

# The Linear Model

Robert A. Miller

Structural Econometrics

November 2020

# Introduction

## Basic setup

- The linear model is defined by the equation:

$$y_n = x_n' \beta_0 + \epsilon_n \quad (1)$$

where  $n \in \{1, 2, \dots\}$  belongs to a population and:

- $y_n$  is a  $1 \times 1$  observed dependent variable
  - $x_n$  is a  $k \times 1$  vector of observed explanatory variables
  - $\beta_0$  is a  $k \times 1$  unknown parameter to be estimated
  - $\epsilon_n$  is a  $1 \times 1$  unobserved idiosyncratic variable.
- The goal is to estimate  $\beta_0$  from a sample  $\{y_n, x_n\}_{n=1}^N$  of size  $N$ .
  - There are essentially three reasons why the linear model has become the workhorse in econometrics:
    - 1 the model is easy to understand
    - 2 the estimator for the unknown coefficient is easy to compute
    - 3 the finite sample properties of the estimator are known
  - To preface nonlinear estimation this lecture reviews the linear model.

# Introduction

## Example 1: differences in differences

- To illustrate one application of the linear model consider a *differences in differences* (DID) framework.
- Here the goal is to decontaminate the effects of a changing a regime, or more generally the effect of a particular factor of interest, from other extraneous factors, such as a time trend.
- We might write:

$$y_n = \beta_{00} + \beta_{01}t_n + \beta_{02}x_n + \beta_{03}x_nt_n + \epsilon_n$$

where  $\beta_0 \equiv (\beta_{00}, \beta_{01}, \beta_{02}, \beta_{03})$  and  $(x_n, t_n) \in \{0, 1\} \times \{0, 1\}$ .

- Intuitively there are  $N$  observations, some of which are sampled in the first period, the others in the second, where a proportion are treated with a factor of interest (setting  $x_n = 1$ ) and a proportion are left untreated (setting  $x_n = 0$ ).
- This model is *saturated* because there are as many coefficients to be estimated as there are different combinations of  $(x, t)$ .

# Introduction

## Example 2: regression discontinuity design

- A second example is the *regression discontinuity design* (RDD) framework.
- Similar in some ways to DID, we seek to separate the effects of a changing a regime from other nonlinear effects that a particular explanatory variable might have on the dependent variable.
- For example let:

$$y_n = \beta_{00} + \beta_{0K} \mathbf{1}\{x_n \leq c\} + \sum_{k=1}^{K-1} \beta_{0k} x_n^k + \epsilon_n$$

where  $\beta_0 \equiv (\beta_{00}, \beta_{01}, \dots, \beta_{0K})$  and  $c \in \mathbb{R}$  is a cut-off value that might be crucial to determining how  $x$  affects  $y$ .

- This framework is used to flexibly model known discontinuities within an otherwise smooth nonlinear equation.

# Introduction

## Example 3: fixed effects

- Models of *fixed effects* (FE) arise when there are multiple observations on each individual  $n \in \{1, 2, \dots, N\}$ , perhaps because they are sampled over time  $t \in \{1, 2, \dots, T\}$ .
- Alternatively there might be several measurements of dependent variable, each of which is measured with error.
- We extend the notation for characterizing the data by writing:

$$y_{nt} = x'_{nt}\beta_0 + \gamma_n + \epsilon_{nt} \quad (2)$$

where:

- $y_{nt}$  is a  $1 \times 1$  observed dependent variable
  - $x_{nt}$  is a  $k \times 1$  vector of observed explanatory variables
  - $\beta_0$  is a  $k \times 1$  unknown parameter to be estimated
  - $\gamma_n$  is a  $k \times 1$  unknown ancillary (or nuisance) parameter
  - $\epsilon_{nt}$  is a  $1 \times 1$  unobserved idiosyncratic variable.
- We estimate  $\beta_0$  from *panel data*  $\{y_{nt}, x_{nt}\}_{n=1}^{NT}$  with  $NT$  observations.

# Introduction

## The ordinary least squares estimator

- The *ordinary least squares* (OLS) estimator of  $\beta_0$  is defined as:

$$\begin{aligned}\beta_{OLS}^{(N)} &\equiv \arg \min_{\beta} \left\{ \sum_{n=1}^N (y_n - x_n' \beta)^2 \right\} \\ &= \arg \min_{\beta} \sum_{n=1}^N [y_n^2 - 2\beta' x_n y_n + (x_n' \beta) (x_n' \beta)]\end{aligned}$$

- The  $k \times 1$  *first order condition* (FOC) for this problem is:

$$0 = -2 \sum_{n=1}^N x_n y_n + 2 \sum_{n=1}^N x_n x_n' \beta_{OLS}^{(N)}$$

- If the  $k \times k$  matrix  $\frac{1}{N} \sum_{n=1}^N x_n x_n'$  has a nonzero determinant, then it is *invertible* and:

$$\beta_{OLS}^{(N)} = \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N x_n y_n \right) \quad (3)$$

- If  $\frac{1}{N} \sum_{n=1}^N x_n x_n'$  is not invertible then the solution to this quadratic minimization problem is not unique.

# Linear Projections

## Metrics for approximations

- Let  $F_{y,x}(y, x)$  denote the joint distribution function of  $(y, x)$  for the population, or data generating process.
- Also define an  $L_p$  space of real valued functions of  $(y, x)$ , with elements  $h(y, x) \in L_p$ , by the condition:

$$\int |h(y, x)|^p dF_{y,x}(y, x) < \infty$$

equipped with norm:

$$\|h(y, x)\|_{L_p} \equiv \left[ \int |h(y, x)|^p dF_{y,x}(y, x) \right]^{\frac{1}{p}}$$

- Given an  $L_p$  space define the *linear projection* of  $y$  on to  $x$  as:

$$\beta_{\|\cdot\|_{L_p}} \equiv \arg \min_{\beta \in \mathbb{R}^k} \|y - x'\beta\|_{L_p} = \arg \min_{\beta \in \mathbb{R}^k} \{E[|y - x'\beta|^p]\} \quad (4)$$

- Thus  $\beta_{\|\cdot\|_{L_p}}$  defines how closely a linear function of  $x$  is to central tendencies of the conditional distribution  $F_{y|x}(y|x)$ .

# Linear Projections

## Projecting $y$ on $x$

- If  $p = 2$ , then  $\beta_{\|\cdot\|_{L^p}}$  becomes:

$$\beta_{OLS} \equiv \arg \min E \left[ (y - x'\beta)^2 \right] = \arg \min E \left[ -2\beta'xy + (x'\beta)^2 \right]$$

with the FOC reducing to:

$$E [yx'] = E [xx'] \beta_{OLS}$$

- If  $E [x_n x_n']$  is invertible then:

$$\beta_{OLS} = E [xx']^{-1} E [xy]$$

- In this case  $\beta_{OLS}^{(N)}$  is the *sample analogue* of  $\hat{\beta}$ , is found by replacing:

$$E [xx'] \text{ with } \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right) \text{ and } E [xy] \text{ with } \left( \frac{1}{N} \sum_{n=1}^N x_n y_n \right)$$



# Linear Projection

Projecting  $y$  on  $x$  with a different norm

- Using a different norm changes the solution to the linear projection.
- For example if  $\|z\| \equiv E[|z|]$  then (4) reduces to:

$$\beta_{LAD} \equiv \arg \min_{\beta} E[|y - x'\beta|] = \arg \min_{\beta} E[\max\{y - x'\beta, x'\beta - y\}]$$

- The sample analogue of  $\beta_{LAD}$ , called the *least absolute deviations* (LAD) estimator, minimizes:

$$\frac{1}{N} \sum_{n=1}^N |y_n - x_n'\beta| \quad (5)$$

- Note (5) is not differentiable with respect to  $\beta$  wherever  $y_n = x_n'\beta$ .
- Nevertheless  $\hat{\beta}_{LAD}^{(N)}$  is the solution to the linear program:

$$\hat{\beta}_{LAD}^{(N)} \equiv \arg \min_{\beta, u_1, \dots, u_N} \frac{1}{N} \sum_{n=1}^N u_n$$

$$\text{such that } u_n \geq y_n - x_n'\beta \text{ and } u_n \geq x_n'\beta - y_n$$

# Quantile Estimators

## Rationale and definition

- The LAD estimator is an example of a *quantile estimator*.
- For any  $\tau \in (0, 1)$  choose  $\beta$  to minimize:

$$E \left[ (\tau - 1) \int_{-\infty}^{x'\beta} (y - x'\beta) dF(y|x) + \tau \int_{x'\beta}^{\infty} (y - x'\beta) dF(y|x) \right]$$

(Note  $y \leq x'\beta$  in the first integral and  $y \geq x'\beta$  in the second.)

- The FOC for the solution  $\beta_\tau$  is:

$$E \left[ (1 - \tau) \int_{-\infty}^{x'\beta_\tau} dF(y|x) = \tau \int_{x'\beta_\tau}^{\infty} dF(y|x) \right]$$

and a sample analogue,  $\beta_\tau^{(N)}$ , minimizes:

$$\frac{1}{N} \sum_{n=1}^N [(\tau - 1) I \{y_n \leq x_n'\beta\} + \tau I \{y_n > x_n'\beta\}] (y_n - x_n'\beta)$$

- Setting  $\tau = 0.5$ , the median, defines the LAD estimator.

# Ordinary Least Squares

## Estimation error

- Substituting (1) into (3) yields:

$$\begin{aligned}\beta_{OLS}^{(N)} &= \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left[ \frac{1}{N} \sum_{n=1}^N x_n (x_n' \beta_0 + \epsilon_n) \right] \\ &= \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right) \beta_0 \\ &\quad + \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N x_n \epsilon_n \right) \\ &= \beta_0 + \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N x_n \epsilon_n \right)\end{aligned}$$

- Thus the estimation error is:

$$\delta_{OLS}^{(N)} \equiv \beta_{OLS}^{(N)} - \beta_0 = \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N x_n \epsilon_n \right) \quad (6)$$

# Ordinary Least Squares

An orthogonality condition assumption

- Denote  $\mathbf{x}^{(N)} \equiv (x_1, \dots, x_N)$  and assume  $E[\epsilon_n | \mathbf{x}^{(N)}] = 0$ .
- Then  $E[\delta_{OLS}^{(N)} | \mathbf{x}^{(N)}] = 0$ , and  $\beta_{OLS}^{(N)}$  is *unbiased*, meaning:

$$E[\beta_{OLS}^{(N)} | \mathbf{x}^{(N)}] = \beta_0$$

- When  $E[\epsilon_n | \mathbf{x}^{(N)}] = 0$  the variance of  $\beta_{OLS}^{(N)}$  is:

$$\begin{aligned} & E \left\{ \left[ \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N x_n \epsilon_n \right) \right] \right. \\ & \quad \left. \times \left[ \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N x_n \epsilon_n \right) \right]' \middle| \mathbf{x}^{(N)} \right\} \quad (7) \\ & = E \left[ \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \times \right. \\ & \quad \left. \left( \frac{1}{N^2} \sum_{n=1}^N \sum_{m=1}^N x_n \epsilon_n \epsilon_m x_m' \right) \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \middle| \mathbf{x}^{(N)} \right] \end{aligned}$$

# Ordinary Least Squares

A further specialization

- Suppose it is also true that:

$$E \left[ \epsilon_n \epsilon_m \mid x^{(N)} \right] = \begin{cases} \sigma^2 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases} \quad (8)$$

- Then (7) simplifies to:

$$\begin{aligned} & E \left[ \delta_{OLS}^{(N)} \delta_{OLS}^{(N)'} \mid x^{(N)} \right] \\ &= \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \times \\ &= \left( \frac{1}{N^2} \sum_{n=1}^N \sum_{m=1}^N x_n E \left[ \epsilon_n \epsilon_m \mid x_n, x_m \right] x_m' \mid x^{(N)} \right) \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \\ &= \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \left( \frac{\sigma^2}{N^2} \sum_{n=1}^N x_n x_n' \right) \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \\ &= \frac{\sigma^2}{N} \left\{ \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} \right\} \end{aligned}$$

# Generalized Least Squares

## A transformation

- Assume  $E \left[ \epsilon_n \mid x^{(N)} \right] = 0$  for all  $n \in \{1, \dots, N\}$ .
- Let  $\epsilon^{(N)} \equiv (\epsilon_1, \dots, \epsilon_N)'$  denote the vector of unobserved variables.
- Denote their covariance matrix by  $\Psi \equiv E \left[ \epsilon^{(N)} \epsilon^{(N)'} \mid x^{(N)} \right]$ .
- Since  $\Psi$  is positive definite,  $\Psi^{-1/2}$  exists and satisfies:

$$\Psi^{-1} = \Psi^{-1/2} \Psi^{-1/2}$$

- Stack the individual equations and premultiply the resulting matrix equation by  $\Psi^{-1/2}$  to obtain a transformation of (1):

$$y_n^* = x_n^{*'} \beta + \epsilon_n^* \quad (9)$$

where:

$$\begin{aligned} (y_1^*, \dots, y_N^*)' &\equiv \Psi^{-1/2} (y_1, \dots, y_N)' \\ (\epsilon_1^*, \dots, \epsilon_N^*)' &\equiv \Psi^{-1/2} (\epsilon_1, \dots, \epsilon_N)' \\ (x_1^*, \dots, x_N^*) &\equiv (x_1, \dots, x_N) \Psi^{-1/2} \end{aligned}$$

# Generalized Least Squares

## Definition

- We define the *generalized least squares* (GLS) estimator by:

$$\begin{aligned}\beta_{GLS}^{(N)} &\equiv \arg \min_{\beta} \left\{ \sum_{n=1}^N (y_n^* - x_n^{*'} \beta)^2 \right\} \\ &= \left( \frac{1}{N} \sum_{n=1}^N x_n^* x_n^{*'} \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N x_n^* y_n^* \right)\end{aligned}$$

- The assumptions in the previous slide imply:

$$\begin{aligned}E \left[ \epsilon_n^* \mid x^{(N)} \right] &= 0 \\ E \left[ \epsilon_n^* \epsilon_m^* \mid x^{(N)} \right] &= \begin{cases} 1 & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases}\end{aligned}$$

- Thus  $\beta_{GLS}^{(N)}$  is unbiased.

# Generalized Least Squares

A random effects estimator for panel data

- Returning to the model of panel data  $\{y_{nt}, x_{nt}\}_{n=1}^{NT}$  with specification (2) we briefly consider the following two estimators.
- The first defines:

$$\hat{\epsilon}_{nt} \equiv \gamma_n + \epsilon_{nt}$$

and treats the equation be estimated as:

$$y_{nt} = x'_{nt}\beta_0 + \hat{\epsilon}_{nt} \quad (10)$$

- A *random effects* estimator (RE) is to conduct OLS or GLS on (10).
- Without loss of generality  $E[\epsilon_{nt} | \gamma_n] = 0$ . The RE estimator is unbiased if:

$$E[\epsilon_{nt} | x^{(N)}] = E[\gamma_n | x^{(N)}] = 0$$



# Generalized Least Squares

A first-difference estimator for panel data

- Alternatively apply the *difference operator* to (2) and obtain:

$$\Delta y_{nt} = \Delta x'_{nt} \beta_0 + \Delta \epsilon_{nt} \quad (11)$$

where:

$$\Delta y_{nt} \equiv y_{n,t+1} - y_{nt}$$

$$\Delta x_{nt} \equiv x_{n,t+1} - x_{nt}$$

$$\Delta \epsilon_{nt} \equiv \epsilon_{n,t+1} - \epsilon_{nt}$$

- Then using (11) estimate  $\beta_0$  from  $\{y_{nt}, x_{nt}\}_{n=1}^{N,T-1}$  with OLS or GLS.
- The FD estimator is unbiased if:

$$E \left[ \epsilon_{nt} \mid x^{(N)} \right] = 0$$

but correlations between  $x_{nt}$  and  $\gamma_n$  do not affect the properties of this estimator.

# Generalized Least Squares

Constructing the covariance matrices for these two GLS estimators

- Without loss of generality  $E[\epsilon_{nt} | \gamma_n] = 0$  and hence:

$$E[\epsilon_{nt} | \gamma_n] = 0 \Rightarrow E[\epsilon_{nt} \gamma_n] = 0$$

- For now we also assume:

- $E[\epsilon_{nt} \epsilon_{ms}] = 0$  for all  $m \neq n$  and all  $(s, t)$
- $E[\epsilon_{nt} \epsilon_{ns}] = 0$  for all  $s \neq t$
- $E[\epsilon_{nt}^2] = \sigma_\epsilon^2$

- If  $E[\gamma_n^2] = \sigma_\gamma^2$  and  $E[\gamma_n \gamma_m] = 0$  for all  $m \neq n$ , then the nonzero elements of  $\Psi_{RE}$  are:

$$E[\widehat{\epsilon}_{nt} \widehat{\epsilon}_{ns}] = \begin{cases} \sigma_\epsilon^2 + \sigma_\gamma^2 & \text{if } s = t \\ \sigma_\gamma^2 & \text{if } s \neq t \end{cases}$$

- By way of contrast the only nonzero elements of  $\Psi_{FD}$  are:

$$E[\Delta \epsilon_{nt} \Delta \epsilon_{ns}] = \begin{cases} 2\sigma_\epsilon^2 & \text{if } s = t \\ -\sigma_\epsilon^2 & \text{if } s = t + 1 \end{cases}$$

# Linear Instrumental Variables

## Motivation

- Rearranging the FOC for the quadratic defining OLS gives:

$$0 = \sum_{n=1}^N x_n \left( y_n - x_n' \beta_{OLS}^{(N)} \right)$$

- As a matter of computation  $\beta_{OLS}^{(N)}$  is obtained by:
  - premultiplying  $\left( y_n - x_n' \beta_{OLS}^{(N)} \right)$  by  $x_n$
  - solving the resulting  $k$  equations in  $k$  unknowns.
- Moreover its unbiasedness stems from the assumption that:

$$0 = E \left[ \epsilon_n \mid x^{(N)} \right] = E \left[ y_n - x_n' \beta_0 \mid x^{(N)} \right]$$

- Instead of premultiplying  $\left( y_n - x_n' \beta_{OLS}^{(N)} \right)$  by  $x_n$  we could premultiply  $\left( y_n - x_n' \beta_{OLS}^{(N)} \right)$  by  $z_n \equiv A w_n$  for some  $k \times l$  matrix  $A$  and some  $l \times 1$  instrument vector, where  $l > k$ , and base the estimator on a different set of equations.

# Linear Instrumental Variables

## Definition

- Accordingly define an instrumental variables (IV) estimator by:

$$0 = \sum_{n=1}^N z_n \left( y_n - x_n' \beta_{IV}^{(N)} \right)$$

- If  $\frac{1}{N} \sum_{n=1}^N z_n x_n'$  is invertible (has a nonzero determinant), then similar to above:

$$\beta_{IV}^{(N)} = \left( \frac{1}{N} \sum_{n=1}^N z_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N z_n y_n \right)$$

- To investigate the finite sample properties of  $\beta_{IV}^{(N)}$  we follow the same reasoning we applied to  $\beta_{OLS}^{(N)}$  by substituting for  $y_n$  to obtain:

$$\begin{aligned} \beta_{IV}^{(N)} &= \left( \frac{1}{N} \sum_{n=1}^N z_n x_n' \right)^{-1} \frac{1}{N} \sum_{n=1}^N z_n (x_n' \beta_0 + \epsilon_n) \\ &= \beta_0 + \left( \frac{1}{N} \sum_{n=1}^N z_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N z_n \epsilon_n \right) \end{aligned}$$

# Linear Instrumental Variables

Conditions for the existence of an unbiased estimator

- In this case the estimation error is:

$$\delta_{IV}^{(N)} \equiv \beta_{IV}^{(N)} - \beta_0 = \left( \frac{1}{N} \sum_{n=1}^N z_n x_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N z_n \epsilon_n \right) \quad (12)$$

- Let  $v^{(N)} \equiv (x^{(N)}, w^{(N)})$ . If  $E[\epsilon_n | v^{(N)}] = 0$  then  $E[\delta_{IV}^{(N)} | v^{(N)}] = 0$  and  $\beta_{IV}^{(N)}$  is unbiased, and (as we show on the next slides):

$$E[\delta_{IV}^{(N)} \delta_{IV}^{(N)'} | v^{(N)}] = \frac{1}{N} Y^{(N)} E[\Omega^{(N)} | v^{(N)}] Y^{(N)'}$$

where:

$$Y^{(N)} \equiv \left( \frac{1}{N} \sum_{n=1}^N z_n x_n' \right)^{-1}$$

$$\Omega^{(N)} \equiv \frac{1}{N} \sum_{n=1}^N z_n z_n' \epsilon_n^2 + \frac{1}{N} \sum_{s=2}^N \sum_{n=1}^{s-1} (z_n z_{n+s}' + z_{n+s} z_n') \epsilon_n \epsilon_{n+s}$$

# Linear Instrumental Variables

## Parsing the covariance

- From (12):

$$\begin{aligned} & E \left[ \delta_{IV}^{(N)} \delta_{IV}^{(N)'} \mid \mathbf{v}^{(N)} \right] \\ &= E \left\{ \left( \frac{1}{N} \sum_{n=1}^N \mathbf{z}_n \mathbf{x}_n' \right)^{-1} \left( \frac{1}{N} \sum_{n=1}^N \mathbf{z}_n \epsilon_n \right) \right. \\ & \quad \left. \times \left( \frac{1}{N} \sum_{n=1}^N \mathbf{z}_m \epsilon_m \right)' \left( \frac{1}{N} \sum_{n=1}^N \mathbf{z}_n \mathbf{x}_n' \right)^{-1'} \mid \mathbf{v}^{(N)} \right\} \\ &= \mathbf{Y}^{(N)} E \left\{ \left( \frac{1}{N} \sum_{n=1}^N \mathbf{z}_n \epsilon_n \right) \left( \frac{1}{N} \sum_{m=1}^N \mathbf{z}_m \epsilon_m \right)' \mid \mathbf{v}^{(N)} \right\} \mathbf{Y}^{(N)'} \end{aligned}$$

# Linear Instrumental Variables

## Parsing the covariance (continued)

- Focusing on the middle terms involving  $\epsilon_n$  and  $\epsilon_m$ :

$$\begin{aligned} & \left( \sum_{n=1}^N z_n \epsilon_n \right) \left( \sum_{m=1}^N z_m \epsilon_m \right)' \\ &= \sum_{n=1}^N \sum_{m=1}^N z_n \epsilon_n \epsilon_m z_m' \\ &= \sum_{n=1}^N z_n \epsilon_n^2 z_n' + \sum_{s=2}^N \sum_{n=1}^{s-1} (z_n z_{n+s}' + z_{n+s} z_n') \epsilon_n \epsilon_{n+s} \end{aligned}$$

- The last line comes from visualizing the matrix of terms:

$$\begin{bmatrix} z_1 \epsilon_1^2 z_1' & \cdots & z_1 \epsilon_1 \epsilon_N z_N' \\ \vdots & \ddots & \vdots \\ z_N \epsilon_N \epsilon_1 z_1' & \cdots & z_N \epsilon_N^2 z_N' \end{bmatrix}$$

- Substituting the expression above back into the formula for the variance gives the result.

# Constrained Least Squares

## Definition and Solution

- Now suppose we have information about the unknown parameter vector  $\beta_0$  that takes the form of a linear constraint,  $q$  equations in  $\beta_0$ :

$$Q\beta_0 = c \quad (13)$$

where as before  $\beta_0$  is a  $k \times 1$  vector and:

- $Q$  is a  $q \times k$  matrix with  $q < k$ ,
  - $c$  is a  $q \times 1$  vector.
- The *constrained least squares* (CLS) estimator is defined by:

$$\begin{aligned} \beta_{CLS}^{(N)} &\equiv \arg \min_{\beta} \left\{ \sum_{n=1}^N (y_n - x_n' \beta)^2 \text{ such that } Q\beta = c \right\} \\ &= \beta_{OLS}^{(N)} - \left[ \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} Q' \right] \times \\ &\quad \left[ Q \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} Q' \right]^{-1} \left( Q\beta_{OLS}^{(N)} - c \right) \end{aligned}$$



# Constrained Least Squares

## Proof of formula for CLS

- Define:

$$\eta \equiv \beta_{CLS}^{(N)} - \beta_{OLS}^{(N)} \quad \gamma \equiv c - Q\beta_{OLS}^{(N)}$$

- From the constraint:

$$0 = Q\beta_{CLS}^{(N)} - c = Q(\beta_{CLS}^{(N)} - \beta_{OLS}^{(N)}) - c + Q\beta_{OLS}^{(N)} = Q\eta - \gamma \quad (14)$$

- The Lagrangian for the optimization problem can be written as:

$$\sum_{n=1}^N (y_n - x_n'\beta)^2 + \lambda(Q\beta - c)$$

and has FOC:

$$\begin{aligned} 0 &= -\left(\frac{2}{N} \sum_{n=1}^N x_n y_n\right) + \left(\frac{2}{N} \sum_{n=1}^N x_n x_n'\right) \beta_{CLS}^{(N)} + Q\lambda \\ &= \left(\frac{2}{N} \sum_{n=1}^N x_n x_n'\right) (\beta_{CLS}^{(N)} - \beta_{OLS}^{(N)}) + Q\lambda \\ &= \left(\frac{2}{N} \sum_{n=1}^N x_n x_n'\right) \eta + Q\lambda \end{aligned} \quad (15)$$

# Constrained Least Squares

## Proof of formula for CLS continued

- From (14) and (15):

$$\gamma = Q\eta \quad \eta = - \left( \frac{2}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} Q' \lambda$$

- Solving for  $\lambda$  in terms of  $\gamma$ :

$$Q\eta = -Q \left( \frac{2}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} Q' \lambda = \gamma$$

and hence:

$$\begin{aligned} \lambda &= - \left[ Q \left( \frac{2}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} Q' \right]^{-1} \gamma \\ \Rightarrow \eta &= \left[ \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} Q' \right] \left[ Q \left( \frac{1}{N} \sum_{n=1}^N x_n x_n' \right)^{-1} Q' \right]^{-1} \gamma \end{aligned}$$

- Using the definitions of  $\eta$  and  $\gamma$  the formula now follows directly.

# Specification Error versus Efficiency

Trading off efficiency with specification error

- Even if  $E[\epsilon_n | x_n] \neq 0$  and hence  $\beta_{OLS}^{(N)}$  is biased, an unbiased estimator  $\beta_{IV}^{(N)}$  can be obtained if there exists some  $z_n$  satisfying:
  - 1 the invertibility assumption for  $\frac{1}{N} \sum_{n=1}^N z_n x_n'$
  - 2 the orthogonality condition  $E[\epsilon_n | z_n] = 0$ .
- This raises the question of why OLS is ever used instead of IV, since the latter seems less restrictive.
- In Assignment 3 you are asked to show that:

$$E \left[ \delta_{OLS}^{(N)} \delta_{OLS}^{(N)'} \right] \leq E \left[ \delta_{IV}^{(N)} \delta_{IV}^{(N)'} \right]$$

- Similarly one can show that:

$$E \left[ \delta_{CLS}^{(N)} \delta_{CLS}^{(N)'} \right] \leq E \left[ \delta_{OLS}^{(N)} \delta_{OLS}^{(N)'} \right]$$

- Comparing the FE and the RE estimators raises similar issues. The former is based on  $N(T-1)$  observations, but the latter requires  $E[\gamma_n | x_{nt}] = 0$  for unbiasedness.

# Specification Error versus Efficiency

## Mean square error

- The *mean square error* (MSE) is one way to evaluate the trade-off between bias and variance.
- Let  $\theta \equiv \sum_{k=0}^{K-1} a_k \beta_k$  be a known linear combination of  $\beta$  defined by  $a \equiv (a_0, \dots, a_{K-1}) \in \mathbb{R}^K$ .
- For any estimator  $\theta^{(N)}$  of  $\theta_0$  we define the MSE as:

$$\begin{aligned} \text{MSE} \left( \theta^{(N)} \right) &\equiv E \left[ \left( \theta^{(N)} - \theta_0 \right)^2 \right] \\ &= E \left[ \left( \theta^{(N)} - E \left[ \theta^{(N)} \right] + E \left[ \theta^{(N)} \right] - \theta_0 \right)^2 \right] \\ &= E \left[ \left\{ \theta^{(N)} - E \left[ \theta^{(N)} \right] \right\}^2 + \left\{ E \left[ \theta^{(N)} \right] - \theta_0 \right\}^2 \right. \\ &\quad \left. + 2 \left\{ \theta^{(N)} - E \left[ \theta^{(N)} \right] \right\} \left\{ E \left[ \theta^{(N)} \right] - \theta_0 \right\} \right] \\ &= E \left[ \left( \left\{ \theta^{(N)} - E \left[ \theta^{(N)} \right] \right\}^2 + \left\{ E \left[ \theta^{(N)} \right] - \theta_0 \right\}^2 \right) \right] \end{aligned}$$