

# **Applications of the Generalized Regression Model**

**Based on Greene's Note 16**

# Two Step Estimation of the Generalized Regression Model

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

1.  $\Omega$  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  $\Omega$  in GLS - Feasible GLS, or FGLS

# General Result for Estimation When $\Omega$ Is Estimated

- True GLS uses  $[X'\Omega^{-1} X]X'\Omega^{-1} y$  which converges in probability to  $\beta$ .
- We seek a vector which converges to the same thing that this does. Call it
- FGLS, based on  $[X'\hat{\Omega}^{-1} X]X'\hat{\Omega}^{-1} y$

The object is to find a set of parameters such that

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

This does not require us to find estimates such that  $\hat{\Omega} - \Omega \rightarrow 0$ .

This does not make sense, actually, because  $\Omega$  is  $n \times n$ , and is growing

in size with the sample size. So, you could not say that  $\hat{\Omega}$  converges to anything. What is required is that the two  $K \times K$  matrices,

$[X'\Omega^{-1}X/n]$  and  $[X'\hat{\Omega}^{-1}X/n]$  converge to the same thing. Note that it is meaningful to speak of “convergence” of these matrices, whereas

it is not to speak of convergence of  $\hat{\Omega}$ . We'll examine Harvey's model in detail. For asymptotic properties, we require that

$(1/\sqrt{n}) [X'\Omega^{-1}\epsilon - X'\hat{\Omega}^{-1}\epsilon]$  converge to zero. This is the difference of two  $K \times 1$  vectors. The logic is the same.

# FGLS

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

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

True GLS would regress  $y/[\sigma \text{Exp}((1/2)\gamma' \mathbf{z}_i)]$  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  $\text{plim } [s, \mathbf{c}] = [\sigma, \gamma]$ , FGLS is as good as true GLS.

# FGLS vs. Full GLS

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

# Heteroscedasticity

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

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

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

The classical model arises if  $\omega_i = 1$ .

A normalization:  $\sum_i \omega_i = 1$ . Not a restriction, just a scaling that is absorbed into  $\sigma^2$ .

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

$$\omega_i / \sum_i \omega_i \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

I.e., no single observation becomes dominant.

$$(1/n) \sum_i \omega_i \rightarrow \text{some stable constant.}$$

(Not a probability limit as such )

# GR Model and Testing

Implications for conventional estimation technique and hypothesis testing:

1. **b** is still unbiased. Proof of unbiasedness did not rely on homoscedasticity
2. Consistent? We need the more general proof. Not difficult.
3. If  $\text{plim } \mathbf{b} = \boldsymbol{\beta}$ , then  $\text{plim } s^2 = \sigma^2$  (with the normalization).

# Inference Based on OLS

What of  $s^2(\mathbf{X}'\mathbf{X})^{-1}$  ? Depends on  $\mathbf{X}'\boldsymbol{\Omega}\mathbf{X} - \mathbf{X}'\mathbf{X}$ . 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  $\omega_i$  is randomly drawn from a distribution with

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

just like  $(1/n)\sum_i x_i^2$ . This is the crux of the

discussion in your text.

# Inference Based on OLS

For the heteroscedasticity to be substantive wrt estimation and inference by LS, the weights must be correlated with  $x$ s and/or their squares. (Text, page 220.)

More likely, the heteroscedasticity will be important. Then,  $\mathbf{b}$  is inefficient. (Later)

The White estimator. **ROBUST** estimation of the variance of  $\mathbf{b}$ .

Implication for testing hypotheses. We will use Wald tests. Why?

# Finding Heteroscedasticity

The central issue is whether  $E[\varepsilon^2] = \sigma^2\omega_i$  is related to the  $x$ s 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  $x$ s and their squares.

# Procedures

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

Breusch and Pagan's Lagrange multiplier test.  
Regress

$[e_i^2 / (\mathbf{e}'\mathbf{e}/n) - 1]$  on  $\mathbf{Z}$  (may be  $\mathbf{X}$ ). Chi-squared.  
Is  $nR^2$  with degrees of freedom rank of  $\mathbf{Z}$ .  
(Very elegant.)

Others described in text for other purposes. E.g., groupwise heteroscedasticity. Wald, LM, and LR tests all examine the dispersion of group specific least squares residual variances

# 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.

We develop a proof based on our asymptotic theory for the asymptotic equivalence of the

# 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

# The AR(1) Model

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

Emphasize, this characterizes the disturbances, not the regressors.

A general characterization of the mechanism producing  
 $\varepsilon$  - history + innovations

Analysis of this model in particular. The mean and variance and autocovariance

“Stationarity.” Some general comments about “time series analysis.” (Not the subject of this course).

Implication: The form of  $\sigma^2\Omega$

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.

# 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

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

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

How to estimate? IV

Should the model be fit in this form? Something missing?

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

# GLS and FGLS

Theoretical result for known  $\Omega$  - i.e., known  $\rho$ . Prais-Winsten vs. Cochrane-Orcutt.

FGLS estimation: How to estimate  $\rho$ ? OLS residuals as usual - first autocorrelation.

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

a. Prais-Winsten vs. Cochrane-Orcutt.

b. The question of dropping the first observation. Should you?

# Testing for Autocorrelation

A general proposition: There are several tests. All are functions of the simple autocorrelation of the least squares residuals.

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."

Durbin's H test when lagged  $y$  is present.

$$H = (1 - d/2) (T/(1 - T \text{ Est.Var}[c]))^{1/2}$$

where  $c$  is the coefficient on the lagged  $y$ . If it is not computable, use Godfrey's test. (Durbin discovered it

# Time Series Regression

Aggregate U.S. Quarterly data on Consumption and Disposable Income (Appendix F5.1, 1950-2000 in your text.) The results below show a regression of  $\text{Log}(\text{Real Consumption})$  on a constant and  $\text{Log}(\text{Real Disposable Income})$ . The fit of the model is extremely good, as one might expect. The figure below is a simple time series plot of the residuals from this regression. The cyclical behavior which is typical of autocorrelated series is evident. (The correlation of the residuals with their one period previous values is about 0.91.)