12490
  HHKIDS|   -.01941***    .00354    -5.49  .0000     -.02634    -.01247
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Pooled Regression

```
--------------------------------------------------------------------
Full pooled sample is used for this iteration.
--------------------------------------------------------------------
+-------------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=HHNINC   Mean                 =          .35208
             Standard deviation   =          .17691
             No. of observations  =           27326  Degrees of freedom
Regression   Sum of Squares       =         91.9565            4
Residual     Sum of Squares       =         763.221        27321
Total        Sum of Squares       =         855.178        27325
             Standard error of e  =          .16714
Fit          R-squared            =          .10753  R-bar squared =    .10740
Model test   F[  4, 27321]        =       822.94044  Prob F > F*   =    .00000
Diagnostic   Log likelihood       =     10112.92716  Akaike I.C.   = -3.57768
             Restricted (b=0)     =      8558.60603  Bayes  I.C.   = -3.57618
             Chi squared [   4]   =      3108.64227  Prob C2 > C2* =    .00000
Model was estimated on Aug 25, 2011 at 10:00:15 AM
--------+----------------------------------------------------------------------
        |                 Standard            Prob.       95% Confidence
  HHNINC|  Coefficient       Error       z    |z|>Z*         Interval
--------+----------------------------------------------------------------------
Constant|     .06617***     .00741     8.93   .0000      .05165     .08070
     AGE|    -.00028***     .00010    -2.75   .0060     -.00048    -.00008
    EDUC|     .02122***     .00044    47.99   .0000      .02035     .02208
 MARRIED|     .08691***     .00260    33.40   .0000      .08181     .09201
  HHKIDS|    -.01987***     .00238    -8.35   .0000     -.02453    -.01520
--------+----------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-------------------------------------------------------------------------------
```

Application

```
namelist ; X = one,age,educ,married,hhkids$
? Get results for females
include  ; new ; female=1$                       Subsample females
regr     ; lhs=hhninc;rhs=x$                     Regression
matrix   ; bf=b ; vf = varb ; xbarf = mean(x) $  Coefficients, variance, mean X
calc     ; meanincf = bf'xbarf $                 Mean prediction for females
? Get results for males
include  ; new ; female=0$                       Subsample males
regr     ; lhs=hhninc;rhs=x$                     Regression
matrix   ; bm=b ; vm = varb ; xbarm = mean(x) $  Coefficients, etc.
calc     ; meanincm = bm'xbarm $                 Mean prediction for males
? Examine difference in mean predicted income
calc     ; list
         ; meanincm ; meanincf                   Display means
         ; diff = xbarm'bm - xbarf'bf $          Difference in means
matrix   ; vdiff = xbarm'[vm]xbarm + xbarf'[vf]xbarf $ Variance of difference
calc     ; list ; diffwald = diff^2 / vdiff $         Wald test of difference = 0
? "Discrimination" component of difference
matrix   ; db = bm-bf  ; discrim = xbarm'db      Difference in coeffs., discrimination
         ; vdb = vm+vf ; vdiscrim = xbarm'[vdb]xbarm $ Variance of discrimination
calc     ; list ; discrim ; dwald = discrim^2 / vdiscrim $ Walt test that D = 0.
? "Difference due difference in X"
matrix   ; dx = xbarm - xbarf $                  Difference in characteristics
matrix   ; qual = dx'bf ; vqual = dx'[vf]dx $    Contribution to total difference
calc     ; list ; qual ; qualwald = qual^2/vqual $  Wald test.
```

Results

```
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
MEANINCM =         .359054
MEANINCF =         .344495
DIFF     =         .014559
DIFFWALD =       52.006502

DISCRIM  =        -.005693
DWALD    =        7.268757

QUAL     =         .020252
QUALWALD =     1071.053640
```

Decompositions

```
----------------------------------------------------------------------
Decomposition of Changes in Average Functions
Model Used in Computations = Linear Regression Function
----------------------------------------------------------------------
                Sample is FEMALE  = 0       FEMALE  = 1     Sample
Estimates Based on                 (0)                (1)      Size
FEMALE    = 0 (a)          .359054 (a,0)      .341020 (a,1)     14243
FEMALE    = 1 (b)          .364747 (b,0)      .344495 (b,1)     13083
Pooled    =** (*)          .361349 (*,0)      .341996 (*,1)     27326
----------------------------------------------------------------------
Wald Test of Difference in the Two Coefficient Vectors
Chi squared[  5] =  211.9299         , P Value =  .0000
----------------------------------------------------------------------
Total Change in Function    (a,0) - (b,1) =      .014559 ( 100.00%)
----------------------------------------------------------------------
Oaxaca     | Due to data is  (a,0) - (a,1) =      .018034 ( 123.87%)
Blinder    | Due to beta is  (a,1) - (b,1) =     -.003475 ( -23.87%)
----------------------------------------------------------------------
Daymont    | Due to data is  (b,0) - (b,1) =      .020252 ( 139.10%)
Andrisani  | Due to beta is  (a,0) - (b,0) =     -.005693 ( -39.10%)
----------------------------------------------------------------------
Daymont    | Due to data is  (b,0) - (b,1) =      .020252 ( 139.10%)
Andrisani  | Due to beta is  (a,1) - (b,1) =     -.003475 ( -23.87%)
(3 Fold)   | Due to function (a,0) - (b,0) -
           |                 (a,1) - (b,1) =     -.002218 ( -15.24%)
----------------------------------------------------------------------
Ransom     | Due to data is  (*,0) - (*,1) =      .019353 ( 132.93%)
Oaxaca     | Due to beta is  (a,0) - (*,0) +     -.004794 ( -32.93%)
Neumark    |                 (*,1) - (b,1)
----------------------------------------------------------------------
```

Likelihood Ratio Test

- The normality assumption
- Does it work 'approximately?'
- For any regression model $y_i = h(\mathbf{x}_i, \beta) + \varepsilon_i$ where $\varepsilon_i \sim N[0,\sigma^2]$, (linear or nonlinear), at the linear (or nonlinear) least squares estimator, however computed, with or without restrictions,

$$\log L(\hat{\beta} \text{ and } \hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/N) = -(N/2)[1 + \log 2\pi + \log\hat{\sigma}^2]$$

This forms the basis for likelihood ratio tests.

$$2[\log L(\hat{\beta}_{unrestricted}) - \log L(\hat{\beta}_{restricted})]$$

$$= N \log \frac{\hat{\sigma}^2_{restricted}}{\hat{\sigma}^2_{unrestricted}} \xrightarrow{d} \chi^2[J]$$

Likelihood Ratio Test

```
-----------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=YIT     Mean                      =     11.57749
            Standard deviation        =       .64344
----------  No. of observations       =         1482
Regression  Sum of Squares            =      584.056
Residual    Sum of Squares            =      29.0957
Total       Sum of Squares            =      613.152
----------  Standard error of e        =       .14035
Fit         R-squared                 =       .95255
Model test  F[  4,  1477]             =  7412.18529
Diagnostic  Log likelihood            =    809.67609
            Restricted (b=0)          =  -1448.90834
            Chi squared [   4]        =   4517.16885
---------+-------------------------------------------------------------
         |               Standard                   Prob.
      YIT|  Coefficient    Error          z        |z|>Z*
---------+-------------------------------------------------------------
 Constant|    11.5775***   .00365   3175.52   .0000
       X1|     .59518***   .01958     30.39   .0000
       X2|     .02305**    .01122      2.05   .0400
       X3|     .02319*     .01303      1.78   .0751
       X4|     .45176***   .01078     41.89   .0000
```

```
-----------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=YIT     Mean                      =     11.57749
            Standard deviation        =       .64344
----------  No. of observations       =         1482
Regression  Sum of Squares            =      584.868
Residual    Sum of Squares            =      28.2836
Total       Sum of Squares            =      613.152
            Standard error of e       =       .13885
----------  R-squared                 =       .95387
Fit         F[ 14,  1467]             =   2166.83407
Model test  Log likelihood            =    830.65291
Diagnostic  Restricted (b=0)          =  -1448.90834
            Chi squared [ 14]         =   4559.12249
---------+-------------------------------------------------------------
         |               Standard                   Prob.
      YIT|  Coefficient    Error          z        |z|>Z*
---------+-------------------------------------------------------------
 Constant|    11.5689***   .00727   1591.61   .0000
       X1|     .60693***   .02186     27.76   .0000
       X2|     .01352      .01169      1.16   .2476
       X3|     .02385*     .01356      1.76   .0786
       X4|     .45379***   .01199     37.84   .0000
      X11|     .47329***   .14310      3.31   .0009
      X22|    -.08046      .04930     -1.63   .1026
      X33|    -.04840      .09251      -.52   .6008
      X44|     .17969***   .04556      3.94   .0001
      X12|    -.08380      .06167     -1.36   .1742
      X13|     .18430**    .07248      2.54   .0110
      X14|    -.28574***   .07560     -3.78   .0002
      X23|    -.00816      .04326      -.19   .8505
      X24|     .05222*     .03096      1.69   .0916
      X34|    -.05821      .04041     -1.44   .1497
```

LR = 2(830.653 – 809.676) = 41.954