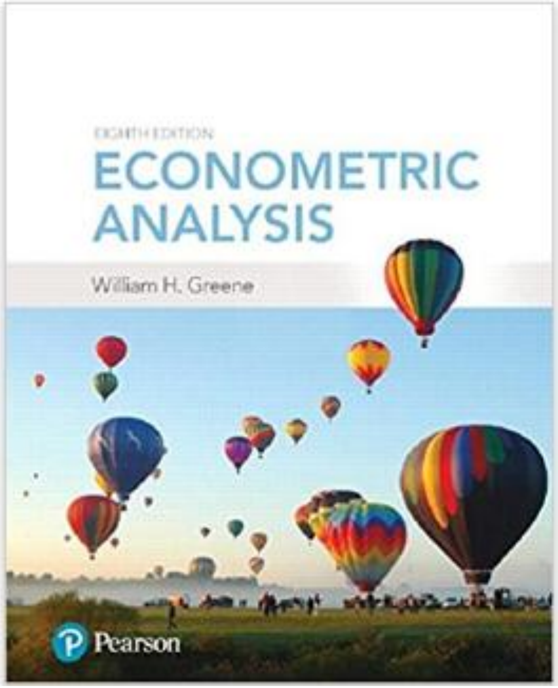
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



Econometrics I

Part 14 – Generalized Regression

Generalized Regression Model

Setting: The classical linear model assumes that $E[\varepsilon\varepsilon'] = Var[\varepsilon] = \sigma^2\mathbf{I}$. That is, observations are uncorrelated and all are drawn from a distribution with the same variance. The **generalized regression** (**GR**) model allows the variances to differ across observations and allows correlation across observations.

Generalized Regression Model

- The generalized regression model:

  $\mathbf{y} = \mathbf{X}\beta + \varepsilon$,

  $E[\varepsilon|\mathbf{X}] = \mathbf{0}$, $Var[\varepsilon|\mathbf{X}] = \sigma^2\mathbf{\Omega}$.

  Regressors are well behaved.

  Trace $\mathbf{\Omega}$ = n.

  This is a 'normalization.'

  Mimics $tr(\sigma^2 I) = n\sigma^2$. Needed since $(\sigma^2 c)\left(\frac{1}{c}\mathbf{\Omega}\right) = \sigma^2\mathbf{\Omega}$ for any c.

- Leading Cases
  - Simple heteroscedasticity
  - Autocorrelation
  - Panel data and heterogeneity more generally.
  - SUR Models for Production and Cost
  - VAR models in Macroeconomics and Finance

Implications of GR Assumptions

- The assumption that Var[$\boldsymbol{\varepsilon}$] = $\sigma^2\mathbf{I}$ is used to derive the result Var[**b**] = $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. If it is not true, then the use of $s^2(\mathbf{X}'\mathbf{X})^{-1}$ to estimate Var[**b**] is inappropriate.
- The assumption was also used to derive the t and F test statistics, so they must be revised as well.
- Least squares gives each observation a weight of 1/n. But, if the variances are not equal, then some observations are more informative than others.
- Least squares is based on simple sums, so the information that one observation might provide about another is never used.

Implications for Least Squares

- Still **unbiased**. (Proof did not rely on $\boldsymbol{\Omega}$)

- For **consistency**, we need the true variance of **b**,

$$\begin{aligned} \text{Var}[\mathbf{b}|\mathbf{X}] &= E[(\mathbf{b}-\boldsymbol{\beta})(\mathbf{b}-\boldsymbol{\beta})'|\mathbf{X}] \\ &= (\mathbf{X}'\mathbf{X})^{-1} E[\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}] (\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}. \end{aligned}$$

(Sandwich form of the covariance matrix.)

Divide all 4 terms by *n*. If the middle one converges to a finite matrix of constants, we have mean square consistency, so we need to examine

$$(1/n)\mathbf{X}'\boldsymbol{\Omega}\mathbf{X} = (1/n)\Sigma_i\Sigma_j \ \omega_{ij} \mathbf{x}_i \mathbf{x}_j'.$$

This will be another assumption of the model.

- **Asymptotic normality**? Easy for heteroscedasticity case, very difficult for autocorrelation case.

Robust Covariance Matrix

- **Robust estimation: Generality**
- How to estimate

  $\text{Var}[\mathbf{b}|\mathbf{X}] = (\mathbf{X'X})^{-1}\,\mathbf{X}'(\sigma^2\,\boldsymbol{\Omega})\mathbf{X}(\mathbf{X'X})^{-1}$ for the LS $\mathbf{b}$?
- The distinction between estimating

  $\sigma^2\boldsymbol{\Omega}$ an n×n matrix

  and estimating the K×K matrix

  $\sigma^2\,\mathbf{X}'\boldsymbol{\Omega}\mathbf{X} = \sigma^2\,\Sigma_i\Sigma_j\omega_{ij}\,\mathbf{x}_i\,\mathbf{x}_j'$
- **NOTE…… VVVIR**s for modern applied econometrics.
  - The White estimator
  - Newey-West estimator.

The White Estimator

$$\text{Est.Var}[\mathbf{b}] = (\mathbf{X'X})^{-1}\left[\sum_{i=1}^{n} e_i^2 \mathbf{x}_i \mathbf{x}_i'\right](\mathbf{X'X})^{-1}$$

(Heteroscedasticity robust covariance matrix.)

- **Meaning of "robust" in this context**
- **Robust standard errors**; (**b** is not "robust")
  - Robust to: Heteroscedasticty
  - Not robust to: (all considered later)
    - Correlation across observations
    - Individual unobserved heterogeneity
    - Incorrect model specification
- **Robust inference** means hypothesis tests and confidence intervals using robust covariance matrices

Inference Based on OLS

What about $s^2(\mathbf{X'X})^{-1}$? Depends on $\mathbf{X'\Omega X}$ - $\mathbf{X'X}$. If they are nearly the same, the OLS covariance matrix is OK. When will they be nearly the same? Relates to an interesting property of weighted averages. Suppose ω_i is randomly drawn from a distribution with $E[\omega_i] = 1$.

Then, $(1/n)\Sigma_i x_i^2 \rightarrow E[x^2]$ and $(1/n)\Sigma_i \omega_i x_i^2 \rightarrow E[x^2]$.

This is the crux of the discussion in your text.

Inference Based on OLS

VIR: For the heteroscedasticity to be substantive wrt estimation and inference by LS, the weights must be correlated with x and/or x^2. (Text, page 305.)

If the heteroscedasticity is substantive. Then, **b** is inefficient.

The White estimator. **ROBUST** estimation of the variance of **b**.

Implication for testing hypotheses. We will use Wald tests.

(**ROBUST TEST STATISTICS**)

Finding Heteroscedasticity

The central issue is whether $E[\varepsilon^2] = \sigma^2\omega_i$ is related to the xs or their squares in the model.

Suggests an obvious strategy. Use residuals to estimate disturbances and look for relationships between e_i^2 and x_i and/or x_i^2. For example, regressions of squared residuals on xs and their squares.

Procedures

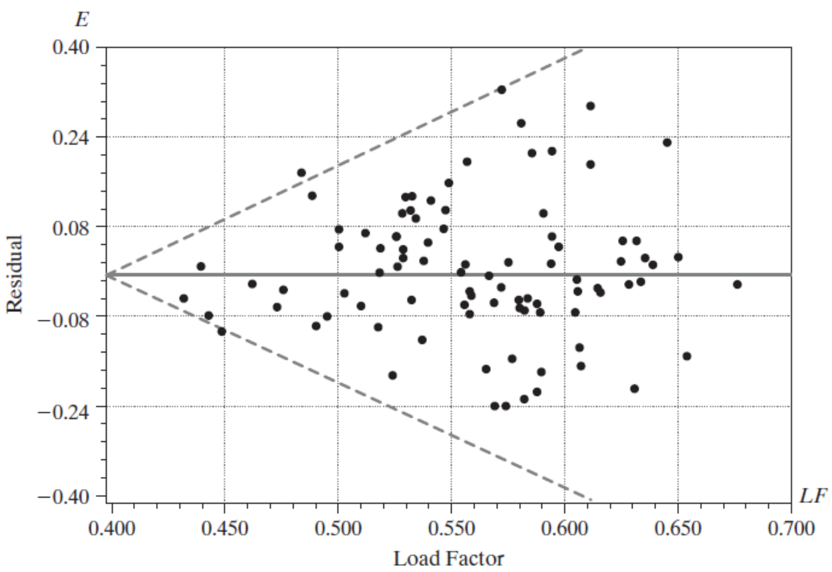
White's general test: nR^2 in the regression of e_i^2 on all unique xs, squares, and cross products. Chi-squared[P]

Breusch and Pagan's Lagrange multiplier test. Regress

$\{[e_i^2 / (\mathbf{e'e}/n)] - 1\}$ on **Z** (may be **X**).

Chi-squared. Is nR^2 with degrees of freedom rank of **Z**.

FIGURE 9.2 Plot of Residuals against Load Factor.



A Heteroscedasticity Robust Covariance Matrix

```
Ordinary      least squares regression ............
LHS=LWAGE     Mean                     =      6.67635
              Standard deviation       =       .46151
              Number of observs.       =         4165
Model size    Parameters               =           11
              Degrees of freedom       =         4154
Residuals     Sum of squares           =      515.950
              Standard error of e      =       .35243
Fit           R-squared                =       .41826
              Adjusted R-squared       =       .41686
Model test    F[ 10,  4154] (prob) =    298.7(.0000)
White heteroscedasticity robust covariance matrix.
Br./Pagan LM Chi-sq [ 10]  (prob) = 105.71 (.0000)
```

LWAGE	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	5.24547***	.07567	69.32	.0000	5.09715	5.39379
ED	.05654***	.00273	20.71	.0000	.05119	.06189
EXP	.04045***	.00219	18.46	.0000	.03616	.04474
EXP*EXP	-.00068***	.4893D-04	-13.92	.0000	-.00078	-.00059
WKS	.00449***	.00116	3.85	.0001	.00220	.00677
OCC	-.14053***	.01508	-9.32	.0000	-.17009	-.11098
SOUTH	-.07210***	.01274	-5.66	.0000	-.09707	-.04714
SMSA	.13901***	.01200	11.59	.0000	.11550	.16252
MS	.06736***	.02099	3.21	.0013	.02622	.10849
FEM	-.38922***	.02395	-16.25	.0000	-.43617	-.34227
UNION	.09015***	.01246	7.23	.0000	.06572	.11458

nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
***, **, * ==> Significance at 1%, 5%, 10% level.

Uncorrected

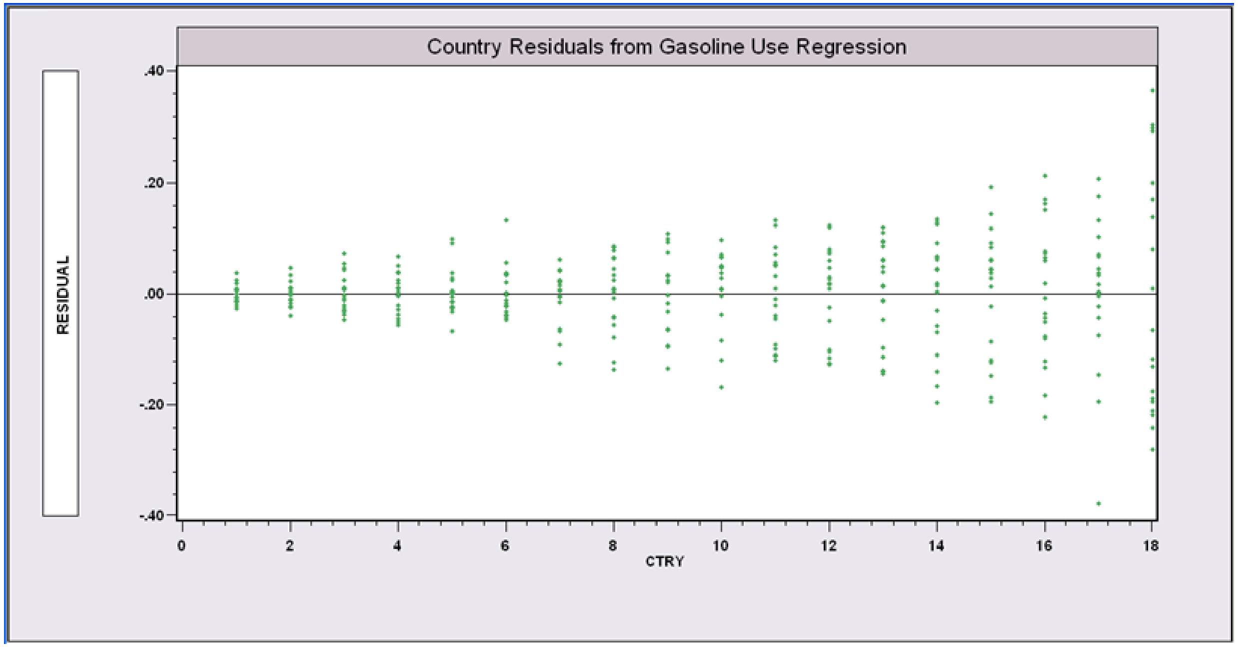
Standard Error	z
.07170	73.15
.00261	21.64
.00217	18.61
.4783D-04	-14.24
.00109	4.12
.01472	-9.54
.01249	-5.77
.01207	11.51
.02063	3.26
.02518	-15.46
.01289	6.99

Note the conflict: Test favors heteroscedasticity.

Robust VC matrix is essentially the same.

Groupwise Heteroscedasticity
Gasoline Demand Model

Countries are ordered by the standard deviation of their 19 residuals.



Regression of log of per capita gasoline use on log of per capita income, gasoline price and number of cars per capita for 18 OECD countries for 19 years. The standard deviation varies by country. The efficient estimator is "weighted least squares."

White Estimator

(Not really appropriate for groupwise heteroscedasticity)

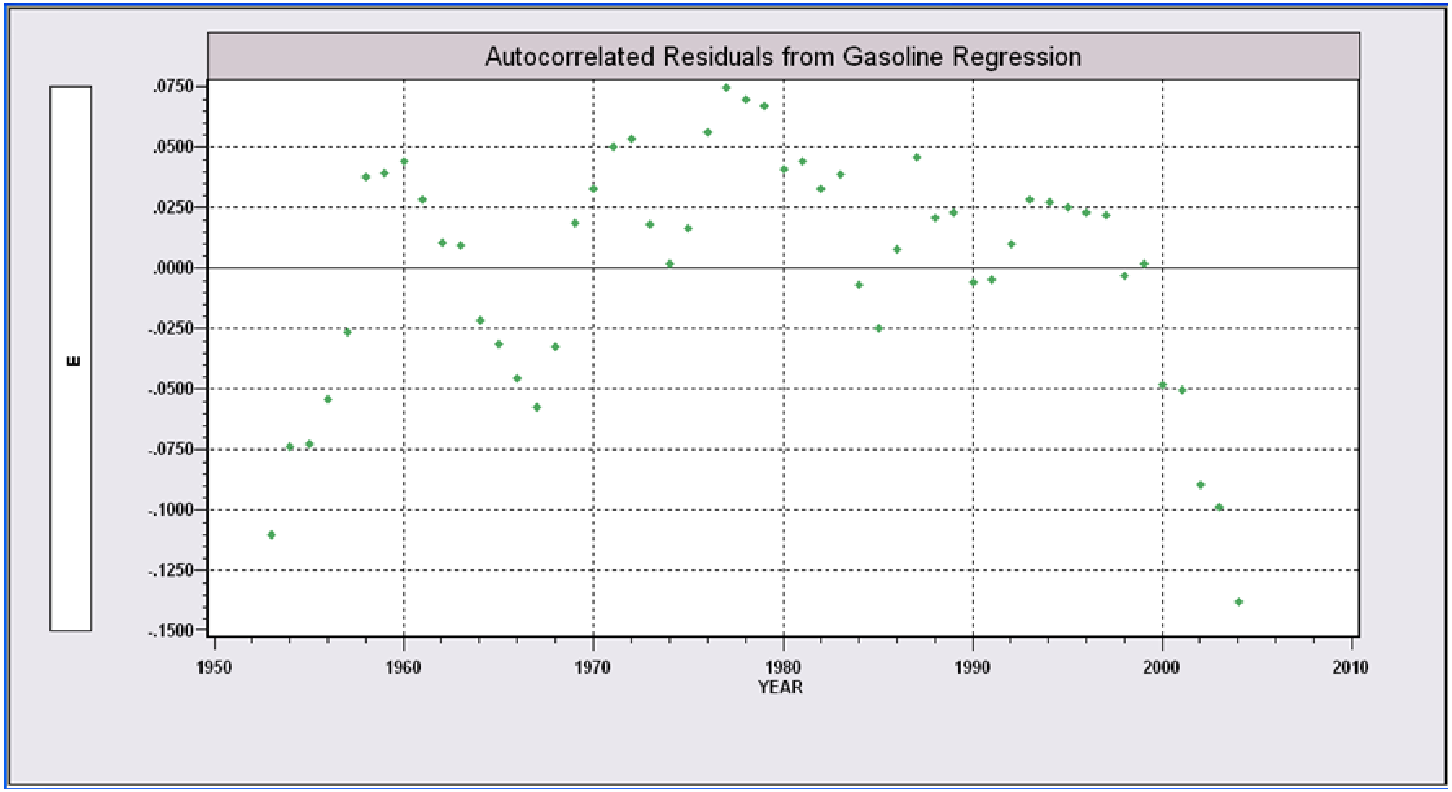
Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	2.39132562	.11693429	20.450	.0000	
LINCOMEP	.88996166	.03580581	24.855	.0000	-6.13942544
LRPMG	-.89179791	.03031474	-29.418	.0000	-.52310321
LCARPCAP	-.76337275	.01860830	-41.023	.0000	-9.04180473

White heteroscedasticity robust covariance matrix

Constant	2.39132562	.11794828	20.274	.0000	
LINCOMEP	.88996166	.04429158	20.093	.0000	-6.13942544
LRPMG	-.89179791	.03890922	-22.920	.0000	-.52310321
LCARPCAP	-.76337275	.02152888	-35.458	.0000	-9.04180473

Autocorrelated Residuals

$\log G = \beta_1 + \beta_2 \log Pg + \beta_3 \log Y + \beta_4 \log Pnc + \beta_5 \log Puc + \varepsilon$



Newey-West Estimator

Heteroscedasticity Component - Diagonal Elements

$$\mathbf{S}_0 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \mathbf{x}_i \mathbf{x}_i'$$

Autocorrelation Component - Off Diagonal Elements

$$\mathbf{S}_1 = \frac{1}{n}\sum_{l=1}^{L}\sum_{t=l+1}^{n} w_l e_t e_{t-l}(\mathbf{x}_t \mathbf{x}_{t-l}' + \mathbf{x}_{t-l}\mathbf{x}_t')$$

$$w_l = 1 - \frac{l}{L+1} = \text{"Bartlett weight"}$$

$$\text{Est.Var}[\mathbf{b}] = \frac{1}{n}\left(\frac{\mathbf{X'X}}{n}\right)^{-1}[\mathbf{S}_0 + \mathbf{S}_1]\left(\frac{\mathbf{X'X}}{n}\right)^{-1}$$

Newey-West Estimate

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	-21.2111***	.75322	-28.160	.0000	
LP	-.02121	.04377	-.485	.6303	3.72930
LY	1.09587***	.07771	14.102	.0000	9.67215
LPNC	-.37361**	.15707	-2.379	.0215	4.38037
LPUC	.02003	.10330	.194	.8471	4.10545

Robust VC Newey-West, Periods = 10

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	-21.2111***	1.33095	-15.937	.0000	
LP	-.02121	.06119	-.347	.7305	3.72930
LY	1.09587***	.14234	7.699	.0000	9.67215
LPNC	-.37361**	.16615	-2.249	.0293	4.38037
LPUC	.02003	.14176	.141	.8882	4.10545

Generalized Least Squares Approach

Aitken theorem. The **Generalized Least Squares** estimator, GLS. Find **P** such that

$$\mathbf{Py} = \mathbf{PX}\beta + \mathbf{P}\varepsilon$$

$$\mathbf{y}^* = \mathbf{X}^*\beta + \varepsilon^*.$$

$$E[\varepsilon^*\varepsilon^{*\prime}|\mathbf{X}^*] = \sigma^2\mathbf{I}$$

Use ordinary least squares in the transformed model. Satisfies the Gauss – Markov theorem.

$$\mathbf{b}^* = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^*$$

Generalized Least Squares – Finding **P**

A transformation of the model:

$$\mathbf{P} = \boldsymbol{\Omega}^{-1/2}.\ \mathbf{P}'\mathbf{P} = \boldsymbol{\Omega}^{-1}$$

$$\mathbf{Py} = \mathbf{PX}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon} \text{ or}$$

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*.$$

We need a noninteger power of a matrix: $\boldsymbol{\Omega}^{-1/2}$.

(Digression) Powers of a Matrix

- (See slides 7:41-42)
- Characteristic Roots and Vectors
  - $\boldsymbol{\Omega} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$
  - $\mathbf{C}$ = Orthogonal matrix of characteristic vectors.
  - $\boldsymbol{\Lambda}$ = Diagonal matrix of characteristic roots

Matrix - R

[6, 6] Cell: 1

	1	2	3	4	5	6
1	1	0.795578	0.908202	0.924205	0.903905	0.886908
2	0.795578	1	0.928756	0.812462	0.802779	0.791689
3	0.908202	0.928756	1	0.963605	0.954187	0.956742
4	0.924205	0.812462	0.963605	1	0.990628	0.989062
5	0.903905	0.802779	0.954187	0.990628	1	0.987139
6	0.886908	0.791689	0.956742	0.989062	0.987139	1

$$\mathbf{R} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}' = \sum_{i=1}^{6} \lambda_i \mathbf{c}_i \mathbf{c}_i'$$

	1	2	3	4	5	6
1	0.399548	-0.121844	-0.895708	-0.0406948	-0.127852	0.0722466
2	0.377099	0.840502	0.067997	0.177137	0.0355656	0.337768
3	0.420955	0.198986	0.132743	-0.413014	-0.104492	-0.764252
4	0.419339	-0.258255	0.101987	0.0247916	0.862514	0.050123
5	0.416351	-0.28231	0.222987	0.750782	-0.325211	-0.166715
6	0.414441	-0.3045	0.339614	-0.481765	-0.348967	0.516048

	λ
1	5.53961
2	.29845
3	.13847
4	.01478
5	.00608
6	.00260

- For positive definite matrix, elements of $\boldsymbol{\Lambda}$ are all positive.
- General result for a power of a matrix: $\boldsymbol{\Omega}^a = \mathbf{C}\boldsymbol{\Lambda}^a\mathbf{C}'$. Characteristic roots are powers of elements of $\boldsymbol{\Lambda}$. $\mathbf{C}$ is the same.
- Important cases:
  - Inverse: $\boldsymbol{\Omega}^{-1} = \mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}'$
  - Square root: $\boldsymbol{\Omega}^{1/2} = \mathbf{C}\boldsymbol{\Lambda}^{1/2}\mathbf{C}'$
  - Inverse of square root: $\boldsymbol{\Omega}^{-1/2} = \mathbf{C}\boldsymbol{\Lambda}^{-1/2}\mathbf{C}'$
  - Matrix to zero power: $\boldsymbol{\Omega}^0 = \mathbf{C}\boldsymbol{\Lambda}^0\mathbf{C}' = \mathbf{C}\mathbf{I}\mathbf{C}' = \mathbf{I}$

Generalized Least Squares – Finding $\mathbf{P}$

(Using powers of the matrix)

$$E[\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}|\mathbf{X}^*] = \mathbf{P}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}^*]\mathbf{P}'$$

$$= \mathbf{P}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}]\mathbf{P}'$$

$$= \sigma^2\mathbf{P}\boldsymbol{\Omega}\mathbf{P}' = \sigma^2\,\boldsymbol{\Omega}^{-1/2}\;\boldsymbol{\Omega}\;\boldsymbol{\Omega}^{-1/2} = \sigma^2\,\boldsymbol{\Omega}^{0}$$

$$= \sigma^2\mathbf{I}$$

Generalized Least Squares

Efficient estimation of β and, by implication, the inefficiency of least squares **b**.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*\prime}\mathbf{X}^{*})^{-1}\mathbf{X}^{*\prime}\mathbf{y}^{*}$$
$$= (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y}$$
$$= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$$

$\hat{\boldsymbol{\beta}} \neq \mathbf{b}$. $\hat{\boldsymbol{\beta}}$ is efficient, so by construction, **b** is not.

Asymptotics for GLS

Asymptotic distribution of GLS. (NOTE. We apply the full set of results of the classical model to the transformed model.)

Unbiasedness

Consistency - “well behaved data”

Asymptotic distribution

Test statistics

Unbiasedness

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X'\Omega^{-1}y}$$
$$= \boldsymbol{\beta} + (\mathbf{X'\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X'\Omega^{-1}}\boldsymbol{\varepsilon}$$

$$\text{E}[\hat{\boldsymbol{\beta}} \mid \mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X'\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X'\Omega^{-1}}\text{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}]$$

$$= \boldsymbol{\beta} \quad \text{if} \quad \text{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}$$

Consistency

Use Mean Square

$$\text{Var}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = \frac{\sigma^2}{n}\left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}\right)^{-1} \to \mathbf{0}\,?$$

Requires $\left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}\right)$ to be "well behaved"

Either converge to a constant matrix or diverge.

Heteroscedasticity case: Easy to establish

$$\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\Omega}^{-1})_{ii}\mathbf{x}_i\mathbf{x}_i' = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega^{ii}}\mathbf{x}_i\mathbf{x}_i'$$

Autocorrelation case: Complicated. Need assumptions

$$\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{\Omega}^{-1})_{ij}\mathbf{x}_i\mathbf{x}_j'.\ n^2 \text{ terms.}$$

Asymptotic Normality

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{n}\left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}\right)^{-1}\frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}$$

Converge to normal with a stable variance O(1)?

$$\left(\frac{\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}}{n}\right)^{-1} \rightarrow \text{ a constant matrix? Assumed.}$$

$$\frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} \rightarrow \text{ a mean to which we can apply the central limit theorem?}$$

Heteroscedasticity case?

$$\frac{1}{n}\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{x}_i}{\sqrt{\omega_i}}\left(\frac{\varepsilon_i}{\sqrt{\omega_i}}\right).\ \text{Var}\left(\frac{\varepsilon_i}{\sqrt{\omega_i}}\right) = \sigma^2, \frac{\mathbf{x}_i}{\sqrt{\omega_i}} \text{ is just data.}$$

Apply Lindeberg-Feller.

Autocorrelation case? More complicated.

Asymptotic Normality (Cont.)

For the autocorrelation case

$$\frac{1}{n}\mathbf{X}'\mathbf{\Omega}^{-1}\boldsymbol{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{\Omega}^{ij}\,\mathbf{x}_i\varepsilon_j$$

Does the double sum converge? Uncertain. Requires elements of $\mathbf{\Omega}^{-1}$ to become small as the distance between i and j increases. (Has to resemble the heteroscedasticity case.)

Test Statistics (Assuming Known **Ω**)

- With known **Ω**, apply all familiar results to the transformed model:
- With normality, t and F statistics apply to least squares based on **Py** and **PX**
- With asymptotic normality, use Wald statistics and the chi-squared distribution, still based on the transformed model.

Unknown Ω

- Ω would be known in narrow heteroscedasticity cases.
- Ω is usually unknown. For now, we will consider two methods of estimation

  - **Two step, or feasible estimation**. Estimate Ω first, then do GLS. Emphasize - same logic as White and Newey-West. We don't need to estimate Ω. We need to find a matrix that behaves the same as $(1/n)\mathbf{X}'\Omega^{-1}\mathbf{X}$.

  - **Full information estimation** of β, σ^2, and Ω all at the same time. Joint estimation of all parameters. Fairly rare. Some generalities.

- We will examine Harvey's model of heteroscedasticity

Specification

- $\mathbf{\Omega}$ must be specified first.
- A full unrestricted $\mathbf{\Omega}$ contains n(n+1)/2 - 1 parameters. (Why minus 1? Remember, tr($\mathbf{\Omega}$) = n, so one element is determined.)
- $\mathbf{\Omega}$ is generally specified in terms of a few parameters. Thus, $\mathbf{\Omega} = \mathbf{\Omega}(\boldsymbol{\theta})$ for some small parameter vector $\boldsymbol{\theta}$. It becomes a question of estimating $\boldsymbol{\theta}$.

Two Step Estimation

The general result for estimation when $\boldsymbol{\Omega}$ is estimated.

GLS uses $[\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}$ which converges in probability to $\boldsymbol{\beta}$.

We seek a vector which converges to the same thing that this does. Call it “Feasible GLS” or FGLS, based on $[\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$

The object is to find a set of parameters such that

$$[\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}]\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y} \;-\; [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} \;\rightarrow \mathbf{0}$$

Two Step Estimation of the Generalized Regression Model

Use the Aitken (Generalized Least Squares - GLS) estimator with an estimate of $\boldsymbol{\Omega}$

1. $\boldsymbol{\Omega}$ is parameterized by a few estimable parameters. Examples, the heteroscedastic model
2. Use least squares residuals to estimate the variance functions
3. Use the estimated $\boldsymbol{\Omega}$ in GLS - Feasible GLS, or FGLS

[4. Iterate? Generally no additional benefit.]

FGLS vs. Full GLS

VVIR (Theorem 9.5)

To achieve full efficiency, we do not need an efficient estimate of the parameters in $\mathbf{\Omega}$, only a consistent one.

Heteroscedasticity

Setting: The regression disturbances have unequal variances, but are still not correlated with each other:

Classical regression with hetero-(different) scedastic (variance) disturbances.

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \quad E[\varepsilon_i] = 0, \quad Var[\varepsilon_i] = \sigma^2 \omega_i, \ \omega_i > 0.$$

A normalization: $\Sigma_i \, \omega_i = n$. The classical model arises if $\omega_i = 1$.

A characterization of the heteroscedasticity: Well defined estimators and methods for testing hypotheses will be obtainable if the heteroscedasticity is "well behaved" in the sense that no single observation becomes dominant.

Generalized (Weighted) Least Squares
Heteroscedasticity Case

$$\text{Var}[\varepsilon \mid \mathbf{X}] = \sigma^2\Omega = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \omega_n \end{bmatrix}$$

$$\Omega^{-1/2} = \begin{bmatrix} 1/\sqrt{\omega_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\omega_2} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1/\sqrt{\omega_n} \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}) = \left(\sum_{i=1}^{n} \frac{1}{\omega_i}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^{n} \frac{1}{\omega_i}\mathbf{x}_i y_i\right)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(\frac{y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}}{\omega_i}\right)^2}{n-K}$$

Estimation: WLS form of GLS

General result - mechanics of weighted least squares.

Generalized least squares - efficient estimation. Assuming weights are known.

Two step generalized least squares:

- Step 1: Use least squares, then the residuals to estimate the weights.
- Step 2: Weighted least squares using the estimated weights.
- (Iteration: After step 2, recompute residuals and return to step 1. Exit when coefficient vector stops changing.)

FGLS – Harvey's Model

Feasible GLS is based on finding an estimator which has the same properties as the true GLS.

Example $\text{Var}[\varepsilon_i|\mathbf{z_i}] = \sigma^2 [\text{Exp}(\gamma'\mathbf{z_i})]^2$.

True GLS would regress $y_i/[\sigma\text{Exp}(\gamma'\mathbf{z_i})]$ on the same transformation of $\mathbf{x_i}$. With a consistent estimator of $[\sigma,\gamma]$, say $[s,\mathbf{c}]$, we do the same computation with our estimates.

So long as plim $[s,\mathbf{c}] = [\sigma,\gamma]$, FGLS is as "good" as true GLS.

- Consistent
- Same Asymptotic Variance
- Same Asymptotic Normal Distribution

Harvey's Model of Heteroscedasticity

- $\text{Var}[\varepsilon_i \mid \mathbf{X}] = \sigma^2 \exp(\boldsymbol{\gamma}'\mathbf{z}_i)$
- $\text{Cov}[\varepsilon_i,\varepsilon_j \mid \mathbf{X}] = 0$

  e.g.: z_i = firm size

  e.g.: $\mathbf{z}_i$ = a set of dummy variables (e.g., countries) (The groupwise heteroscedasticity model.)
- $[\sigma^2 \boldsymbol{\Omega}]$ = diagonal $[\exp(\theta + \boldsymbol{\gamma}'\mathbf{z}_i)]$,

  $\theta = \log(\sigma^2)$

Harvey's Model

Methods of estimation:

Two step FGLS: Use the least squares residuals to estimate $(\boldsymbol{\theta},\boldsymbol{\gamma})$, then use

$$\hat{\hat{\boldsymbol{\beta}}} = \left\{\mathbf{X}'\left[\boldsymbol{\Omega}\left(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\gamma}}\right)\right]^{-1}\mathbf{X}\right\}^{-1}\mathbf{X}'\left[\boldsymbol{\Omega}\left(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\gamma}}\right)\right]^{-1}\mathbf{y}$$

Full maximum likelihood estimation. Estimate all parameters simultaneously.

A handy result due to Oberhofer and Kmenta - the "zig-zag" approach. Iterate back and forth between $(\boldsymbol{\theta},\boldsymbol{\gamma})$ and $\boldsymbol{\beta}$.

Harvey's Model for Groupwise Heteroscedasticity

Groupwise sample, y_{ig}, x_{ig},…

N groups, each with n_g observations.

$Var[\varepsilon_{ig}] = \sigma_g^2$

Let d_{ig} = 1 if observation i,g is in group g, 0 else.

= group dummy variable. (Drop the first.)

$Var[\varepsilon_{ig}] = \sigma_g^2 \exp(\theta_2 d_2 + \ldots \theta_G d_G)$

$Var_1 = \sigma_g^2$, $Var_2 = \sigma_g^2 \exp(\theta_2)$ and so on.

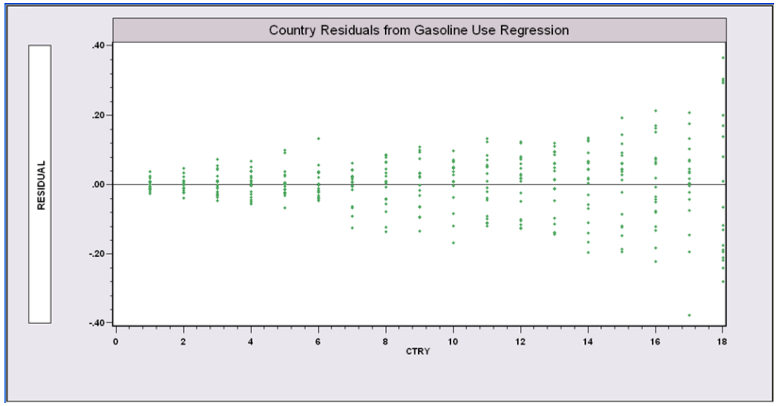
Estimating Variance Components

- OLS is still consistent:
- $\text{Est.Var}_1 = e_1'e_1/n_1$ estimates σ_g^2
- $\text{Est.Var}_2 = e_2'e_2/n_2$ estimates $\sigma_g^2 \exp(\theta_2)$, etc.
- Estimator of θ_2 is $\ln[(e_2'e_2/n_2)/(e_1'e_1/n_1)]$
- (1) Now use FGLS – weighted least squares
- Recompute residuals using WLS slopes
- (2) Recompute variance estimators
- Iterate to a solution… between (1) and (2)

Baltagi and Griffin's Gasoline Data

World Gasoline Demand Data, 18 OECD Countries, 19 years
Variables in the file are

COUNTRY = name of country
YEAR = year, 1960-1978
LGASPCAR = log of consumption per car
LINCOMEP = log of per capita income
LRPMG = log of real price of gasoline
LCARPCAP = log of per capita number of cars



See Baltagi (2001, p. 24) for analysis of these data. The article on which the analysis is based is Baltagi, B. and Griffin, J., "Gasoline Demand in the OECD: An Application of Pooling and Testing Procedures," European Economic Review, 22, 1983, pp. 117-137. The data were downloaded from the website for Baltagi's text.

Least Squares First Step

```
----------------------------------------------------------------------
Multiplicative Heteroskedastic Regression Model...
Ordinary     least squares regression ............
LHS=LGASPCAR Mean                 =         4.29624
             Standard deviation   =          .54891
             Number of observs.   =             342
Model size   Parameters           =               4
             Degrees of freedom   =             338
Residuals    Sum of squares       =        14.90436
B/P LM statistic [17 d.f.] =        111.55 (.0000)  (Large)
Cov matrix for b is sigma^2*inv(X'X)(X'WX)inv(X'X)  (Robust)
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
Constant|    2.39133***       .20010       11.951   .0000
LINCOMEP|     .88996***       .07358       12.094   .0000    -6.13943
   LRPMG|    -.89180***       .06119      -14.574   .0000     -.52310
LCARPCAP|    -.76337***       .03030      -25.190   .0000    -9.04180
--------+-------------------------------------------------------------
```

Variance Estimates = ln[e(i)'e(i)/T]

Sigma	.48196***	.12281	3.924	.0001	
D1	-2.60677***	.72073	-3.617	.0003	.05556
D2	-1.52919**	.72073	-2.122	.0339	.05556
D3	.47152	.72073	.654	.5130	.05556
D4	-3.15102***	.72073	-4.372	.0000	.05556
D5	-3.26236***	.72073	-4.526	.0000	.05556
D6	-.09099	.72073	-.126	.8995	.05556
D7	-1.88962***	.72073	-2.622	.0087	.05556
D8	.60559	.72073	.840	.4008	.05556
D9	-1.56624**	.72073	-2.173	.0298	.05556
D10	-1.53284**	.72073	-2.127	.0334	.05556
D11	-2.62835***	.72073	-3.647	.0003	.05556
D12	-2.23638***	.72073	-3.103	.0019	.05556
D13	-.77641	.72073	-1.077	.2814	.05556
D14	-1.27341*	.72073	-1.767	.0773	.05556
D15	-.57948	.72073	-.804	.4214	.05556
D16	-1.81723**	.72073	-2.521	.0117	.05556
D17	-2.93529***	.72073	-4.073	.0000	.05556

OLS vs. Iterative FGLS

```
Looks like a substantial gain in reduced standard errors
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Ordinary Least Squares
        |Robust Cov matrix for b is sigma^2*inv(X'X)(X'WX)inv(X'X)
Constant|    2.39133***       .20010       11.951   .0000
LINCOMEP|     .88996***       .07358       12.094   .0000     -6.13943
   LRPMG|    -.89180***       .06119      -14.574   .0000      -.52310
LCARPCAP|    -.76337***       .03030      -25.190   .0000     -9.04180
--------+-------------------------------------------------------------
        |Regression (mean) function
Constant|    1.56909***       .06744       23.267   .0000
LINCOMEP|     .60853***       .02097       29.019   .0000     -6.13943
   LRPMG|    -.61698***       .01902      -32.441   .0000      -.52310
LCARPCAP|    -.66938***       .01116      -59.994   .0000     -9.04180
```

Seemingly Unrelated Regressions

The classical regression model, $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$. Applies to each of M equations and T observations. Familiar example: The capital asset pricing model:

$$(\mathbf{r}_m - \mathbf{r}_f) = \alpha_m \mathbf{i} + \beta_m(\mathbf{r}_{market} - \mathbf{r}_f) + \varepsilon_{\mathbf{m}}$$

Not quite the same as a panel data model. M is usually small - say 3 or 4. (The CAPM might have M in the thousands, but it is a special case for other reasons.)

Formulation

Consider an extension of the groupwise heteroscedastic model: We had

$$\mathbf{y_i} = \mathbf{X_i}\beta + \boldsymbol{\varepsilon_i} \text{ with } E[\boldsymbol{\varepsilon_i}|\mathbf{X}] = \mathbf{0}, \text{ Var}[\boldsymbol{\varepsilon_i}|\mathbf{X}] = \sigma_i^2\mathbf{I}.$$

Now, allow two extensions:

Different coefficient vectors for each group,

Correlation across the observations at each specific point in time (think about the CAPM above. Variation in excess returns is affected both by firm specific factors and by the economy as a whole).

Stack the equations to obtain a GR model.

SUR Model

Two Equation System

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

or

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

$$\mathbf{y} \quad = \quad \mathbf{X}\boldsymbol{\beta} \; + \; \boldsymbol{\varepsilon}$$

$$\text{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \text{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}] = \text{E}\begin{bmatrix} \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1' & \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2' \\ \boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_1' & \boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2' \end{bmatrix} \mathbf{X} = \begin{bmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} \\ \sigma_{12}\mathbf{I} & \sigma_{22}\mathbf{I} \end{bmatrix}$$

$$= \sigma^2\boldsymbol{\Omega}$$

OLS and GLS

Each equation can be fit by OLS ignoring all others. Why do GLS? Efficiency improvement.

Gains to GLS:

None if identical regressors - **NOTE THE CAPM ABOVE!**

Implies that GLS is the same as OLS. This is an application of a strange special case of the GR model. “If the K columns of **X** are linear combinations of K characteristic vectors of $\boldsymbol{\Omega}$, in the GR model, then OLS is algebraically identical to GLS.” We will forego our opportunity to prove this theorem. This is our only application. (Kruskal’s Theorem)

Efficiency gains increase as the cross equation correlation increases (of course!).

The Identical **X** Case

Suppose the equations involve the same **X** matrices. (Not just the same variables, the same data. Then GLS is the same as equation by equation OLS.

Grunfeld's investment data are not an example - each firm has its own data matrix. (Text, p. 371, Example 10.3, Table F10.4)

The 3 equation model on page 344 with Berndt and Wood's data give an example. The three share equations all have the constant and logs of the price ratios on the RHS. Same variables, same years. The CAPM is also an example.

(Note, because of the constraint in the B&W system (the same δ parameters in more than one equation), the OLS result for identical Xs does not apply.)

Estimation by FGLS

Two step FGLS is essentially the same as the groupwise heteroscedastic model.

(1) OLS for each equation produces residuals $\mathbf{e}_i$.

(2) $\mathbf{S}_{ij} = (1/n)\mathbf{e}_i'\mathbf{e}_j$ then do FGLS

Maximum likelihood estimation for normally distributed disturbances: Just iterate FLS.

(This is an application of the Oberhofer-Kmenta result.)

Example 10.3 *A Cost Function for U.S. Manufacturing*

A number of studies using the translog methodology have used a four-factor model, with capital K, labor L, energy E, and materials M, the factors of production. Among the studies to employ this methodology was Berndt and Wood's (1975) estimation of a translog cost function for the U.S. manufacturing sector. The three factor shares used to estimate the model are

$$s_K = \beta_K + \delta_{KK}\ln\left(\frac{p_K}{p_M}\right) + \delta_{KL}\ln\left(\frac{p_L}{p_M}\right) + \delta_{KE}\ln\left(\frac{p_E}{p_M}\right),$$

$$s_L = \beta_L + \delta_{KL}\ln\left(\frac{p_K}{p_M}\right) + \delta_{LL}\ln\left(\frac{p_L}{p_M}\right) + \delta_{LE}\ln\left(\frac{p_E}{p_M}\right),$$

$$s_E = \beta_E + \delta_{KE}\ln\left(\frac{p_K}{p_M}\right) + \delta_{LE}\ln\left(\frac{p_L}{p_M}\right) + \delta_{EE}\ln\left(\frac{p_E}{p_M}\right).$$

Berndt and Wood's data are reproduced in Appendix Table F10.2. Constrained FGLS estimates of the parameters presented in Table 10.4 were obtained by constructing the pooled regression in (10-20) with data matrices

$$\mathbf{y} = \begin{bmatrix} s_K \\ s_L \\ s_E \end{bmatrix}, \qquad (10\text{-}35)$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{i} & 0 & 0 & \ln P_K/P_M & \ln P_L/P_M & \ln P_E/P_M & 0 & 0 & 0 \\ 0 & \mathbf{i} & 0 & 0 & \ln P_K/P_M & 0 & \ln P_L/P_M & \ln P_K/P_M & 0 \\ 0 & 0 & \mathbf{i} & 0 & 0 & \ln P_K/P_M & 0 & \ln P_L/P_M & \ln P_E/P_M \end{bmatrix},$$

$$\boldsymbol{\beta}' = (\beta_K, \beta_L, \beta_E, \delta_{KK}, \delta_{KL}, \delta_{KE}, \delta_{LL}, \delta_{LE}, \delta_{EE}).$$

TABLE 10.5 Parameter Estimates for Aggregate Translog Cost Function (Standard errors in parentheses)

	Constant	*Capital*	*Labor*	*Energy*	*Materials*
Capital	0.05689	0.02949	−0.00005	−0.01067	−0.01877*
	(0.00135)	(0.00580)	(0.00385)	(0.00339)	(0.00971)
Labor	0.25344		0.07543	−0.00476	−0.07063*
	(0.00223)		(0.00676)	(0.00234)	(0.01060)
Energy	0.04441			0.01835	−0.00294*
	(0.00085)			(0.00499)	(0.00800)
Materials	0.64526*				0.09232*
	(0.00330)				(0.02247)

Vector Autoregression

The vector autoregression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting. It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models. Forecasts from VAR models are quite flexible because they can be made conditional on the potential future paths of specified variables in the model.

In addition to data description and forecasting, the VAR model is also used for structural inference and policy analysis. In structural analysis, certain assumptions about the causal structure of the data under investigation are imposed, and the resulting causal impacts of unexpected shocks or innovations to specified variables on the variables in the model are summarized. These causal impacts are usually summarized with impulse response functions and forecast error variance decompositions.

Eric Zivot: http://faculty.washington.edu/ezivot/econ584/notes/varModels.pdf

VAR

$$y_1(t) = \gamma_{11}y_1(t-1) + \gamma_{12}y_2(t-1) + \gamma_{13}y_3(t-1) + \gamma_{14}y_4(t-1) + \delta_1 x(t) + \varepsilon_1(t)$$
$$y_2(t) = \gamma_{21}y_1(t-1) + \gamma_{22}y_2(t-1) + \gamma_{23}y_3(t-1) + \gamma_{24}y_4(t-1) + \delta_2 x(t) + \varepsilon_2(t)$$
$$y_3(t) = \gamma_{31}y_1(t-1) + \gamma_{32}y_2(t-1) + \gamma_{33}y_3(t-1) + \gamma_{34}y_4(t-1) + \delta_3 x(t) + \varepsilon_3(t)$$
$$y_4(t) = \gamma_{41}y_1(t-1) + \gamma_{42}y_2(t-1) + \gamma_{43}y_3(t-1) + \gamma_{44}y_5(t-1) + \delta_4 x(t) + \varepsilon_4(t)$$

(In Zivot's examples,

1. Exchange rates
2. y(t)=stock returns, interest rates, indexes of industrial production, rate of inflation

VAR Formulation

$\mathbf{y}(t) = \Gamma\mathbf{y}(t\text{-}1) + \boldsymbol{\delta}x(t) + \boldsymbol{\varepsilon}(t)$

SUR with identical regressors.

Granger Causality: Nonzero off diagonal elements in Γ

$$y_1(t) = \gamma_{11}y_1(t-1) + \gamma_{12}y_2(t-1) + \gamma_{13}y_3(t-1) + \gamma_{14}y_4(t-1) + \delta_1 x(t) + \varepsilon_1(t)$$

$$y_2(t) = \gamma_{21}y_1(t-1) + \gamma_{22}y_2(t-1) + \gamma_{23}y_3(t-1) + \gamma_{24}y_4(t-1) + \delta_2 x(t) + \varepsilon_2(t)$$

$$y_3(t) = \gamma_{31}y_1(t-1) + \gamma_{32}y_2(t-1) + \gamma_{33}y_3(t-1) + \gamma_{34}y_4(t-1) + \delta_3 x(t) + \varepsilon_3(t)$$

$$y_4(t) = \gamma_{41}y_1(t-1) + \gamma_{42}y_2(t-1) + \gamma_{43}y_3(t-1) + \gamma_{44}y_5(t-1) + \delta_4 x(t) + \varepsilon_4(t)$$

Hypothesis: y_2 does not Granger cause y_1: $\gamma_{12}=0$

Impulse Response

$\mathbf{y}(t) = \Gamma\mathbf{y}(t\text{-}1) + \boldsymbol{\delta}x(t) + \boldsymbol{\varepsilon}(t)$

By backward substitution or using the lag operator (text, 1022-1024)

$$\mathbf{y}(t) = \boldsymbol{\delta}x(t) + \Gamma\boldsymbol{\delta}x(t\text{-}1) + \Gamma^2\boldsymbol{\delta}x(t\text{-}2) + \ldots \text{ (ad infinitum)}$$

$$+ \boldsymbol{\varepsilon}(t) + \Gamma\boldsymbol{\varepsilon}(t\text{-}1) + \Gamma^2\boldsymbol{\varepsilon}(t\text{-}2) + \ldots$$

[Γ^P must converge to **0** as P increases. Roots inside unit circle.]

Consider a one time shock (impulse) in the system, $\lambda = \Delta\varepsilon_2$ in period t

Consider the effect of the impulse on $y_1(s)$, s=t, t+1,...

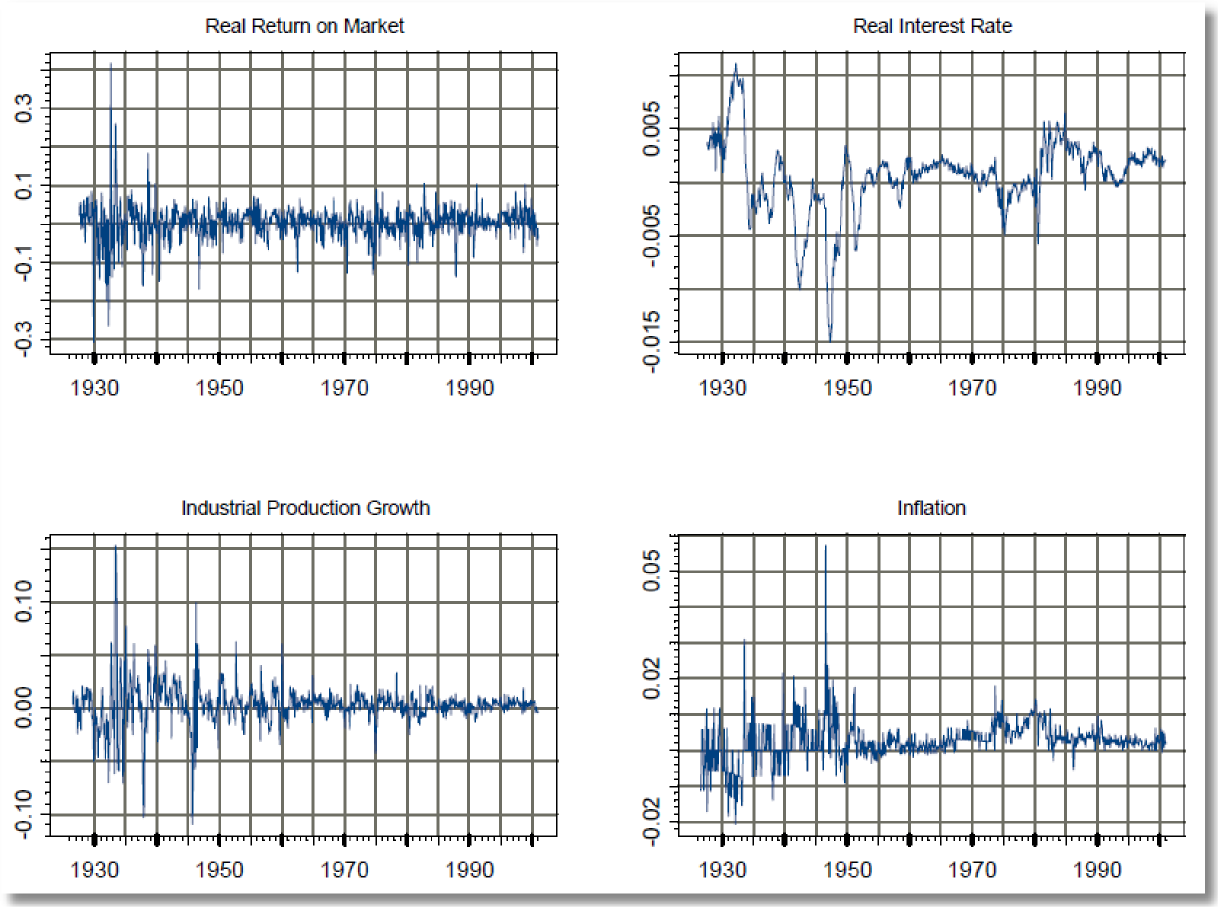
Effect in period t is 0. ε_2 is not in the y1 equation.

$\Delta\varepsilon_2$ affects y2 in period t, which affects y1 in period t+1. Effect is $\gamma_{12} \times \lambda$

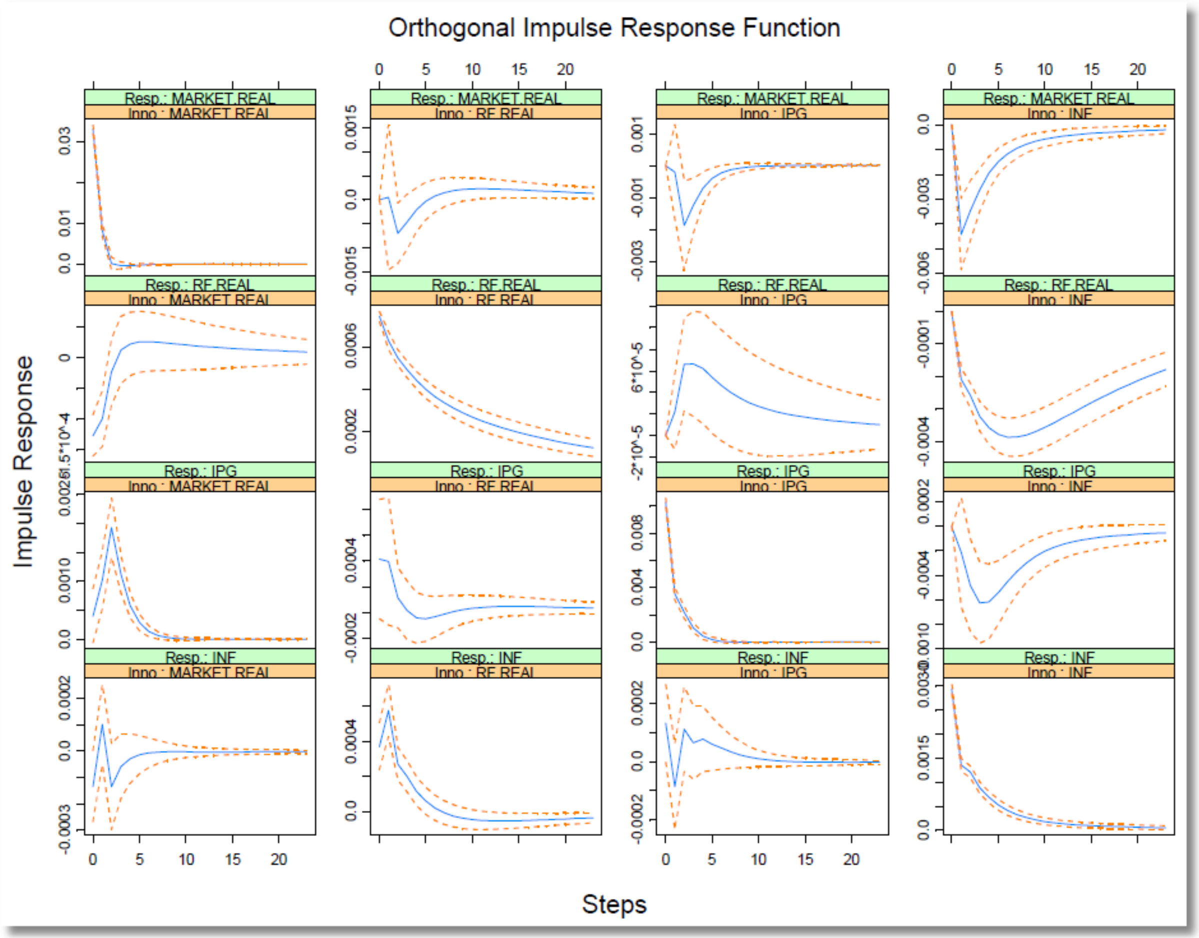
In period t+2, the effect from 2 periods back is $(\Gamma^2)_{12} \times \lambda$

... and so on.

Zivot's Data


Real Return on Market
0.3
0.1
-0.1
-0.3
1930
1950
1970
1990
Real Interest Rate
0.005
-0.005
-0.015
Industrial Production Growth
0.10
0.00
-0.10
Inflation
0.05
0.02
-0.02

Impulse Responses



Appendix: Autocorrelation in Time Series

Autocorrelation

The analysis of “autocorrelation” in the narrow sense of correlation of the disturbances across time largely parallels the discussions we’ve already done for the GR model in general and for heteroscedasticity in particular. One difference is that the relatively crisp results for the model of heteroscedasticity are replaced with relatively fuzzy, somewhat imprecise results here. The reason is that it is much more difficult to characterize meaningfully “well behaved” data in a time series context. Thus, for example, in contrast to the sharp result that produces the White robust estimator, the theory underlying the Newey-West robust estimator is somewhat ambiguous in its requirement of a bland statement about “how far one must go back in time until correlation becomes unimportant.”

Autocorrelation Matrix

$$\sigma^2\boldsymbol{\Omega} = \left(\frac{\sigma_u^2}{1-\rho^2}\right)\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$$

(Note, trace $\boldsymbol{\Omega}$ = n as required.)

Autocorrelation

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

('First order autocorrelation.' How does this come about?)

Assume $-1 < \rho < 1$. Why?

u_t = 'nonautocorrelated white noise'

$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ (the autoregressive form)

$= \rho(\rho\varepsilon_{t-2} + u_{t-1}) + u_t$

$= ...$ (continue to substitute)

$= u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \rho^3 u_{t-3} + ...$

= (the moving average form)

Autocorrelation

$$\text{Var}[\varepsilon_t] = \text{Var}[u_t + \rho u_{t-1} + \rho^2 u_{t-1} + \ldots]$$

$$= \text{Var}\left[\sum_{i=0}^{\infty} \rho^i u_{t-i}\right]$$

$$= \sum_{i=0}^{\infty} \rho^{2i} \sigma_u^2 = \frac{\sigma_u^2}{1-\rho^2}$$

An easier way: Since $\text{Var}[\varepsilon_t] = \text{Var}[\varepsilon_{t-1}]$ and $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$

$$\text{Var}[\varepsilon_t] = \rho^2 \text{Var}[\varepsilon_{t-1}] + \text{Var}[u_t] + 2\rho \text{Cov}[\varepsilon_{t-1}, u_t]$$

$$= \rho^2 \text{Var}[\varepsilon_t] + \sigma_u^2$$

$$= \frac{\sigma_u^2}{1-\rho^2}$$

Autocovariances

Continuing...

$$\begin{aligned}
\text{Cov}[\varepsilon_t,\varepsilon_{t-1}] &= \text{Cov}[\rho\varepsilon_{t-1}+u_t,\varepsilon_{t-1}] \\
&= \rho\text{Cov}[\varepsilon_{t-1},\varepsilon_{t-1}]+\text{Cov}[u_t,\varepsilon_{t-1}] \\
&= \rho\text{Var}[\varepsilon_{t\text{-}1}] = \rho\text{Var}[\varepsilon_t] \\
&= \frac{\rho\sigma_u^2}{(1-\rho^2)}
\end{aligned}$$

$$\begin{aligned}
\text{Cov}[\varepsilon_t,\varepsilon_{t-2}] &= \text{Cov}[\rho\varepsilon_{t-1}+u_t,\varepsilon_{t-2}] \\
&= \rho\text{Cov}[\varepsilon_{t-1},\varepsilon_{t-2}]+\text{Cov}[u_t,\varepsilon_{t-2}] \\
&= \rho\text{Cov}[\varepsilon_t,\varepsilon_{t-1}] \\
&= \frac{\rho^2\sigma_u^2}{(1-\rho^2)} \text{ and so on.}
\end{aligned}$$

Generalized Least Squares

$$\boldsymbol{\Omega}^{-1/2} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \ldots & 0 \\ -\rho & 1 & 0 & \ldots & 0 \\ 0 & -\rho & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & -\rho & 0 \end{bmatrix}$$

$$\boldsymbol{\Omega}^{-1/2}\mathbf{y} = \begin{pmatrix} \left(\sqrt{1-\rho^2}\right) y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \ldots \\ y_T - \rho_{T-1} \end{pmatrix}$$

GLS and FGLS

Theoretical result for known $\mathbf{\Omega}$ - i.e., known ρ. Prais-Winsten vs. Cochrane-Orcutt.

FGLS estimation: How to estimate ρ? OLS residuals as usual - first autocorrelation.

Many variations, all based on correlation of e_t and e_{t-1}

The Autoregressive Transformation

$$y_t = \mathbf{x_t}'\boldsymbol{\beta} + \varepsilon_t \qquad \varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

$$\rho y_{t-1} = \rho\mathbf{x_{t\text{-}1}}'\boldsymbol{\beta} + \rho\varepsilon_{t-1}$$

$$- - - - - - - - - - -$$

$$y_t - \rho y_{t-1} = (\mathbf{x_t} - \rho\mathbf{x_{t\text{-}1}})'\boldsymbol{\beta} + (\varepsilon_t - \rho\varepsilon_{t-1})$$

$$y_t - \rho y_{t-1} = (\mathbf{x_t} - \rho\mathbf{x_{t\text{-}1}})'\boldsymbol{\beta} + u_t$$

(Where did the first observation go?)

Estimated AR(1) Model

```
----------------------------------------------------------------------
AR(1) Model:      e(t) = rho * e(t-1) + u(t)
Initial value of rho       =          .87566
Maximum iterations         =               1
Method = Prais - Winsten
Iter=  1, SS=        .022, Log-L=    127.593
Final value of Rho    =              .959411
Std. Deviation:  e(t) =              .076512
Std. Deviation:  u(t) =              .021577
Autocorrelation: u(t) =              .253173
N[0,1] used for significance levels
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+------------------------------------------------------------------------
Constant|   -20.3373***       .69623      -29.211   .0000                   FGLS
      LP|    -.11379***       .03296       -3.453   .0006      3.72930
      LY|     .87040***       .08827        9.860   .0000      9.67215
    LPNC|     .05426          .12392         .438   .6615      4.38037
    LPUC|    -.04028          .06193        -.650   .5154      4.10545
     RHO|     .95941***       .03949       24.295   .0000
--------+------------------------------------------------------------------------
Constant|   -21.2111***       .75322      -28.160   .0000                   OLS
      LP|    -.02121          .04377        -.485   .6303      3.72930
      LY|    1.09587***       .07771       14.102   .0000      9.67215
    LPNC|    -.37361**        .15707       -2.379   .0215      4.38037
    LPUC|     .02003          .10330         .194   .8471      4.10545
```

The Familiar AR(1) Model

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, |\rho| < 1.$$

This characterizes the disturbances, not the regressors.

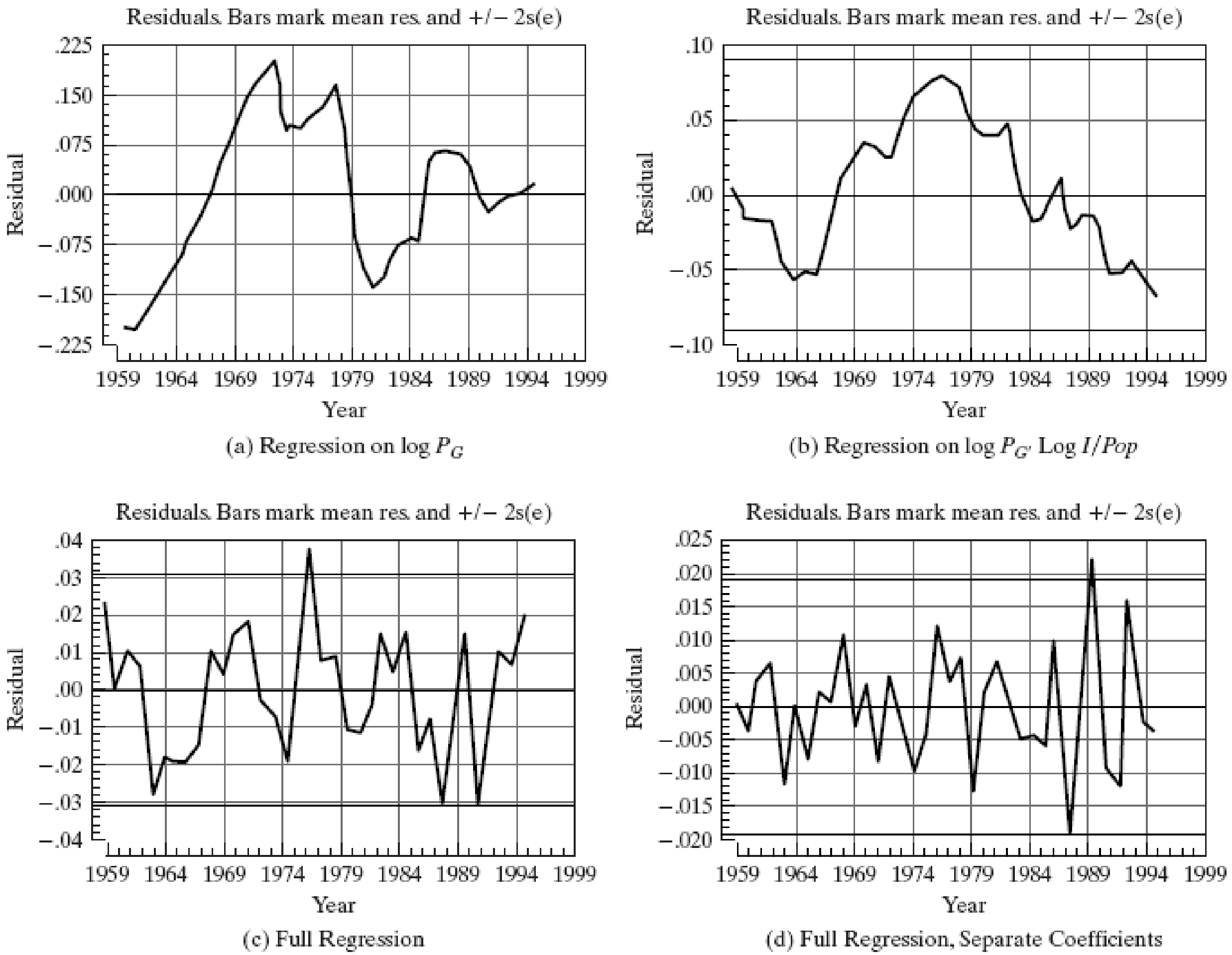
- A general characterization of the mechanism producing ε
  history + current innovations
- Analysis of this model in particular. The mean and variance and autocovariance
- Stationarity. Time series analysis.
- Implication: The form of $\sigma^2\boldsymbol{\Omega}$; Var[$\varepsilon$] vs. Var[u].
- Other models for autocorrelation - less frequently used – AR(1) is the workhorse.

Building the Model

- Prior view: A feature of the data
  - "Account for autocorrelation in the data."
  - Different models, different estimators

- Contemporary view: Why is there autocorrelation?
  - What is missing from the model?
  - Build in appropriate dynamic structures
  - Autocorrelation should be "built out" of the model
  - Use robust procedures (Newey-West) instead of elaborate models specifically for the autocorrelation.

Model Misspecification



Implications for Least Squares

Familiar results: Consistent, unbiased, inefficient, asymptotic normality

The inefficiency of least squares:

- Difficult to characterize generally. It is worst in “low frequency” i.e., long period (year) slowly evolving data.
- Can be extremely bad. GLS vs. OLS, the efficiency ratios can be 3 or more.

A very important exception - the lagged dependent variable

$$y_t = \beta x_t + \gamma y_{t-1} + \varepsilon_t. \ \varepsilon_t = \rho\varepsilon_{t-1} + u_t,.$$

Obviously, Cov[y_{t-1} ,ε_t] $\neq$ 0, because of the form of ε_t.

- How to estimate? IV
- Should the model be fit in this form? Is something missing?

Robust estimation of the covariance matrix - the Newey-West estimator.

Testing for Autocorrelation

A general proposition: There are several tests. All are functions of the simple autocorrelation of the least squares residuals. Two used generally, Durbin-Watson and Lagrange Multiplier

The Durbin - Watson test. $d \approx 2(1 - r)$. Small values of d lead to rejection of

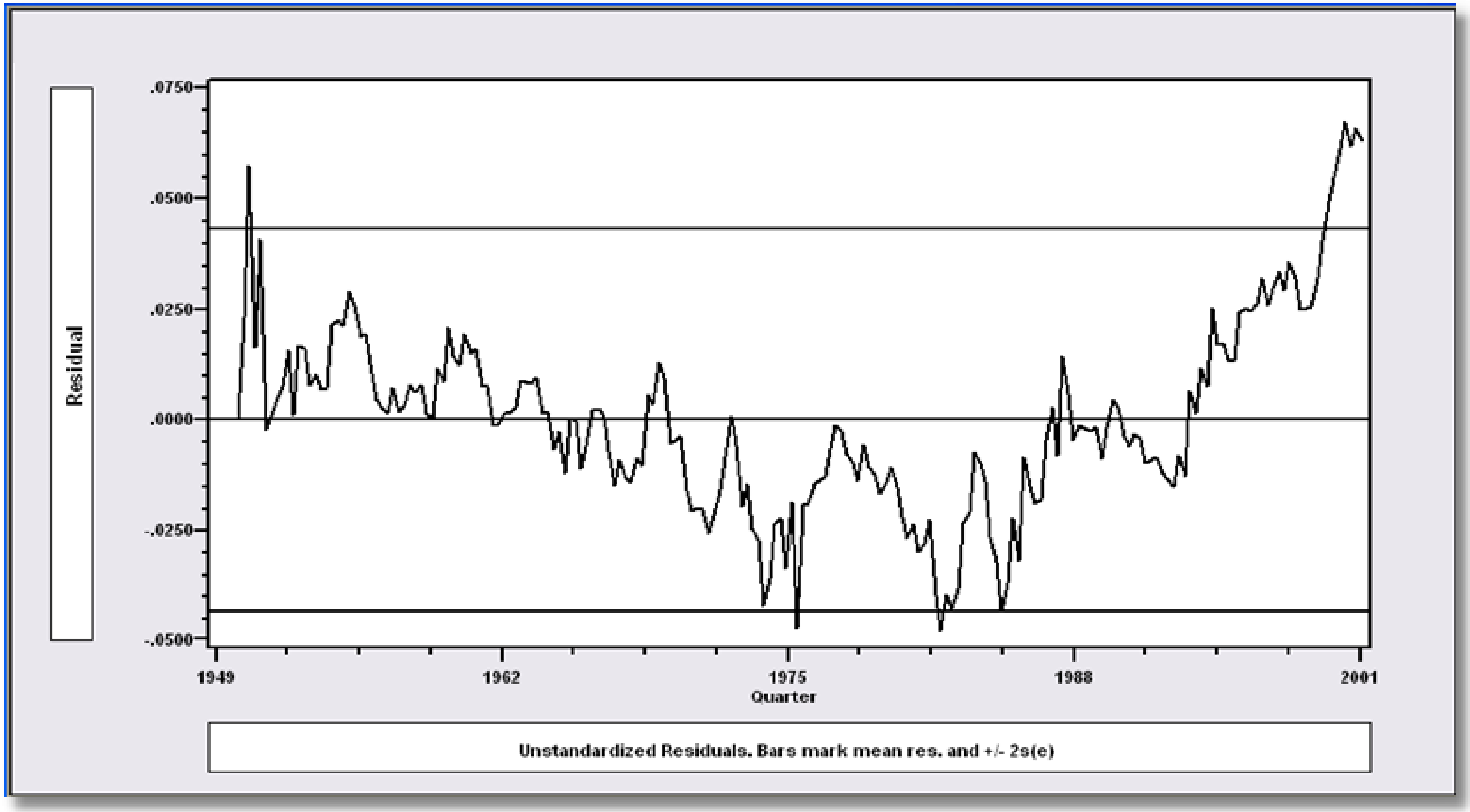
NO AUTOCORRELATION: Why are the bounds necessary?

Godfrey's LM test. Regression of e_t on e_{t-1} and $\mathbf{x}_t$. Uses a "partial correlation."

Consumption “Function”
Log real consumption vs. Log real disposable income
(Aggregate U.S. Data, 1950I – 2000IV. Table F5.2 from text)

```
----------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGC     Mean                 =        7.88005
             Standard deviation   =         .51572
             Number of observs.   =            204
Model size   Parameters           =              2
             Degrees of freedom   =            202
Residuals    Sum of squares       =         .09521
             Standard error of e  =         .02171
Fit          R-squared            =         .99824  <<<***
             Adjusted R-squared   =         .99823
Model test   F[  1,   202] (prob) =114351.2(.0000)
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  t-ratio  P[|T|>t]   Mean of X
--------+-------------------------------------------------------------
Constant|    -.13526***       .02375       -5.695   .0000
    LOGY|    1.00306***       .00297      338.159   .0000      7.99083
--------+-------------------------------------------------------------
```

Least Squares Residuals: r = .91



Conventional vs. Newey-West

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	-.13525584	.02375149	-5.695	.0000	
LOGY	1.00306313	.00296625	338.159	.0000	7.99083133

Newey-West Robust Covariance Matrix

Variable	Coefficient	Standard Error	t-ratio	P[|T|>t]	Mean of X
Constant	-.13525584	.07257279	-1.864	.0638	
LOGY	1.00306313	.00938791	106.846	.0000	7.99083133

FGLS

```
+---------------------------------------------+
| AR(1) Model:     e(t) = rho * e(t-1) + u(t) |
| Initial value of rho       =         .90693 | <<<***
| Maximum iterations         =            100 |
| Method = Prais - Winsten                    |
| Iter=  1, SS=       .017, Log-L= 666.519353 |
| Iter=  2, SS=       .017, Log-L= 666.573544 |
| Final value of Rho     =            .910496 | <<<***
| Iter=  2, SS=       .017, Log-L= 666.573544 |
| Durbin-Watson:   e(t) =             .179008 |
| Std. Deviation:  e(t) =             .022308 |
| Std. Deviation:  u(t) =             .009225 |
| Durbin-Watson:   u(t) =            2.512611 |
| Autocorrelation: u(t) =            -.256306 |
| N[0,1] used for significance levels         |
+---------------------------------------------+
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	-.08791441	.09678008	-.908	.3637	
LOGY	.99749200	.01208806	82.519	.0000	7.99083133
RHO	.91049600	.02902326	31.371	.0000	

Sorry to bother you again, but an important issue has come up. I am using LIMDEP to produce results for my testimony in a utility rate case. I have a time series sample of 40 years, and am doing simple OLS analysis using a primary independent variable and a dummy. There is serial correlation present. The issue is what is the BEST available AR1 procedure in LIMDEP for a sample of this type?? I have tried Cochrane-Orcott, Prais-Winsten, and the MLE procedure recommended by Beach-MacKinnon, with slight but meaningful differences.

By modern constructions, your best choice if you are comfortable with AR1 is Prais-Winsten. Noone has ever shown that iterating it is better or worse than not. Cochrance-Orcutt is inferior because it discards information (the first observation). Beach and MacKinnon would be best, but it assumes normality, and in contemporary treatments, fewer assumptions is better. If you are not comfortable with AR1, use OLS with Newey-West and 3 or 4 lags.

Feasible GLS

For FGLS estimation, we do not seek an estimator of $\mathbf{\Omega}$ such that

$$\hat{\mathbf{\Omega}} - \mathbf{\Omega} \rightarrow \mathbf{0}$$

This makes no sense, since $\hat{\mathbf{\Omega}}$ is nxn and does not "converge" to anything. We seek a matrix $\mathbf{\Omega}$ such that

$$(1/n)\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X} - (1/n)\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X} \rightarrow \mathbf{0}$$

For the asymptotic properties, we will require that

$$(1/\sqrt{n})\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\varepsilon - (1/n)\mathbf{X}'\mathbf{\Omega}^{-1}\varepsilon \rightarrow \mathbf{0}$$

Note in this case, these are two random vectors, which we require to converge to the same random vector.