# Econometrics Introduction to Mathematical Statistics

Paul P. Momtaz

The Anderson School UCLA

#### **Econometrics**

### Paul P. Momtaz

Ordinary Least Squares

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell)

First Thoughts

Measurement Erro Omitted Variables Rias

Simultane

Generalized Leas

Cholesky Decomposition

Maximum

Introduction

MLE Properties
The Three Classical

### **Ordinary Least Squares**

The Bivariate Model
Gauss-Markov Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell) Theorem
First Thoughts on Endogeneity
Measurement Error
Omitted Variables Bias
Simultaneity

### Generalized Least Squares

Cholesky Decomposition Feasible GLS

### Maximum Likelihood Estimation

Introduction
MLE Properties
The Three Classical Tests of MLE

### **Econometrics**

### Paul P. Momtaz

Ordinary Least Squares

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh (-Lovell Theorem

Endogeneity
Measurement Erro

Omitted Variables Bias

Generalized Least

Cholesky Decomposition

Maximum Likelihood

Introduction MLE Propert

MLE Properties
The Three Classic
Tests of MLE



# Ordinary Least Squares The Bivariate Model

Gauss-Markov Assumptions

Goodness-Of-Fit

Hypothesis Testing

Frisch-Waugh(-Lovell) Theorem

First Thoughts on Endogeneity

Measurement Error

Omitted Variables Bias

Simultaneity

### Generalized Least Squares

Cholesky Decomposition

Feasible GLS

### Maximum Likelihood Estimation

Introduction

MLE Properties

The Three Classical Tests of MLE

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

### Squares The Rivariate Model

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell

First Thoughts on Endogeneity

Omitted Variables Bias

Generalized Least

Cholesky Decomposition Feasible GLS

Maximum Likelihood

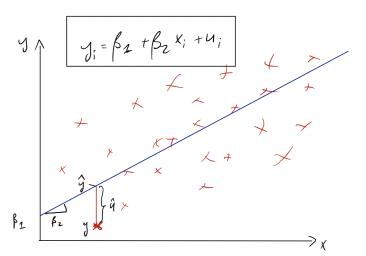
Introduction
MLE Propertie

MLE Properties The Three Classica Tests of MLE



### The basic OLS idea

### Intuition



#### Econometrics

### Paul P. Momtaz

Ordinary Least

### The Rivariate Model

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell)

> First Thoughts on Endogeneity Measurement Error

Omitted Variables Bias

Generalized Least

Cholesky Decomposition Feasible GLS

Maximum Likelihood Estimation

Introduction
MLE Properties
The Three Classical
Tests of MLE

### The basic OLS idea

### Bivariate Model

Assumptions: 
$$\mathbb{E}[u_i] = 0$$
,  $Var[u_i] = \sigma^2$ ,  $\mathbb{E}[u_i, u_j] = 0$  for  $i \neq j$ ,  $x_i$  nonstochastic

From 
$$y_i = \beta_1 + \beta_2 x_i + u_i$$
, want to minimize 
$$RSS = \sum_{i=1}^{N} (y_i - \beta_1 - \beta_2 x_i)^2$$

FOCs: 
$$\frac{\partial RSS}{\partial \beta_1} = -2 \sum_{i=1}^{N} \tilde{u}_i = 0, \qquad \frac{\partial RSS}{\partial \beta_2} = -2 \sum_{i=1}^{N} x_i \tilde{u}_i = 0$$

From first FOC: 
$$\hat{\beta}_1 = \sum_{i=1}^N \frac{y_i}{N} + \hat{\beta}_2 \sum_{i=1}^N \frac{x_i}{N} = \overline{y} - \hat{\beta}_2 \overline{x}$$

Substituting into second FOC: 
$$\hat{\beta}_2 = \frac{\sum_{i=1}^{N} (y_i - \overline{y}) \sum_{i=1}^{N} (x_i - \overline{x})}{\sum_{i=1}^{N} (x_i - \overline{x})^2}$$

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

### The Rivariate Model

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell) Theorem

ndogeneity Measurement Erro Omitted Variables

Omitted Variables Bias Simultaneity

Generalized Least

Cholesky

Maximum Likelihood

Introduction
MLE Properties
The Three Classical

4□ → 4周 → 4 章 → 4 章 → 9 Q P

### The basic OLS idea

### Bivariate Model in Matrix Notation

$$y = X\beta + u$$

Assumptions: 
$$\mathbb{E}[u] = 0$$
  $\mathbb{E}[uu'] = \sigma^2 I_N$ ,  $X$  nonstochastic

Minimize: 
$$RSS = \tilde{u}'\tilde{u} = (y - X\hat{\beta})'(y - X\hat{\beta})$$

FOC: 
$$\frac{\partial RSS}{\partial \hat{\beta}} = -2x'y + 2X'X\hat{\beta} \Rightarrow \hat{\beta} = (X'X)^{-1}X'y$$

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

### The Rivariate Model

iauss-Markov issumptions ioodness-Of-Fit lypothesis Testing irisch-Waugh(-Lovell

Endogeneity

Measurement Erro

Omitted Variables Bias

Conoralized Lo

# Squares

Decomposition Feasible GLS

Maximum Likelihood

Introduction
MLE Properties
The Three Classical

### Outille

### **Ordinary Least Squares**

The Bivariate Mode

### Gauss-Markov Assumptions

Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell) Theorem
First Thoughts on Endogeneity
Measurement Error
Omitted Variables Bias
Simultaneity

### Generalized Least Squares

Cholesky Decomposition

### Maximum Likelihood Estimation

Introduction
MLE Properties
The Three Classical Tests of MLE

#### **Econometrics**

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Mod Gauss-Markov

Assumptions Goodness-Of-Fit Hypothesis Testing

Theorem
First Thoughts or

Measurement Erro Omitted Variables

Simultaneity

Generalized Least Squares

Cholesky Decomposition Feasible GLS

Maximum Likelihood

Introduction
MLE Properties
The Three Classic

Given a linear mode,  $\hat{\beta}_{OLS}$  is a consistent estimator of  $\beta$  if

- (1)No perfect collinearity in X
- (2)No selection bias
- (3) $\mathbb{E}[u]=0$
- (4) $\mathbb{E}[Xu]=0$

(drop redundant observations)

(a representative sample)

(guaranteed if we include an intercept) Measurement Error

(unbiased if  $\mathbb{E}[u|X] = 0$ )

$$(3) + (4) \Rightarrow Var[\varepsilon] = \sigma^2 I_N.$$

$$(2) \Rightarrow \mathbb{E}[\varepsilon|x] = \mathbb{E}[\varepsilon] = 0 \text{ and } Var[\varepsilon|x] = Var[\varepsilon].$$

Assumption (4) is the focus of much attention, as we will discuss. But even without (4), we can always interpret  $\hat{\beta}_{OLS}$  as summaries of correlations in the data, rather than as parameter estimates.

### Gauss-Markov Assumptions

### Unbiasedness and Variance

### Unbiasedness:

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

$$\mathbb{E}[b|x] = \beta + \mathbb{E}[(x'x)^{-1}x'\varepsilon|x]$$
 (1)

$$\stackrel{\text{by 2}}{=} \beta + \mathbb{E}[(x'x)^{-1}x'|x]\mathbb{E}[\varepsilon|x] \tag{2}$$

$$\stackrel{\text{by 1}}{=} \beta \tag{3}$$

Variance:

$$Var[b|x] = \mathbb{E}[(b - \mathbb{E}[b])(b - \mathbb{E}[b])'|x]$$
(4)

$$= \mathbb{E}[(b-\beta)(b-\beta)'|x] \text{ from unbiasedness}$$
 (5)

$$= \mathbb{E}[(b-\beta)(b-\beta)|x| \text{ from unbiasedness}$$

$$= \mathbb{E}[(x'x)^{-1}x'\varepsilon\varepsilon'x(x'x)^{-1}|x|$$
(6)

$$= \mathbb{E}[(x'x)^{-1}x'\varepsilon\varepsilon'x(x'x)^{-1}|x] \tag{6}$$

$$= (x'x)^{-1}x'(\sigma^2 I_N)x(x'x)^{-1} \text{ by 3 and 4}$$
 (7)

$$= \sigma^2 (x'x)^{-1} \tag{8}$$

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Mod

Assumptions
Goodness-Of-Fit
Hypothesis Testing

neorem rst Thoughts on adogeneity

Aeasurement Erro Omitted Variables Bias

Simultaneity

Squares Squares

Pecomposition Peasible GLS

Maximum ikelihood

Estimation

Introduction
MLE Properties



## Gauss-Markov Assumptions

DF Adjustment, Consistency, and Asymptotics

Significance Tests:

Assumption 5: 
$$\varepsilon \sim N(0, \sigma^2 I_N) \Rightarrow b \sim N(\beta, \sigma^2 (x'x)^{-1})$$

$$\Rightarrow$$
 unbiased estimator of  $\sigma^2$ :  $S^2 = \frac{1}{N-k} \sum_{i=1}^N \varepsilon_i^2$ 

Consistency:

$$\begin{array}{c} \mathsf{Plim} = \beta \text{ as } \mathsf{N} \to \infty \text{ (Asymptotics)} \\ \mathsf{Assumption 6:} \ \ \frac{1}{\mathsf{N}} \sum_{i=1}^{\mathsf{N}} \mathsf{x}_i \mathsf{x}_i' \text{ converges to finite nonsingular matrix } \Sigma_{\mathsf{xx}}. \\ \mathsf{Assumption 7:} \ \mathbb{E}[\mathsf{x}_i \varepsilon_i] = 0 \end{array}$$

Distribution:

As 
$$N \to \infty$$
,  $\sqrt{N}(b-\beta) \to N(0, \sigma^2 \Sigma_{xx}^{-1})$   
Approximate in finite samples  $b \sim N(\beta, S^2(x'x)^{-1})$ 

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate M

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing

First Thoughts on Endogeneity Measurement Erro

Omitted Variables
Bias
Simultaneity

Generalized Least

Decomposition Feasible GLS

Likelihood Estimation

Introduction
MLE Properties
The Three Classica



### **Econometrics**

### Paul P. Momtaz

### Goodness-Of-Fit

### Ordinary Least Squares

### Goodness-Of-Fit

### Goodness-Of-Fit

$$R^{2} = \frac{\widehat{Var}(\hat{y}_{i})}{\widehat{Var}(y_{i})} = \frac{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_{i} - \overline{y}_{i})^{2}}{\frac{1}{N} \sum_{i=1}^{N} (y_{i} - \overline{y}_{i})^{2}} = \frac{\sum_{i=1}^{N} (\hat{y}_{i} - \overline{y}_{i})^{2}}{\sum_{i=1}^{N} (y_{i} - \overline{y}_{i})^{2}} = \frac{ESS}{TSS}$$

$$\widehat{Var}(y_{i}) = \widehat{Var}(\hat{y}_{i}) + \widehat{Var}(\varepsilon_{i})$$

$$R^{2} = 1 - \frac{\widehat{Var}(\varepsilon_{i})}{\widehat{Var}(y_{i})} = 1 - \frac{\sum_{i=1}^{N} \varepsilon_{i}^{2}}{\sum_{i=1}^{N} (y_{i} - \overline{y}_{i})^{2}} = 1 - \frac{RSS}{TSS}$$

Adjusted 
$$R^2=1-rac{rac{1}{N-k}\sum_{i=1}^{N}arepsilon_i^2}{rac{1}{N-1}\sum_{i=1}^{N}(y_i-\overline{y_i})^2}$$

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate N

### Goodness-Of-Fit

Hypothesis Testing
Frisch-Waugh(-Lovell

irst Thoughts on ndogeneity Measurement Error

Omitted Variables Bias

Simultaneity

#### Generalized Least Squares

Decomposition Feasible GLS

Maximum Likelihood

Introduction

MLE Properties
The Three Classic

### Econometrics

### Paul P. Momtaz

### Ordinary Least Squares

The Bivariate Model
Gauss-Markov Assumptions
Goodness-Of-Fit

### Hypothesis Testing

Frisch-Waugh(-Lovell) Theorem First Thoughts on Endogeneity Measurement Error Omitted Variables Bias Simultaneity

### Generalized Least Square

Cholesky Decomposition

### Maximum Likelihood Estimation

Introduction
MLE Properties
The Three Classical Tests of MLE

## Ordinary Least Squares

The Bivariate Mode Gauss-Markov Assumptions Goodness-Of-Fit

#### Hypothesis Testing Frisch-Waugh(-Lov

Theorem First Though

Endogeneity

Measurement Err

Omitted Variables Bias

Community

### Generalized Lea Squares

Cholesky Decomposition Feasible GLS

### Maximum Likelihood

Introduction
MLE Properties

Standard Error of 
$$b_j = SE(b_j) = \sqrt{Var(b_j)} = S\sqrt{(x'x)_{jj}^{-1}}$$

$$H_0: eta_j = eta_j^0 \qquad t_j = rac{b_j - eta_j^0}{\mathsf{SE}(b_j)} \sim t_{N-k}$$

Two-sided:  $\alpha = 5\% \rightarrow 1.96 < t_j$ ,

One sided:  $\alpha = 5\% \rightarrow 1.64 < t_j$ 

Confidence Intervals:  $b_j - T_{N-k,\frac{\alpha}{2}}SE(b_j) < \beta_j < b_j + t_{N-k,\frac{\alpha}{2}}SE(b_j)$ 

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model
Gauss-Markov
Assumptions
Goodness-Of-Fit

### Hypothesis Testing

Theorem First Thoughts on

ndogeneity Measurement Erro Omitted Variables

Omitted Variables Bias

Simultan

Generalized Least Squares

Cholesky Decomposition Feasible GLS

Maximum Likelihood

Introduction

MLE Properties
The Three Classical

# Hypothesis Testing

F test

$$H_0: \beta_{k-l+1} = \cdots = \beta_k = 0$$

Compare RSS of full and restricted model ( $S_1$  and  $S_0$ )

$$f = \frac{(S_0 - S_1/J)}{S_1/(N-k)} \sim F_{N-k}^J$$
 or  $f = \frac{(R_1^2 - R_0^2)/J}{(1 - R_1^2)/(N-k)}$ 

### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit

Hypothesis Testing Frisch-Waugh(-Lovell

First Thoughts

Measurement Erro

Omitted Variables Bias

Simultane

Generalized Least

Cholesky Decomposition Feasible GLS

Maximum Likelihood

Introduction
MLE Properties
The Three Classical
Tests of MLE

### o a cinite

### Ordinary Least Squares

The Bivariate Model
Gauss-Markov Assumptions
Goodness-Of-Fit
Hypothesis Testing

### Frisch-Waugh(-Lovell) Theorem

First Thoughts on Endogeneity Measurement Error Omitted Variables Bias Simultaneity

### Generalized Least Square

Cholesky Decomposition

### Maximum Likelihood Estimation

Introduction
MLE Properties
The Three Classical Tests of MLI

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Mode Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing

### Frisch-Waugh(-Lovell) Theorem

Endogeneity

Measurement Erro

Omitted Variables Bias

Simultaneity

### Generalized Lea Squares

Cholesky Decomposition Feasible GLS

Maximum Likelihood

Introduction
MLE Properties

## Frisch-Waugh(-Lovell) Theorem

### Partitioned Regressions

$$y = x_1 \beta_1 + x_2 \beta_2 + \varepsilon$$
 or  $\begin{bmatrix} x_1 & x_2 \\ x_1 & x_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} + \begin{bmatrix} \varepsilon \\ \varepsilon \end{bmatrix} = \begin{bmatrix} y \\ y \end{bmatrix}$ 

Premultiply by  $x'_1$  and  $x'_2$  yo get rid of  $\varepsilon$ .

Premultiply lower equation by  $(x_1'x_2)(x_2'x_2)^{-1}$  to get

$$(x_1'x_2)(x_2'x_2)^{-1}(x_2'x_1)\hat{\beta}_1 + (x_1'x_2)\hat{\beta}_2 = (x_1'x_2)(x_2'x_2)^{-1}x_2'y$$

Subtract this from upper equation to get rid of  $\hat{\beta}_2$  and define  $P_2 = x_2(x_2'x_2)^{-1}$ .

$$\hat{\beta}_1^{OLS} = [x_1'(I - P_2)x_1]^{-1}x_1'(I - P_2)y$$

So, we purged  $x_1$  off its correlation with  $x_2$ . The regression "controls" for  $x_2$ .

Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing

#### Frisch-Waugh(-Lovell) Theorem

First Thoughts on Endogeneity

Measurement Erro Omitted Variables Bias

Generalized Least

Cholesky Decomposition

Maximum Likelihood

Introduction
MLE Properties
The Three Classic

Tests of MLE



# Frisch-Waugh(-Lovell) Theorem

### Omitted Variable Bias

What did we do?

- 1. Regress  $x_1$  on  $x_2$ .
- 2. Residual matrix  $E_{1\cdot 2} = (I P_2)x_1$ .
- 3.  $y \sim E_{1.2}$

Another way to see this:

Assume 
$$\mathbb{E}[x_2|x_1] = \pi x_1$$

$$y = \beta_1 x_1 + (\beta_2 x_2 + \varepsilon)$$
  
 
$$\mathbb{E}[y|x_1] = \beta_1 x_1 + \mathbb{E}[\beta_2 x_2 | x_1] + \mathbb{E}[\varepsilon | x_1] = \beta_1 x_1 + \beta_2 \pi x_1$$

The multivariate regression decomposes the overall effect of changes in  $x_1$  into a direct effect,  $\beta_1$ , and an indirect effect associated with changes in  $x_2$ ,  $\beta_2\pi$ .

In contrast, a univariate regression would have given the overall effect  $\hat{\beta}_1^* = \hat{\beta}_1 + \pi \hat{\beta}_1$ .

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit

#### Frisch-Waugh(-Lovell) Theorem

Endogeneity
Measurement Erro
Omitted Variables

Omitted Variables Bias Simultaneity

Generalized Least

Cholesky Decomposition Feasible GLS

Maximum Likelihood

Introduction
MLE Properties
The Three Classical



### Econometrics

### Paul P. Momtaz

### Ordinary Least Squares

The Bivariate Model
Gauss-Markov
Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell)

#### First Thoughts on Endogeneity

Measurement Erro Omitted Variables Bias

# Generalized Least

Cholesky Decomposition Feasible GLS

### Maximum Likelihood

Likelihood Estimatioi

Introduction
MLE Properties
The Three Classic

### Ordinary Least Squares

The Bivariate Model
Gauss-Markov Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell) Theorem

### First Thoughts on Endogeneity

Measurement Error Omitted Variables Bias Simultaneity

### Generalized Least Square

Cholesky Decomposition Feasible GLS

### Maximum Likelihood Estimation

Introduction
MLE Properties
The Three Classical Tests of MLE

Measurement error in dependent variable not a problem for OLS consistency or validity, but problematic if in independent variable.

$$y_i = \beta x_i^* + \varepsilon_i$$

But we only observe 
$$x_i = x_i^* + v_i, \qquad v_i \sim N(0, \sigma_v^2).$$

Errors flatten out OLS line due to scattered point estimates.

$$\begin{aligned} \textit{Plim}(b) &= \textit{plim}\left(\frac{\sum x_i y_i / n}{\sum x_i^2 / n}\right) = \frac{\textit{Cov}(x^* + v, \beta x^* + \varepsilon)}{\textit{Var}(x)} \quad \text{mean } \cdot \\ &= \beta \cdot \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_v^2} = \beta \cdot \frac{1}{1 + \frac{\sigma_v^2}{\sigma_{x^*}^2}} \quad \text{Signal-to-noise-ratio} \end{aligned}$$

### Paul P. Momtaz

Ordinary Least Squares

> The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell)

First Thoughts on Endogeneity

Measurement Error

Omitted Variables

Committeed I cont

Squares

Feasible GLS

mean - zero variables

Introduction

MLE Properties
The Three Classical
Tests of MLE

We observe  $x_i = x_i^* + v_i$  where  $x_i^*$  is true value

$$y_i = \alpha + \beta x_i^* + \varepsilon_i$$
  
=  $\alpha + \beta (x_i - v_i) + \varepsilon_i$   
=  $\alpha + \beta x_i + (\varepsilon_i - \beta v_i)$   
=  $\alpha + \beta x_i + u_i$ 

Both  $x_i$  and  $u_i$  depend on  $v_i$ , so correlated  $\Rightarrow$  OLS estimation downward biased.

Measurement error in  $y_i$  increase variance in error term but does not cause endogeneity.

### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model
Gauss-Markov
Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell

First Thoughts on Endogeneity

Measurement Error

Omitted Variables Bias

Generalized Least

quares Cholesky Decomposition

Maximum Likelihood

Introduction

MLE Properties
The Three Classical
Tests of MLE

$$y = x_1 \beta_1^* + v$$
 and  $y = x_1 \beta_1 + x_2 \beta_2 + \varepsilon$ 

$$\beta_2 = (x_1'x_1)^{-1}x_1'(y - x_2\beta_2)$$

$$= (x_1'x_1)^{-1}x_1'y - (x_1'x_1)^{-1}x_1'x_2\beta_2$$

$$= \beta_1^* - (x_1'x_1)^{-1}x_1'x_2\beta_2$$

$$\Rightarrow \beta_1^* = \beta_1 + G\beta_2 \Rightarrow$$
 Asymptotic bias (unless  $G = 0$  or  $\beta_2 = 0$ )

- No OVB if  $corr(x_1, x_2) = 0$ .
- ▶ OVB does not cause  $corr(x_1, v) \neq 0$  as  $\mathbb{E}[v|x_1] = 0$  in short regression.
- ▶ But if OVB  $x_1$  corr to composite error term  $u = x_2\beta_2 + \varepsilon$ .
- If you do not control for  $x_2$ , you assume that  $x_2$  varies with changes in  $x_1$ .

#### **Econometrics**

### Paul P. Momtaz

Omitted Variables Bias

### Simultaneity

Given two structural equations  $y_i = \beta_1 x_i + \gamma_1 w_i + u_i$  and  $w_i = \beta_2 x_i + \gamma_2 y_i + v_i.$ 

$$w_i = \frac{\beta_2 + \gamma_2 \beta_1}{1 - \gamma_1 \gamma_2} x_i + \frac{1}{1 - \gamma_1 \gamma_2} v_i + \frac{\gamma_2}{1 - \gamma_1 \gamma_2} u_i$$

Assuming

$$Corr(x_i, u_i) = Corr(v_i, u_i) = 0 \Rightarrow \mathbb{E}[w_i u_i] = \frac{\gamma_2}{1 - \gamma_1 \gamma_2} \mathbb{E}[u_i u_i] \neq 0 \Rightarrow \mathbb{E}[u|w] \neq 0$$

#### **Econometrics**

### Paul P. Momtaz

### Simultaneity

# Generalized Least Squares (GLS)

$$y_t = x_t' \beta + \varepsilon_t$$
 where  $\varepsilon_t = u_t + u_{t-1} \Rightarrow \text{ error structure not homoskedastic.}$ 

$$\mathbb{E}[\varepsilon\varepsilon'] = \Sigma = \sigma^2 D = \sigma^2 \begin{bmatrix} 2 & 1 & 0 & \dots & 0 \\ 1 & 2 & 1 & \dots & \vdots \\ 0 & 1 & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & 2 \end{bmatrix}$$

#### **Econometrics**

### Paul P. Momtaz

#### Generalized Least Squares

### Econometrics

### Paul P. Momtaz

### Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell Theorem

First Thoughts on Endogeneity Measurement Erro

Omitted Variables

Generalized Least

#### Cholesky Decomposition

Feasible GLS

Maximum \_ikelihood

Introduction
MLE Properties
The Three Classic

### Ordinary Least Squares

The Bivariate Model
Gauss-Markov Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell) Theorem
First Thoughts on Endogeneity
Measurement Error
Omitted Variables Bias

### Generalized Least Squares

Cholesky Decomposition

Feasible GLS

### Maximum Likelihood Estimation

Introduction
MLE Properties
The Three Classical Tests of MLE

# Generalized Least Squares (GLS)

### Cholesky Decomposition

To demonstrate properties of GLS transform to CLRM and show that  $\mathsf{BLUE}.$ 

1. Transform

Cholesky Factorization: 
$$D = (L')^{-1}L^{-1} \Rightarrow L'DL = I_t$$
, given  $D$  symm, pos - def  
So:  $L'v = L'x\beta + L'\varepsilon \Rightarrow v^* + x^*\beta + \varepsilon^* \Rightarrow \mathsf{CLRM}$ 

2. Show Properties

By Gauss - Markov, 
$$\hat{\beta}_{OLS}^* = (x^{*'}x^*)^{-1}x^{*'}y^*$$
 is BLUE, so

$$\hat{\beta}_{GLS} = (x'LL'x)^{-1}x'LL'y$$

$$= (x'D^{-1}x)^{-1}x'D^{-1}y$$

$$= (x'(\sigma^2D)^{-1}x)^{-1}x'(\sigma^2D)^{-1}y$$

$$= (x'\Sigma^{-1}x)^{-1}x'\Sigma^{-1}y$$
Note That
$$= (x'\Sigma^{-1}x)^{-1}x'\Sigma^{-1}(x\beta + \varepsilon)$$

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

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model
Gauss-Markov
Assumptions
Goodness-Of-Fit
Hypothesis Testing
Friesh-Waugh(Lovell)

rst Thoughts o idogeneity

Measurement Error Omitted Variables Bias

Generalized Least Squares

#### Cholesky Decomposition Feasible GLS

Likelihood
Estimation
Introduction
MLE Properties
The Three Classical

# Generalized Least Squares (GLS)

So:

$$Var[\hat{\beta}_{GLS}] = \mathbb{E}[(\hat{\beta}_{GLS} - \beta)(\hat{\beta}_{GLS} - \beta)'] = (x'\Sigma^{-1}x)^{-1}$$

Intuition: GLS weights observations by the inverse square root of the variance of the error term since  $L'=D^{\frac{-1}{2}}\propto \Sigma^{-\frac{1}{2}}$ 

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell

First Thought Endogeneity

Measurement Erro Omitted Variables Bias

Generalized Least

#### Cholesky Decomposition

Decomposition Feasible GLS

Maximum Likelihood Estimation

Introduction
MLE Properties
The Three Classical
Tests of MLE

### Generalized Least Squares

Feasible GLS

### **Econometrics**

### Paul P. Momtaz

### Feasible GLS

# Feasible Generalized Least Squares (Feasible GLS)

Often  $\Sigma$  is unknown but  $\hat{\Sigma}$  exists.

Sufficient conditions for consistency and asymptotic normality of FGLS:

1. 
$$plim \frac{1}{N} x' \hat{\Sigma}^{-1} x \rightarrow Q_{\Sigma} > 0$$
.

2. 
$$plim \frac{1}{N} x' \hat{\Sigma}^{-1} \varepsilon \to 0$$

why?

$$\hat{\beta}_{FGLS} = (x'\hat{\Sigma}^{-1}x)^{-1}x'\hat{\Sigma}^{-1}y = \beta + (x'\hat{\Sigma}^{-1}x)^{-1}x'\hat{\Sigma}^{-1}\varepsilon$$

$$Plim(\hat{\beta}_{FGLS} - \beta) = Plim(x'\hat{\Sigma}^{-1}x)^{-1}plim(x'\hat{\Sigma}^{-1}\varepsilon) \to (Q_{\Sigma})^{-1} \cdot 0 = 0$$

If heteroskedasticity of unknown form, i.e.  $\mathbb{E}[\varepsilon_t^2] = \sigma_t^2$  and  $\Sigma = diag(\sigma_t^2)$ ,

We can estimate  $\hat{\sigma}_t^2 = \hat{\varepsilon}_t^2$  from a first - stage regression. But  $\hat{\Sigma} \nrightarrow \Sigma$ .

Note that  $\frac{1}{N}x'\hat{\Sigma}^{-1}x = \frac{1}{N}\sum_t \frac{1}{\hat{\sigma}_t^2}x_tx_t'$  where  $\hat{\Sigma}$  is  $k \times k$  fixed since  $N \to \infty$ . So:  $plim \frac{1}{N}x'\hat{\Sigma}^{-1}x \to Q_{\Sigma} > 0$  and  $plim \frac{1}{N}x'\hat{\Sigma}^{-1}\hat{\varepsilon} \to 0$ .

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

auss-Markov ssumptions oodness-Of-Fit ypothesis Testing risch-Waugh(-Lovell) heorem

dogeneity

leasurement Error

mitted Variables

Simultaneity

Generalized Least

Squares Cholesky

Feasible GLS

Likelihood
Estimation
Introduction
MLE Properties
The Three Classical



### Ordinary Least Square

The Bivariate Model
Gauss-Markov Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell) Theorem
First Thoughts on Endogeneity
Measurement Error
Omitted Variables Bias

### Generalized Least Square

Cholesky Decomposition

### Maximum Likelihood Estimation

### Introduction

MLE Properties
The Three Classical Tests of MI

#### Econometrics

### Paul P. Momtaz

# Ordinary Least Squares

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell Theorem

Endogeneity
Measurement Erro

Omitted Variables Bias Simultaneity

### Generalized Least

Cholesky Decomposition

Maximum Likelihood

#### Introduction

MLE Properties
The Three Classica
Tests of MLE



### MLF Introduction

Intuition: Find  $\theta$  that maximizes the likelihood of observations.

$$\max_{\theta} L(\theta; y) = \max_{\theta} \prod_{i=1}^{N} f(y_i; \theta) \text{ or } \max_{\theta} \log L(\theta; y) = \max_{\theta} \sum_{i=1}^{N} \log f(y_i; \theta)$$

Efficient score:  $S(\theta) = \frac{\partial \log L}{\partial \theta}$ . MLE is solution to  $S(\hat{\theta}) = 0$ .

### Example:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \qquad \varepsilon_i \sim IN(0, \sigma^2), \qquad y_i \sim IN(\alpha + \beta x_i, \sigma^2)$$

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (y_i - \alpha - \beta x_i)^2\right]$$
$$\log L(\alpha, \beta, \sigma^2; data) = \sum_{i=1}^{N} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (y_i - \alpha - \beta x_i)^2\right]$$

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell)

rst Thoughts on ndogeneity Measurement Erro

Omitted Variables Bias Simultaneity

Generalized Least Squares

Decomposition Feasible GLS

> Maximum Likelihood Estimation

### Introduction

MLE Properties
The Three Classica

### MLE Introduction

FOC's:

$$[\alpha] \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0$$
$$[\beta] \sum_{i=1}^{N} x_i (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0$$
$$[\sigma^2] \underbrace{\mathcal{N}\hat{\sigma}^2}_{\text{not} (N-2)!} = \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

Plug in  $\hat{\sigma}^2$  in  $\log L(\alpha, \beta, \sigma^2; data)$  to see maximized  $\log L$  is

$$\log L(\hat{\theta}) = \text{constant} - \frac{N}{2} \log \hat{\sigma}^2 = \text{constant} - \frac{N}{2} \log \left( \frac{RSS}{N} \right)$$

Individual Score: 
$$S_i(\theta) = \frac{\partial \log L_i(\theta)}{\partial \theta}$$

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell) Theorem

Endogeneity

Measurement Erro

Omitted Variables Bias

Simultaneity

Squares Cholesky

Decomposition Feasible GLS

Maximum Likelihood

### Introduction

MLE Properties
The Three Classical

### Maximum Likelihood Estimation

### **MLE Properties**

#### **Econometrics**

### Paul P. Momtaz

MLE Properties The Three Classical



### MLE Properties

Average information matrix :  $\overline{J}_N(\theta) = -\mathbb{E}\left[\frac{1}{N}\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\right]$ Limiting information matrix:  $J(\theta) \equiv \lim_{N \to \infty} \overline{J}_N(\theta)$ 

### $MLE, \hat{\theta}, is$

- Consistent
- Asymptotically normal since  $\sqrt{N}(\hat{\theta} \theta) \rightarrow N(0, [J(\theta)]^{-1})$ .
- Asymptotically efficient (Cramer Rao lower bound).
- Invariant (continuous function theorem: MLE of  $g(\theta)$  for any  $g(\hat{\theta})$ .
- $\triangleright$   $Var[S_i(\theta)] = \mathbb{E}[S_i(\theta)S_i(\theta)'] = J_i(\theta).$

Can be estimated outer product of gradients (G)

### **Econometrics**

### Paul P. Momtaz

MLE Properties

$$\hat{V}_{G} = \left[\frac{1}{N}\sum_{i=1}^{N}S_{i}(\hat{\theta})S_{i}(\hat{\theta})'\right]^{-1}$$

or Hessian

$$\hat{V}_H = \left[ -rac{1}{N} \sum_{i=1}^N rac{\partial^2 \log L_i( heta)}{\partial heta \partial heta'} \Big|_{\hat{ heta}} 
ight]^{-1}$$

### Ordinary Least Square

The Bivariate Model
Gauss-Markov Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell) Theorem
First Thoughts on Endogeneity
Measurement Error
Omitted Variables Bias

### Generalized Least Square

Cholesky Decomposition

### Maximum Likelihood Estimation

Introduction
MLE Properties

The Three Classical Tests of MLF

#### **Econometrics**

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell) Theorem

Endogeneity
Measurement Erro

Omitted Variables Bias Simultaneity

Generalized Least

Cholesky Decomposition

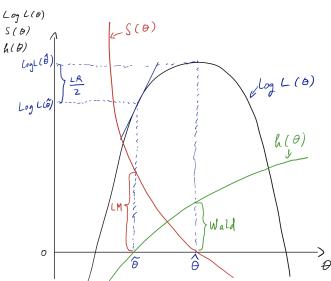
Maximum Likelihood

Introduction

MLE Propertie

### The Three Classical Tests of MLE

Idea: Set of restrictions to be tested  $H_0: h(\theta) = 0$   $\hat{\theta}$  and  $\tilde{\theta}$  are unrestricted and restricted MLE, respectively.



#### Econometrics

#### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell

Endogeneity
Measurement Erro

Omitted Variables Bias

Generalized Least Squares

Squares Cholesky Decomposition

Maximum
Likelihood
Estimation
Introduction

MLE Properties
The Three Classical
Tests of MLE

### Wald Test

- ▶ Approximate  $Var[h(\hat{\theta})] = G(\theta)' Var[\hat{\theta}] G(\theta)$  where  $G(\theta) = \frac{\partial h(\theta)'}{\partial \theta}$ .
- ► Test Statistic  $\xi_W = Nh(\hat{\theta})'[G(\theta)'[J(\theta)]^{-1}G(\theta)]^{-1}h(\hat{\theta}), \xi_W \sim_a \chi^2(d).$
- ▶ Estimate  $[J(\theta)]^{-1}$  by  $\hat{V}_H$  or  $\hat{V}_G$  evaluated at  $\hat{\theta}$ , evaluate  $G(\theta)$  at  $\hat{\theta}$ .
- ► Shortcoming: Wald test not invariant to how restrictions (linear/non-linear) are formulated.
- Overall question: How close is  $h(\hat{\theta})$  to zero since  $h(\tilde{\theta}) = 0$ ?

#### Econometrics

#### Paul P. Momtaz

Ordinary Least Squares

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell) Theorem

Endogeneity

Measurement Erro

Omitted Variables Bias Simultaneity

. Generalized Lea

Squares

Cholesky Decomposition

Maximum Likelihood

Estimation Introduction

### Likelihood Ratio Test

- ► How close are  $L(\hat{\theta})$  and  $L(\tilde{\theta})$ ?

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

Gauss-Markov
Assumptions
Goodness-Of-Fit
Hypothesis Testing
Frisch-Waugh(-Lovell

irst Thoughts of Indogeneity

Omitted Variables Bias

Simultaneity

Generalized Least Squares

> Cholesky Decomposition Feasible GLS

Maximum Likelihood

troduction

# Lagrange Multiplier or Score Test

- ▶ How close is  $S(\tilde{\theta})$  to zero given  $S(\hat{\theta}) = 0$ ?

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell

First Thoughts of Endogeneity

Measurement Erro Omitted Variables Bias

Simultanei

Generalized Least

Cholesky Decomposition

Maximum

Estimation

ntroduction

## Lagrange Multiplier or Score Test

### Alternative Computation

### Alternative for $\xi_{IM}$

- ▶ Supplementary regression  $\tilde{u} \sim x$
- $ightharpoonup RSS = N\hat{\sigma}^2$ .
- $ightharpoonup TSS = N\tilde{\sigma}^2$

$$\Rightarrow R_u^2 = 1 - rac{\hat{\sigma}^2}{\tilde{\sigma}^2}, \qquad \xi_{LM} = NR_u^2$$

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing Frisch-Waugh(-Lovell)

First Thou Endogene

> Measurement Erro Omitted Variables Bias

Simultar

Generalized Least Squares

Cholesky Decomposition

Maximum Likelihood

Estimation Introduction

ntroduction MLE Properties

### The Three Classical Tests of MLE

### Application

$$y = x\beta + u$$
  $N(0, \sigma^2 I)$ 

Unrestricted Estimation:  $\log L(\hat{\beta}, \hat{\sigma}^2) = \text{constant} - \frac{N}{2} \log \hat{\sigma}^2$ 

Restricted Estimation:  $\log L(\tilde{\beta}, \tilde{\sigma}^2) = \text{constant} - \frac{N}{2} \log \tilde{\sigma}^2$ 

$$\xi_{LR} = N \log \left( \frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right), \qquad \xi_N = N \left( \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \right), \qquad \xi_{LM} = N \left( \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\tilde{\sigma}^2} \right)$$

Note that test statistics are functions of each other

$$\xi_{\mathit{LR}} = \mathit{N} \log \left( 1 + \frac{\xi_{\mathit{w}}}{\mathit{N}} \right), \qquad \xi_{\mathit{LM}} = \frac{\xi_{\mathit{w}}}{1 + \frac{\xi_{\mathit{w}}}{\mathit{N}}}$$

Linear Case:  $\xi_w \geq \xi_{LR} \geq \xi_{LM}$ 

#### Econometrics

### Paul P. Momtaz

Ordinary Least Squares

The Bivariate Model Gauss-Markov Assumptions Goodness-Of-Fit Hypothesis Testing

First Thoughts on Endogeneity

Omitted Variables Bias Simultaneity

Generalized Least

Squares Cholesky Decomposition

Maximum

Estimation
Introduction
MLE Properties