

## Week 8: Generalized Method of Moments

ResEcon 703: Topics in Advanced Econometrics

Matt Woerman  
University of Massachusetts Amherst

# Agenda

## Last week

- Logit estimation using maximum likelihood

## This week's topics

- Method of moments overview
- Method of moments estimator
- Method of moments examples
- GMM estimator
- GMM example
- Properties of the GMM estimator
- Optimal GMM estimator
- Specification and hypothesis tests
- Logit model with GMM
- GMM R example

## This week's reading

- Generalized method of moments supplement

# Method of Moments Overview

# (Generalized) Method of Moments

Method of moments (MM) estimation is a special case of generalized method of moments (GMM) estimation

- Common in modern empirical economics
  - ▶ Most common in industrial organization, macroeconomics, and finance
- Less parametric than maximum likelihood
  - ▶ Requires assumptions about moment conditions, not assumptions about the full distribution of data
- Most of the estimators you have learned so far are (G)MM estimators
  - ▶ OLS, 2SLS, GLS, MLE, etc.

Overview of (generalized) method of moments estimation

- MM and GMM use population moment conditions to estimate parameters
  - ▶ Moments can come from economic models, statistical assumptions, etc.
- MM and GMM estimators “solve” the sample moment conditions that are analogous to these population moment conditions

# Moment Conditions

Moment conditions are functions of parameters and data that equal zero in expectation when evaluated at the true parameter values

$$\text{Mean: } \mu_y = E[y] \quad \Rightarrow \quad E[y - \mu_y] = 0$$

$$\text{Var: } \sigma_y^2 = E[(y - \mu_y)^2] \quad \Rightarrow \quad E[(y - \mu_y)^2 - \sigma_y^2] = 0$$

$$\text{Cov: } \sigma_{xy} = E[(y - \mu_y)(x - \mu_x)] \quad \Rightarrow \quad E[(y - \mu_y)(x - \mu_x) - \sigma_{xy}] = 0$$

How do we normally generate moment conditions?

- Economic models
  - ▶ Example: first-order conditions
- Econometric or statistical model
  - ▶ Example: instruments must be uncorrelated with errors
- Model fit
  - ▶ Example: predicted market shares must equal realized market shares

## Method of Moments Intuition

Suppose we have five random draws, but we do not know the distribution

$$\mathbf{y} = \{5, 10, 9, 14, 7\}$$

Suppose the mean of the unknown distribution is given by  $\mu$ , which gives us a population moment condition

$$E[y_i] = \mu \Rightarrow E[y_i - \mu] = 0$$

If this moment condition holds in the population, then we should expect an analogous condition to hold in our sample

$$\frac{1}{n} \sum_{i=1}^n (y_i - \mu) = 0$$

We can estimate the population parameter,  $\hat{\mu}$ , by finding the value that makes this empirical expression true in our sample

# Method of Moments Estimator

# Method of Moments Assumptions

Data,  $\{y, \mathbf{x}, \mathbf{z}\}$ , are drawn from a population with  $K$  moment conditions that are functions of  $K$  parameters,  $\theta$

$$E[\mathbf{m}(y, \mathbf{x}, \mathbf{z}, \theta)] = \mathbf{0}$$

- $\mathbf{m}(\cdot)$ : vector of  $K$  functions
- $\theta$ : vector of  $K$  parameters
- $y$ : dependent variable
- $\mathbf{x}$ : vector of independent variables
- $\mathbf{z}$ : vector of exogenous instruments

We make assumptions about the moment conditions of the population from which the data are drawn, but we do not have to make an assumption about the full underlying distribution



# Sample Moments

These moment conditions hold for the population from which the data are drawn

$$E[\mathbf{m}(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta})] = \mathbf{0}$$

But we do not observe the full population, only a sample

We replace the population expectation with its empirical analog—the sample mean—to generate the corresponding  $K$  sample moments as a function of the  $K$  parameters

$$\frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}) = \mathbf{0}$$

## Method of Moments Estimator

The method of moments estimators is the set of parameters,  $\hat{\theta}$ , that solves the system of equation given by the sample moments

$$\frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \hat{\theta}) = \mathbf{0}$$

In some case,  $\hat{\theta}$  will have a closed-form expression, but in other cases we can reformulate  $\hat{\theta}$  as the solution to a minimization problem and use numerical optimization

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} Q(\theta)$$

where  $Q(\theta)$  is given by

$$Q(\theta) = \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta) \right]' \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta) \right]$$

## Method of Moments Examples

## Method of Moments Population Mean Example

We have five data points from a population, but what is the mean,  $\mu$ , of this population?

$$\mathbf{y} = \{5, 10, 9, 14, 7\}$$

If  $y$  is i.i.d. with mean  $\mu$ , then a population moment condition is

$$E[y - \mu] = 0$$

Replacing the expectation with the sample analog gives

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}) = 0$$

Solving this one equation for the one parameter yields

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i = 9$$

# Method of Moments OLS Regression Example

The OLS regression estimator is a special case of a MM estimator

Consider the general OLS regression model

$$y_i = \beta' \mathbf{x}_i + \varepsilon_i$$

- $\beta$ : vector of  $K$  parameters
- $\mathbf{x}_i$ : vector of  $K$  variables

We assume the error term,  $\varepsilon_i$ , is orthogonal to the data,  $\mathbf{x}_i$

$$E[\mathbf{x}_i \varepsilon_i] = 0$$

Replacing  $\varepsilon_i$  with  $y_i - \beta' \mathbf{x}_i$  gives  $K$  population moment conditions

$$E[\mathbf{x}_i (y_i - \beta' \mathbf{x}_i)] = \mathbf{0}$$

# Method of Moments OLS Regression Example

We have  $K$  moments that are functions of the  $K$  parameters,  $\beta$

$$E[\mathbf{x}_i(y_i - \beta' \mathbf{x}_i)] = \mathbf{0}$$

We replace these  $K$  expectations with the  $K$  sample analogs

$$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i(y_i - \hat{\beta}' \mathbf{x}_i) = \mathbf{0}$$

Solving these  $K$  equations for the  $K$  parameters yields the method of moments estimator

$$\hat{\beta} = \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left( \sum_{i=1}^n \mathbf{x}_i y_i \right)$$

This method of moments estimator is equivalent to the traditional OLS estimator

## Method of Moments Maximum Likelihood Example

The MLE can also be motivated as a special case of a MM estimator

Consider a random variable  $y$  with conditional density

$$f(y_i | \mathbf{x}_i, \boldsymbol{\theta})$$

Then the log-likelihood function of  $\boldsymbol{\theta}$  conditional on  $\mathbf{y}$  and  $\mathbf{X}$  is

$$\ln L(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X}) = \sum_{i=1}^n \ln f(y_i | \mathbf{x}_i, \boldsymbol{\theta})$$

Maximizing this log-likelihood function yields  $K$  first-order conditions, one for each of the  $K$  parameters

$$\mathbf{0} = \frac{\partial \ln L(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \frac{\partial \ln f(y_i | \mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

## Method of Moments Maximum Likelihood Example

From these  $K$  first-order conditions, we have that

$$\sum_{i=1}^n \frac{\partial \ln f(y_i | \mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

Dividing by  $n$  yields an expression that looks like sample moments

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \ln f(y_i | \mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

These  $K$  sample moments can be viewed as the empirical analogs of  $K$  population moment conditions

$$E \left[ \frac{\partial \ln f(y | \mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] = \mathbf{0}$$

MLE can be motivated as a MM estimator with these population moments



# GMM Estimator

# Generalized Method of Moments Assumptions

Data,  $\{y, \mathbf{x}, \mathbf{z}\}$ , are drawn from a population with  $L$  moment conditions that are functions of  $K$  parameters,  $\boldsymbol{\theta}$ , with  $L \geq K$

$$E[\mathbf{m}(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta})] = \mathbf{0}$$

- $\mathbf{m}(\cdot)$ : vector of  $L$  functions
- $\boldsymbol{\theta}$ : vector of  $K$  parameters
- $y$ : dependent variable
- $\mathbf{x}$ : vector of independent variables
- $\mathbf{z}$ : vector of exogenous instruments

We make assumptions about the moment conditions of the population from which the data are drawn, but we do not have to make an assumption about the full underlying distribution

# Source of Moment Conditions

Generalized method of moments allows for more moment conditions than the number of parameters

- Method of moments requires the number of moment conditions to equal the number of parameters

Additional moment conditions can come from many sources

- Economic model: Demand parameter appears in demand and supply first-order conditions
- Instruments: Multiple instruments for one endogenous variable
- Model fit: Predicted market shares equal realized market shares
- Statistical assumptions: Errors are orthogonal to data

# Sample Moments

These  $L$  moment conditions hold for the population from which the data are drawn

$$E[\mathbf{m}(y, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta})] = \mathbf{0}$$

But we do not observe the full population, only a sample

We replace the population expectation with its empirical analog—the sample mean—to generate the corresponding  $L$  sample moments as a function of the  $K$  parameters

$$\frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}) = \mathbf{0}$$

# Generalized Method of Moments Estimator

With only  $K$  parameters, we cannot solve for  $L$  unique sample moments when  $L > K$

- Instead, we will seek to simultaneously get as close as possible to solving all  $L$  sample moments

The generalized method of moments estimators is the set of parameters,  $\hat{\theta}$ , that minimizes the weighted sum of squared sample moments

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} Q(\theta)$$

where  $Q(\theta)$  is given by

$$Q(\theta) = \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta) \right]' \mathbf{W} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta) \right]$$

and  $\mathbf{W}$  is a  $L \times L$  positive definite weighting matrix

## GMM Example

# Generalized Method of Moments 2SLS Regression Example

The 2SLS regression estimator is a special case of a GMM estimator

Consider the linear regression model

$$y_i = \beta' \mathbf{x}_i + \varepsilon_i$$

- $\beta$ : vector of  $K$  parameters
- $\mathbf{x}_i$ : vector of  $K$  variables, but some are endogenous

We have a vector of  $L$  instruments,  $\mathbf{z}_i$ , that are correlated with  $\mathbf{x}_i$  but are orthogonal to the error term,  $\varepsilon_i$

$$E[\mathbf{z}_i \varepsilon_i] = 0$$

Replacing  $\varepsilon_i$  with  $y_i - \beta' \mathbf{x}_i$  gives  $L$  population moment conditions

$$E[\mathbf{z}_i (y_i - \beta' \mathbf{x}_i)] = \mathbf{0}$$

# Generalized Method of Moments 2SLS Regression Example

We have  $L$  moments that are functions of the  $K$  parameters,  $\beta$

$$E[\mathbf{z}_i(y_i - \beta' \mathbf{x}_i)] = \mathbf{0}$$

We replace these  $L$  expectations with the  $L$  sample analogs

$$\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i(y_i - \beta' \mathbf{x}_i) = \mathbf{0}$$

We cannot solve all  $L$  sample moment conditions with only  $K$  parameters (if  $L > K$ ), so instead we find the set of parameters,  $\hat{\beta}$ , that minimizes the weighted sum of squared sample moments

$$Q(\beta) = \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i(y_i - \beta' \mathbf{x}_i) \right]' \mathbf{W} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i(y_i - \beta' \mathbf{x}_i) \right]$$



# Generalized Method of Moments 2SLS Regression Example

The GMM estimator for this instrumental variables regression,  $\hat{\beta}$ , is

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} Q(\beta)$$

where  $Q(\beta)$  is given by

$$Q(\beta) = \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i (y_i - \hat{\beta}' \mathbf{x}_i) \right]' \mathbf{W} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i (y_i - \hat{\beta}' \mathbf{x}_i) \right]$$

The GMM estimator will be equivalent to the 2SLS estimator if we use a weighting matrix,  $\mathbf{W}$ , of

$$\mathbf{W} = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i'$$

But another weighting matrix might yield a “better” estimator

## Properties of the GMM Estimator

# Asymptotic Properties of GMM Estimator

If some assumptions about the empirical moments and the weighting matrix hold, the generalized method of moments estimator has these properties

- 1 Consistency
- 2 Asymptotic normality

Note that the GMM estimator is consistent and asymptotically normal, but it is not efficient, in contrast to the MLE

- The distributional assumption of maximum likelihood provides additional information that makes the MLE efficient

# GMM Empirical Moment Assumptions

In these assumptions and properties, the 0 subscript denotes an object evaluated at the true parameter values,  $\theta_0$

- 1 The empirical moments obey the law of large numbers

$$\frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta_0) \xrightarrow{p} \mathbf{0}$$

- 2 The derivatives of the empirical moments converge

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta)}{\partial \theta'} \bigg|_{\theta=\theta_0} \xrightarrow{p} \mathbf{G}_0$$

# GMM Empirical Moment Assumptions

- 3 The empirical moments obey the central limit theorem

$$\frac{\sqrt{n}}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \mathbf{S}_0)$$

where

$$\mathbf{S}_0 = \text{plim} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0) \mathbf{m}(y_j, \mathbf{x}_j, \mathbf{z}_j, \boldsymbol{\theta}_0)'$$

- 4 The parameters are identified

$$\frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_1) = \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_2) \Leftrightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$$

# GMM Weighting Matrix Assumption

- 5 The weighting matrix converges to a finite symmetric positive definite matrix

$$\mathbf{W} \xrightarrow{P} \mathbf{W}_0$$

# Consistency of the GMM Estimator

The GMM estimator,  $\hat{\theta}$ , is a consistent estimator of the true parameter values,  $\theta_0$

$$\hat{\theta} \xrightarrow{P} \theta_0$$

- As our sample size grows (to infinity), the GMM estimator becomes vanishingly close to the true parameter value(s)

# Asymptotic Normality of the GMM Estimator

The GMM estimator,  $\hat{\theta}$ , is asymptotically normal with mean  $\theta_0$  and known variance

$$\hat{\theta} \overset{a}{\sim} \mathcal{N} \left( \theta_0, \frac{1}{n} (\mathbf{G}_0' \mathbf{W}_0 \mathbf{G}_0)^{-1} (\mathbf{G}_0' \mathbf{W}_0 \mathbf{S}_0 \mathbf{W}_0 \mathbf{G}_0) (\mathbf{G}_0' \mathbf{W}_0 \mathbf{G}_0)^{-1} \right)$$

where

$$\mathbf{G}_0 = \text{plim} \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

$$\mathbf{W}_0 = \text{plim} \mathbf{W}$$

$$\mathbf{S}_0 = \text{plim} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}_0) \mathbf{m}(y_j, \mathbf{x}_j, \mathbf{z}_j, \boldsymbol{\theta}_0)'$$



## Optimal GMM Estimator

# GMM Variance-Covariance Matrix

Every weighting matrix yields a GMM estimator that is consistent

- But the choice of weighting matrix,  $\mathbf{W}$ , will affect the variance of the GMM estimator,  $\hat{\theta}$ , because its probability limit,  $\mathbf{W}_0$ , enters the calculation of the variance-covariance matrix

$$\text{Var}(\hat{\theta}) = \frac{1}{n}(\mathbf{G}'_0 \mathbf{W}_0 \mathbf{G}_0)^{-1}(\mathbf{G}'_0 \mathbf{W}_0 \mathbf{S}_0 \mathbf{W}_0 \mathbf{G}_0)(\mathbf{G}'_0 \mathbf{W}_0 \mathbf{G}_0)^{-1}$$

If every weighting matrix yields a GMM estimator that is consistent, then we would ideally like to use the weighting matrix that minimizes the variance of the estimator

## Optimal Weighting Matrix

$$\text{Var}(\hat{\theta}) = \frac{1}{n}(\mathbf{G}'_0 \mathbf{W}_0 \mathbf{G}_0)^{-1}(\mathbf{G}'_0 \mathbf{W}_0 \mathbf{S}_0 \mathbf{W}_0 \mathbf{G}_0)(\mathbf{G}'_0 \mathbf{W}_0 \mathbf{G}_0)^{-1}$$

The variance of the GMM estimator is minimized when

$$\mathbf{W}_0 = \mathbf{S}_0^{-1} = \left[ \text{plim} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta_0) \mathbf{m}(y_j, \mathbf{x}_j, \mathbf{z}_j, \theta_0)' \right]^{-1}$$

which gives a variance-covariance matrix of

$$\text{Var}(\hat{\theta}) = \frac{1}{n}(\mathbf{G}'_0 \mathbf{S}_0^{-1} \mathbf{G}_0)^{-1}$$

$\mathbf{S}_0$  is the variance-covariance matrix of the empirical moments

- We minimize the variance of the GMM estimator by weighting each moment inversely to its variance

## Optimal (or Two-Step) GMM Estimator

We want to use a weighting matrix,  $\mathbf{W}$ , that converges in probability to the variance of the empirical moments,  $\mathbf{S}_0^{-1}$

- But we do not know  $\mathbf{S}_0^{-1}$  because we do not know the true parameter values,  $\theta_0$

We can use a two-step procedure, performing GMM estimation twice, to obtain this optimal GMM estimator

- ① Step 1: Perform GMM with any arbitrary weighting matrix and estimate the variance of the moments,  $\tilde{\mathbf{S}}$ , using this GMM estimator
- ② Step 2: Perform GMM again with  $\mathbf{W} = \tilde{\mathbf{S}}^{-1}$

This procedure is a generalization of the FGLS regression

We will use the `gmm()` function from the `gmm` package in R to find this optimal GMM estimator

# Optimal GMM Estimator: Step 1

Step 1: Perform GMM with any arbitrary weighting matrix to obtain the first-step GMM estimator,  $\tilde{\theta}$

- The GMM estimator is consistent for any weighting matrix, so we can use any arbitrary weighting matrix to obtain a consistent estimate of the true parameter values,  $\theta_0$
- The identity matrix—equally weighting each moment—is a simple one to use

Use this first-step GMM estimator,  $\tilde{\theta}$ , to construct an estimator of  $S_0$

- Also assume observations are independent to simplify this estimator

$$\tilde{S} = \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \tilde{\theta}) \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \tilde{\theta})'$$

Then use the inverse of this estimate,  $\tilde{S}$ , as the weighting matrix in the second step

## Optimal GMM Estimator: Step 2

Step 2: Perform GMM with weighting matrix  $\mathbf{W} = \tilde{\mathbf{S}}^{-1}$

Find the GMM estimator,  $\hat{\boldsymbol{\theta}}$ , that minimizes the objective function

$$Q(\boldsymbol{\theta}) = \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}) \right]' \tilde{\mathbf{S}}^{-1} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\theta}) \right]$$

This second-step GMM estimator,  $\hat{\boldsymbol{\theta}}$ , also gives a consistent estimate of the true parameter values,  $\boldsymbol{\theta}_0$ , and it has the minimum variance among all GMM estimators

- It is not efficient (when compared to MLE), but it is efficient among GMM estimators

This optimal GMM estimator has the variance-covariance matrix

$$\text{Var}(\hat{\boldsymbol{\theta}}) = \frac{1}{n} (\mathbf{G}_0' \mathbf{S}_0^{-1} \mathbf{G}_0)^{-1}$$

## Optimal GMM Variance Estimator

The variance-covariance matrix of the GMM estimator is evaluated at the true parameter values,  $\theta_0$ , which we do not know

- We construct an estimator of this variance-covariance matrix by evaluating at the optimal GMM estimator,  $\hat{\theta}$

$$\widehat{Var}(\hat{\theta}) = \frac{1}{n} \left( \hat{\mathbf{G}}' \hat{\mathbf{S}}^{-1} \hat{\mathbf{G}} \right)^{-1}$$

where the components of this matrix are estimated by

$$\begin{aligned} \hat{\mathbf{G}} &= \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta)}{\partial \theta'} \right|_{\theta=\hat{\theta}} \\ \hat{\mathbf{S}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \hat{\theta}) \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \hat{\theta})' \end{aligned}$$

# Specification and Hypothesis Tests



## Overidentifying Restrictions Test

When we have more moment conditions than parameters, the model is “overidentified” and we cannot ensure all moments equal zero simultaneously, but we can test if all moment conditions are sufficiently close to zero

$$H_0 : E[\mathbf{m}(y, \mathbf{x}, \mathbf{z}, \theta_0)] = \mathbf{0}$$

When  $\hat{\theta}$  is estimated by optimal GMM, the overidentifying restrictions test statistic is equal to the objective function evaluated at the optimal GMM estimator,  $Q(\hat{\theta})$

$$OIR = \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \hat{\theta}) \right]' \tilde{\mathbf{S}}^{-1} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{m}(y_i, \mathbf{x}_i, \mathbf{z}_i, \hat{\theta}) \right]$$

which is asymptotically distributed  $\chi^2$  with  $L - K$  degrees of freedom

$$OIR \stackrel{a}{\sim} \chi^2(L - K)$$

# Hypothesis Tests

We discussed three hypothesis testing procedures for use with maximum likelihood estimation

- Likelihood ratio test
- Wald test
- Lagrange multiplier test

Each of these test procedures has an analogous test using the results of GMM estimation

- The intuition of each test is the same as its MLE counterpart, but using the GMM objective function in place of the likelihood or log-likelihood function

Refer to an econometrics textbook for the details of these hypothesis tests

## Logit Model with GMM

# Logit Model First-Order Conditions

When estimating the parameters of the logit model by maximum likelihood, the first-order conditions for the MLE are

$$\frac{\partial \ln L(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{X})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

where the log-likelihood function is given by

$$\ln L(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{X}) = \sum_{n=1}^N \sum_{i=1}^J y_{ni} \ln P_{ni}(\mathbf{x}_n, \boldsymbol{\theta})$$

We can alternatively express these first-order conditions as sample moments conditions

- See the Train textbook for the proof of this equality

# Logit Model Moment Conditions

If we assume representative utility is linear,  $V_{ni} = \beta' \mathbf{x}_{ni}$ , then the logit MLE first-order conditions are equivalent to

$$\frac{1}{NJ} \sum_{n=1}^N \sum_{i=1}^J (y_{ni} - P_{ni}(\mathbf{x}_n, \beta)) \mathbf{x}_{ni} = \mathbf{0}$$

These sample moments can be viewed as the empirical analogs of the population moment conditions

$$E[(y_{ni} - P_{ni}(\mathbf{x}_n, \beta)) \mathbf{x}_{ni}] = \mathbf{0}$$

These population moment conditions state that the data,  $\mathbf{x}_n$ , are orthogonal to the model residuals,  $y_{ni} - P_{ni}(\mathbf{x}_{ni}, \beta)$

- Analogous to the data and errors being orthogonal in OLS regressions

# Logit Model with Method of Moments

If this assumption of exogeneity holds, then we can also estimate the parameters of the logit model by MM using the sample moments

$$\frac{1}{NJ} \sum_{n=1}^N \sum_{i=1}^J \left( y_{ni} - P_{ni}(\mathbf{x}_n, \hat{\beta}) \right) \mathbf{x}_{ni} = \mathbf{0}$$

- MM because we have the same number of moments and parameters

Because logit choice probabilities are nonlinear, this MM estimator does not have a closed-form solution

- Instead, we will use numerical optimization to find the set of parameters,  $\hat{\beta}$ , that minimizes the MM objective function

$$Q(\beta) = \left[ \frac{1}{NJ} \sum_{n=1}^N \sum_{i=1}^J (y_{ni} - P_{ni}(\mathbf{x}_n, \beta)) \mathbf{x}_{ni} \right]' \left[ \frac{1}{NJ} \sum_{n=1}^N \sum_{i=1}^J (y_{ni} - P_{ni}(\mathbf{x}_n, \beta)) \mathbf{x}_{ni} \right]$$

## Logit Model Moment Conditions with Instruments

If the assumption of exogeneity does not hold, but we instead have some exogenous instruments,  $\mathbf{Z}$ , that are correlated with our data,  $\mathbf{X}$ , then we can create a different set of sample moment conditions that must hold

Our exogenous instruments,  $\mathbf{Z}$ , are orthogonal to the model residuals

$$E[(y_{ni} - P_{ni}(\mathbf{x}_n, \beta)) \mathbf{z}_{ni}] = \mathbf{0}$$

Replacing the expectation with the sample analog gives

$$\frac{1}{NJ} \sum_{n=1}^N \sum_{i=1}^J (y_{ni} - P_{ni}(\mathbf{x}_n, \hat{\beta})) \mathbf{z}_{ni} = \mathbf{0}$$

Our instruments,  $\mathbf{Z}$ , can be any exogenous variables that are correlated with our data,  $\mathbf{X}$ , including the exogenous elements of  $\mathbf{X}$

- We must have at least as many instruments (and moments) as parameters in order to identify all parameters

## Logit Model with GMM

We can estimate the parameters of the logit model with instruments by GMM using the instrument sample moments

$$\frac{1}{NJ} \sum_{n=1}^N \sum_{i=1}^J \left( y_{ni} - P_{ni}(\mathbf{x}_n, \hat{\beta}) \right) \mathbf{z}_{ni} = \mathbf{0}$$

We will use the two-step procedure described earlier—based on this objective function—to find the optimal GMM estimator

$$Q(\beta) = \left[ \frac{1}{NJ} \sum_{n=1}^N \sum_{i=1}^J (y_{ni} - P_{ni}(\mathbf{x}_n, \beta)) \mathbf{z}_{ni} \right]' \mathbf{W} \left[ \frac{1}{NJ} \sum_{n=1}^N \sum_{i=1}^J (y_{ni} - P_{ni}(\mathbf{x}_n, \beta)) \mathbf{z}_{ni} \right]$$

- 1 Estimate  $\tilde{\beta}$  using any arbitrary weighting matrix in order to estimate  $\tilde{S}$  and find the optimal weighting matrix
- 2 Use that weighting matrix to find the optimal GMM estimator,  $\hat{\beta}$



## GMM R Example

## Binary Choice Example

We are studying how consumers make choices about expensive and highly energy-consuming appliances in their homes.

- We have (simulated) data on 600 households that rent apartments without air conditioning. These households must choose whether or not to purchase a window air conditioning unit. (To simplify things, we assume there is only one “representative” air conditioner for each household and its price and operating cost are exogenous.)
- We observe the following data about each household and its “representative” air conditioner
  - ▶ An indicator if they purchase the air conditioner (TRUE/FALSE)
  - ▶ The purchase price of the air conditioner (\$)
  - ▶ The annual operating cost of the air conditioner (\$ per year)
  - ▶ The household's electricity price (cents per kWh)
  - ▶ The size of the household's apartment (square feet)
  - ▶ The household's annual income (\$1000s)
  - ▶ The number of residents in the household (people)
  - ▶ An indicator for the household's city (1, 2, or 3)

## Random Utility Model for Air Conditioner Choice

We model the utility to household  $n$  of not purchasing an air conditioned ( $j = 0$ ) or purchasing an air conditioner ( $j = 1$ ) as

$$U_{n0} = V_{n0} + \varepsilon_{n0}$$

$$U_{n1} = V_{n1} + \varepsilon_{n1}$$

where  $V_{nj}$  depends on the data about alternative  $j$  and household  $n$

The probability that household  $n$  purchases an air conditioner is

$$P_{n1} = \Pr(\varepsilon_{n0} - \varepsilon_{n1} < V_{n1} - V_{n0})$$

- Only differences in utility—not the actual values of utility—affect this probability
- What is the difference in utility to household  $n$  from purchasing an air conditioner vs. not purchasing an air conditioner?

# Load Dataset

`read_csv()` is a tidyverse function to read a .csv file into a tibble

```
## Load tidyverse
library(tidyverse)
## Load dataset
ac_data <- read_csv('ac_renters.csv')
```

# Dataset

```
## Look at dataset
ac_data
## # A tibble: 600 x 8
##   air_conditioning cost_system cost_operating elec_price square_feet
##   <lgl>           <dbl>         <dbl>         <dbl>         <dbl>
## 1 FALSE          513          247          12.8          541
## 2 FALSE          578          138           9.6          384
## 3 TRUE           658          171          10.7          619
## 4 FALSE          615          198          11.5          624
## 5 FALSE          515          165          10.5          365
## 6 FALSE          588          143           9.7          411
## 7 TRUE           643          153          10.1          529
## 8 FALSE          676          182           11          694
## 9 TRUE           516          137           9.6          305
## 10 TRUE          544          185          11.1          454
## # ... with 590 more rows, and 3 more variables: income <dbl>,
## #   residents <dbl>, city <dbl>
```

## Representative Utility for Air Conditioner Choice

$$P_{n1} = \Pr(\varepsilon_{n0} - \varepsilon_{n1} < V_{n1} - V_{n0})$$

What is the difference in utility to household  $n$  from purchasing an air conditioner vs. not purchasing an air conditioner?

- They gain utility from having air conditioning
- They lose utility from paying the purchase price of the air conditioner
- They lose utility from paying the annual operating cost of the air conditioner

We can model the difference in utility as

$$V_{n1} - V_{n0} = \beta_0 + \beta_1 P_n + \beta_2 C_n$$

where

- $P_n$  is the purchase price of the air conditioner
- $C_n$  is the annual operating cost of the air conditioner
- $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  are utility parameters to be estimated

## Binary Logit Model of Air Conditioner Choice

Under the logit assumption, the choice probability of purchasing an air conditioner becomes

$$\begin{aligned} P_{n1} &= \frac{1}{1 + e^{-(V_{n1} - V_{n0})}} \\ &= \frac{1}{1 + e^{-(\beta_0 + \beta_1 P_n + \beta_2 C_n)}} \end{aligned}$$

When estimating the parameters of this model by maximum likelihood, the first-order conditions for the MLE are

$$\frac{\partial \ln L(\beta \mid \mathbf{y}, \mathbf{X})}{\partial \theta} = \mathbf{0}$$

where the log-likelihood function is given by

$$\ln L(\beta \mid \mathbf{y}, \mathbf{X}) = \sum_{n=1}^N \sum_{i=0}^1 y_{ni} \ln P_{ni}(\mathbf{x}_n, \beta)$$

# Moment Conditions for Air Conditioner Choice

These MLE first-order conditions are equivalent to sample moments

$$\frac{1}{2N} \sum_{n=1}^N \sum_{i=0}^1 (y_{ni} - P_{ni}(\mathbf{x}_n, \beta)) \mathbf{x}_n = \mathbf{0}$$

With a binary choice model, the model residuals for each alternative,  $y_{ni} - P_{ni}(\mathbf{x}_n, \beta)$ , will always be the negative of one another

- We can consider the moments for only one alternative because the second set of moments are redundant

We will estimate the parameters of the logit model using method of moments with the sample moment conditions

$$\frac{1}{N} \sum_{n=1}^N (y_{n1} - P_{n1}(\mathbf{x}_n, \beta)) \mathbf{x}_n = \mathbf{0}$$



# Method of Moments Estimation for Air Conditioner Choice

The method of moments estimator is the set of parameters,  $\hat{\beta}$ , that solves these sample moment conditions

$$\frac{1}{N} \sum_{n=1}^N \left( y_{n1} - P_{n1}(\mathbf{x}_n, \hat{\beta}) \right) \mathbf{x}_n = \mathbf{0}$$

Because the logit choice probabilities are nonlinear, we will find  $\hat{\beta}$  by minimizing the objective function

$$Q(\beta) = \left[ \frac{1}{N} \sum_{n=1}^N (y_{n1} - P_{n1}(\mathbf{x}_n, \beta)) \mathbf{x}_n \right]' \left[ \frac{1}{N} \sum_{n=1}^N (y_{n1} - P_{n1}(\mathbf{x}_n, \beta)) \mathbf{x}_n \right]$$

We will use the `gmm()` function from the `gmm` package in R to find this method of moments estimator

# Generalized Method of Moments in R

```
## Load the gmm package
library(gmm)

## Help file for the gmm function
?gmm

## Arguments for gmm function
gmm(g, x, t0, ...)
```

`gmm()` requires that you create a function, `g`, that

- 1 Takes a set of parameters and a data *matrix* as inputs
- 2 Calculates the empirical moments for each observation
- 3 Returns a  $N \times L$  matrix of empirical moments

You also have to give `gmm()` arguments for

- `x`: your data *matrix*
- `t0`: parameter starting values
- Many other optional arguments
  - ▶ `gmm()` can be very finicky with the optimization arguments

# Steps to Calculate the Logit Moments

The logit sample moments are

$$\frac{1}{N} \sum_{n=1}^N (y_{n1} - P_{n1}(\mathbf{x}_n, \beta)) \mathbf{x}_n$$

where the choice probabilities for our model are

$$P_{n1} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 P_n + \beta_2 C_n)}}$$

Steps to calculate the logit moments conditional on  $\beta$

- 1 Calculate the representative utility of AC for each household
- 2 Calculate the choice probability of AC for each household
- 3 Calculate the econometric residual for each household
- 4 Calculate the vector of moments for each household
- 5 Return the resulting  $N \times K$  matrix of empirical moments

# Organize Data Into a Matrix

```
## Create dataset for use in MM moment function
mm_data <- ac_data %>%
  mutate(air_conditioning = 1 * air_conditioning,
         constant = 1) %>%
  select(air_conditioning, constant, cost_system, cost_operating) %>%
  as.matrix()

## Look at data matrix
head(mm_data)
```

	air_conditioning	constant	cost_system	cost_operating
## [1,]	0	1	513	247
## [2,]	0	1	578	138
## [3,]	1	1	658	171
## [4,]	0	1	615	198
## [5,]	0	1	515	165
## [6,]	0	1	588	143

# Function to Calculate Logit Moments

```
## Function to calculate moments for air conditioning choice
mm_fn <- function(params, data){
  ## Select data for X [N x K]
  X <- data[, -1]
  ## Select data for y [N x 1]
  y <- data[, 1]
  ## Calculate representative utility of air conditioning [N x 1]
  utility <- X %*% params
  ## Calculate logit choice probability of air conditioning [N x 1]
  prob <- 1 / (1 + exp(-utility))
  ## Calculate econometric residuals [N x 1]
  residuals <- y - prob
  ## Create moment matrix [N x K]
  moments <- c(residuals) * X
  return(moments)
}
```

# Estimate Binary Logit Model

```
## Arguments for gmm function  
gmm(g, x, t0, ...)
```

```
## Find the MM estimator  
model_1 <- gmm(g = mm_fn, x = mm_data, t0 = rep(0, 3),  
               vcov = 'iid', method = 'Nelder-Mead',  
               control = list(reltol = 1e-25, maxit = 10000))
```

# Model Summary

`summary()` summarizes the results of the model

```
## Summarize model results
summary(model_1)
##
## Call:
## gmm(g = mm_fn, x = mm_data, t0 = rep(0, 3), vcov = "iid", method = "Nelder-Mead",
##     control = list(reltol = 1e-25, maxit = 10000))
##
##
## Method: twoStep
##
## Coefficients:
##           Estimate      Std. Error    t value      Pr(>|t|)
## Theta[1]  4.4366e+00    9.4473e-01    4.6962e+00    2.6506e-06
## Theta[2]  -2.9743e-03    1.4865e-03   -2.0010e+00    4.5398e-02
## Theta[3]  -1.5428e-02    2.4017e-03   -6.4236e+00    1.3312e-10
##
## J-Test: degrees of freedom is 0
##           J-test           P-value
## Test E(g)=0:    2.37195931661001e-20    *****
##
## #####
## Information related to the numerical optimization
## Convergence code = 0
## Function eval. = 1460
## Gradient eval. = NA
```

# Interpreting Parameters

`coef()` is the R function to display only the model parameters

```
## Display model coefficients
coef(model_1)
##      Theta[1]      Theta[2]      Theta[3]
## 4.436646632 -0.002974325 -0.015427739
```

How do we interpret these parameters?

- Air conditioning generates 4.44 “utils” of utility
- An additional \$100 of purchase price reduces utility by 0.30
- An additional \$100 of annual operating cost reduces utility by 1.54



# Endogeneity in Air Conditioner Choice

This interpretation of the parameters is correct only if the data are exogenous

- But if the data are correlated with the random utility term, then we cannot give the parameters this kind of “causal” interpretation

Is there any reason to worry about endogeneity in this setting?

- Why might the price be endogenous?
- Why might the annual operating cost be endogenous?

When the data are endogenous, we need to use exogenous instruments to recover parameters with a “causal” interpretation

# Instruments for Air Conditioner Choice

Could any of the other variables in the dataset be good instruments?

- Relevant, or correlated with the data in the model
- Exogenous, or not correlated with the random utility term

```
## Look at variable names to use as instruments
```

```
names(ac_data)
```

```
## [1] "air_conditioning" "cost_system"      "cost_operating"  
## [4] "elec_price"       "square_feet"      "income"  
## [7] "residents"        "city"
```

## GMM Estimation for Air Conditioner Choice

The GMM estimator is the set of parameters,  $\hat{\beta}$ , that solves these sample moment conditions

$$\frac{1}{N} \sum_{n=1}^N \left( y_{n1} - P_{n1}(\mathbf{x}_n, \hat{\beta}) \right) \mathbf{z}_n = \mathbf{0}$$

Because the logit choice probabilities are nonlinear, we will find  $\hat{\beta}$  by minimizing the objective function

$$Q(\beta) = \left[ \frac{1}{N} \sum_{n=1}^N (y_{n1} - P_{n1}(\mathbf{x}_n, \beta)) \mathbf{z}_n \right]' \mathbf{W} \left[ \frac{1}{N} \sum_{n=1}^N (y_{n1} - P_{n1}(\mathbf{x}_n, \beta)) \mathbf{z}_n \right]$$

We will use the `gmm()` function from the `gmm` package in R to perform the two-step procedure to find the optimal GMM estimator

# Organize Data Into a Matrix

```
## Create dataset for use in GMM moment function
gmm_data <- ac_data %>%
  mutate(air_conditioning = 1 * air_conditioning,
         constant = 1) %>%
  select(air_conditioning, constant, cost_system, cost_operating,
         elec_price, square_feet, residents) %>%
  as.matrix()
```

```
## Look at data matrix
head(gmm_data)
##      air_conditioning constant cost_system cost_operating elec_price square_feet residents
## [1,]                0         1         513           247        12.8         541         2
## [2,]                0         1         578           138         9.6         384         1
## [3,]                1         1         658           171        10.7         619         2
## [4,]                0         1         615           198        11.5         624         2
## [5,]                0         1         515           165        10.5         365         1
## [6,]                0         1         588           143         9.7         411         2
```

# Function to Calculate Logit Moments with Instruments

```
## Function to calculate moments for AC choice with instruments
gmm_fn <- function(params, data){
  ## Select data for X [N x K]
  X <- data[, 2:4]
  ## Select data for y [N x 1]
  y <- data[, 1]
  ## select data for Z [N x L]
  Z <- data[, c(2, 5:7)]
  ## Calculate representative utility of air conditioning [N x 1]
  utility <- X %*% params
  ## Calculate logit choice probability of air conditioning [N x 1]
  prob <- 1 / (1 + exp(-utility))
  ## Calculate econometric residuals [N x 1]
  residuals <- y - prob
  ## Create moment matrix [N x L]
  moments <- c(residuals) * Z
  return(moments)
}
```

# Estimate Binary Logit Model with Instruments

```
## Find the GMM estimator  
model_2 <- gmm(g = gmm_fn, x = gmm_data, t0 = rep(0, 3),  
               vcov = 'iid', method = 'Nelder-Mead',  
               control = list(reltol = 1e-25, maxit = 10000))
```

# Model Summary

```
## Summarize model results
summary(model_2)
##
## Call:
## gmm(g = gmm_fn, x = gmm_data, t0 = rep(0, 3), vcov = "iid", method = "Nelder-Mead",
##     control = list(reltol = 1e-25, maxit = 10000))
##
##
## Method: twoStep
##
## Coefficients:
##           Estimate      Std. Error  t value      Pr(>|t|)
## Theta[1]  4.8181e+00   1.2242e+00   3.9357e+00   8.2951e-05
## Theta[2]  1.9330e-04   1.9520e-03   9.9023e-02   9.2112e-01
## Theta[3] -3.0459e-02   3.6807e-03  -8.2756e+00   1.2786e-16
##
## J-Test: degrees of freedom is 1
##           J-test      P-value
## Test E(g)=0:   1.2643e+02  2.4738e-29
##
## Initial values of the coefficients
##      Theta[1]  Theta[2]  Theta[3]
##  1.76601518  0.04517918 -0.17724070
##
## #####
## Information related to the numerical optimization
## Convergence code = 0
## Function eval. = 578
## Gradient eval. = NA
```

# Interpreting Parameters

```
## Display model coefficients
coef(model_2)
##      Theta[1]      Theta[2]      Theta[3]
## 4.8180601696 0.0001932951 -0.0304594326
```

How do we interpret these parameters?

- Air conditioning generates 4.82 “utils” of utility
- Purchase price does not affect utility
- An additional \$100 of annual operating cost reduces utility by 3.05



# Test of Overidentifying Restrictions

`specTest()` is the `gmm` function to perform a test of overidentifying restrictions

- Test if all of our population moments equal zero

$$H_0 : E[(y_{n1} - P_{n1}(\mathbf{x}_n, \beta)) \mathbf{z}_n] = \mathbf{0}$$

```
## Test overidentifying restrictions
specTest(model_2)
##
## ## J-Test: degrees of freedom is 1 ##
##
##          J-test          P-value
## Test E(g)=0:    1.2643e+02    2.4738e-29
```

We can reject the null hypothesis and conclude that our model is overidentified and not correctly specified