# Method of Moments, Generalized Method of Moments, and Simulated Method of Moments

Jesús Bueren

EUI

# Introduction

- Most papers that we are going to cover in this course estimate parameters using the method of simulated moments.
  - For this purpose, we are going to revise the general method of moments.
  - Application to life-cycle heterogeneous agents models.
- These slides are based on Greene Chapter 13, Hayashi chapter 3, and Arellano Appendix A

# The Method of Moments

- GMM estimators move away from parametric assumptions about the data generating process made when using maximum likelihood.
- GMM exploits the fact that sample statistics each have a counterpart in the population:
  - e.g. sample mean and population expected value
- Is it a good idea to use sample data to infer characteristics of the population?

# The Method of Moments

- Considear i.i.d random sampling from distribution $f(y|\theta_1, \theta_2, \ldots, \theta_K)$ with finite moments $E[y^{2K}]$.

- The $k^{th}$ "raw" uncentered moment is given by:

$$\bar{m}_k(\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} y_i^k \tag{1}$$

- By the LLN we have:

$$E[\bar{m}_k(\boldsymbol{y})] = \mu_k = E[y_i^k] \tag{2}$$

$$Var[\bar{m}_k(\boldsymbol{y})] = \frac{1}{n}Var[y_i^k] = \frac{1}{n}(\mu_{2k} - \mu_k^2) \tag{3}$$

- By the CLT:

$$\sqrt{n}(\bar{m}_k(\boldsymbol{y}) - \mu_k) \xrightarrow{d} N(0, \mu_{2k} - \mu_k^2) \tag{4}$$

# The Method of Moments

## General Idea

- In general, $\mu_k$ is going to be a function of the underlying parameters.
- By computing $K$ raw moments in the data and equating them to the functions implied by the population moments:
  - We obtain $K$ equations with $K$ unknowns.
  - In principle, we could solve this system of equations to provide estimates of the $K$ unknown parameters.

# The Method of Moments

- Moments based on powers of $y$ provide a natural source of information about the parameter.

- Instead, functions of the data may also be useful.

- Let $m_k(.)$ be a continuous and differentiable function.

- We could construct the following data moment:

$$\bar{m}_k(\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} m_k(y_i), k = 1, 2, \ldots, K \tag{5}$$

- By the LLN:

$$plim_{n\to\infty}\, \bar{m}_k(\boldsymbol{y}) = E[m_k(y_i)] = \mu_k(\theta_1, \ldots, \theta_K)$$

# The Method of Moments

- We define a moment conditions as a function of the model and data, such that their expectation is zero at the true parameter values:

$$E(m_k(y, \boldsymbol{\theta}_0)) = 0$$

- With $K$ parameters, the method of moments estimator can be defined as parameter vector $\hat{\boldsymbol{\theta}}$ that solves for the sample analog of the population moment conditions:

$$\begin{aligned} \bar{m}_1(\boldsymbol{y}, \hat{\boldsymbol{\theta}}) &= \frac{1}{N}\sum_{i=1}^{n} m_1(y_i, \hat{\boldsymbol{\theta}}) = 0 \\ &\vdots \\ \bar{m}_k(\boldsymbol{y}, \hat{\boldsymbol{\theta}}) &= \frac{1}{N}\sum_{i=1}^{n} m_k(y_i, \hat{\boldsymbol{\theta}}) = 0 \end{aligned}$$

# The Method of Moments

## Example 1: Method of Moments for $N(\mu, \sigma^2)$

- By LLN:

$$m_1(y, \mu) = E[y - \mu] = 0$$
$$m_2(y, \mu, \sigma) = E[(y - \mu)^2 - \sigma^2] = 0$$

- Their corresponding sample analogs give us the moment estimator:

$$\bar{m}_1(\boldsymbol{y}, \hat{\mu}) = \frac{1}{n}\sum_{n=1}^{N}(y_i - \hat{\mu}) = 0$$
$$\bar{m}_2(\boldsymbol{y}, \hat{\mu}, \hat{\sigma}^2) = \frac{1}{n}\sum_{n=1}^{N}(y_i - \hat{\mu})^2 - \hat{\sigma}^2 = 0$$

# The Method of Moments

## Example 2: Gamma Distribution

- The gamma distribution is

$$f(y) = \frac{\lambda^p}{\Gamma(P)} e^{-\lambda y} y^{P-1}, y \geq 0, P > 0, \lambda > 0$$

- Imagine you had $n$ i.i.d random draws from $f(y)$.
- By the properties of the gamma distribution we have:

$$E \begin{bmatrix} y - P/\lambda \\ y^2 - P(P+1)/\lambda^2 \\ \ln y - \Psi(P) - \ln \lambda \\ 1/y - \lambda/(P-1) \end{bmatrix} = 0$$

- Depending on the targeted moments you will obtain different solutions (see code)

# Identification

- We have a set of moment condition that hold in the population:

$$E[\boldsymbol{m}(\boldsymbol{y}, \boldsymbol{\theta_0})] = 0 \tag{6}$$

- Let $\hat{\boldsymbol{\theta}}$ be a a vector of parameter such that:

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

- We say that the coefficient vector is identified if $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$

- Conditions for identification:

  1. Number of moment conditions equal to number of parameters.
  2. The matrix of derivatives, $\bar{\boldsymbol{G}}(\boldsymbol{\theta}_0)$, will have full rank i.e. rank K. Question: Is it a problem if two moments are linearly dependent?
  3. If $\boldsymbol{m}(\boldsymbol{y}, \boldsymbol{\theta})$ is continuous, the parameter vector that satisfies the population moments conditions is unique.

# The Method of Moments

## Asymptotic Properties

- In a few cases, we can obtain the exact distribution of the method of moments estimator.
  - For example, in sampling from the normal distribution, $\hat{\mu} \sim N(\mu, \sigma^2/n)$
- In general we don't know the distribution of the estimated parameters.
  - We are going to use the CLT to construct asymptotic approximation of distributions of the estimated parameters.

# The Method of Moments

## Asymptotic Properties

- From the application of the central limit theorem we know that:

$$\sqrt{N}\bar{\boldsymbol{m}}(\boldsymbol{y},\boldsymbol{\theta}_0) = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{m}(\boldsymbol{y_i},\boldsymbol{\theta}_0) \xrightarrow{d} N(0,\boldsymbol{\Phi}),$$

  where $\boldsymbol{\Phi} = E[\boldsymbol{m}(\boldsymbol{y},\boldsymbol{\theta}_0)\boldsymbol{m}(\boldsymbol{y},\boldsymbol{\theta}_0)']$ is the asymptotic variance covariance matrix of the moment conditions.

- Let's denote $\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)$ the gradient of the moment conditions:

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}_0) = \frac{\partial\bar{\boldsymbol{m}}(\boldsymbol{y},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}$$

# The Method of Moments

## Asymptotic Properties

- Empirically $\hat{\boldsymbol{\theta}}$ us found by solving the system of equations:

$$\bar{\boldsymbol{m}}(\boldsymbol{y}, \hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}(y_i, \hat{\boldsymbol{\theta}}) = \mathbf{0}$$

  a consistent estimator of the asymptotic covariance of the moment conditions can be computed using:

$$\boldsymbol{F}_{jk} = \frac{1}{n} \sum_{i=1}^{n} m_j(y_i, \hat{\boldsymbol{\theta}}) m_k(y_i, \hat{\boldsymbol{\theta}})$$

- The estimator provides the asymptotic covariance matrix of the moments.

$$\boldsymbol{F} \xrightarrow{p} \boldsymbol{\Phi},$$

# The Method of Moments

## Asymptotic Properties

- Under our assumption of random sampling, although the precise distribution of the parameters is likely to be unknown, we can appeal to the CLT to obtain asymptotic approximation.

- Let $\bar{\boldsymbol{G}}(\boldsymbol{\theta})$ denote the $K \times K$ matrix whose $k$th row is the vector of partial derivatives,

$$\bar{\boldsymbol{G}}_k(\bar{\boldsymbol{\theta}})' = \frac{\partial \bar{m}_k(y, \bar{\boldsymbol{\theta}})}{\partial \bar{\boldsymbol{\theta}}}$$

- Assuming that the functions in the moment conditions are continuous and functionally independent,

$$\bar{\boldsymbol{G}}_k(\hat{\boldsymbol{\theta}})' \xrightarrow{p} \boldsymbol{\Gamma}_k(\boldsymbol{\theta}_0)'$$

# The Method of Moments

## Asymptotic Properties

- Assuming moment conditions are continuous and continuously differentiable, by the mean value theorem, there exists a point $\bar{\boldsymbol{\theta}}$ in $(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)$ such that:

$$\bar{\boldsymbol{m}}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$$

$$\bar{\boldsymbol{m}}(\boldsymbol{\theta_0}) + \bar{\boldsymbol{G}}'(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) = \mathbf{0}$$

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\bar{\boldsymbol{G}}'(\bar{\boldsymbol{\theta}})^{-1}\sqrt{N}\,\bar{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta_0})$$

- Given that we know the asymptotic distribution of $\sqrt{N}\,\bar{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta_0})$ and that $\hat{\theta}$ is consistent, then $\bar{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$ and $\boldsymbol{G}(\bar{\boldsymbol{\theta}}) \to \boldsymbol{G}(\boldsymbol{\theta_0})$, thus:

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, [\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)]^{-1}\boldsymbol{\Phi}[\boldsymbol{\Gamma}'(\boldsymbol{\theta}_0)]^{-1})$$

- Then the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_0$ may be estimated with:

$$\text{Est.Asy.Var}[\hat{\boldsymbol{\theta}}] = \frac{1}{n}[\bar{\boldsymbol{G}}(\hat{\boldsymbol{\theta}})]^{-1}\boldsymbol{F}[\bar{\boldsymbol{G}}'(\hat{\boldsymbol{\theta}})]^{-1}$$

# The Method of Moments

## Example: The Normal Distribution

- We know that in the specific case of estimating the parameters of a normal distribution:
  - the distribution of the mean is exactly normal
  - the distribution of the variance is a chi-square.
  - the two distributions are independent
- The joint is a mixture of two independent distributions: a normal and a chi-square.
- For teaching purposes, let's ignore this and assume the general case where we don't know the distribution of the estimated parameters.

# The Method of Moments

## Example: The Normal Distribution

- We rewrite the moment conditions:

$$\bar{m}_1(\boldsymbol{y}, \hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\mu} = 0$$

$$\bar{m}_2(\boldsymbol{y}, \hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{\mu})^2 - \hat{\sigma}^2 = 0,$$

where,

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{bmatrix}, \bar{\boldsymbol{m}}(\boldsymbol{y}, \hat{\boldsymbol{\theta}}) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\mu} \\ \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{\mu})^2 - \hat{\sigma}^2 \end{bmatrix}$$

# The Method of Moments

## Example: The Normal Distribution

- So let's derive again for this case the large sample properties:

$$
\begin{aligned}
&\sqrt{n}\bar{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Phi}) \\
&\bar{\boldsymbol{m}}(\boldsymbol{y}, \hat{\boldsymbol{\theta}}) = 0 \\
&\bar{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta}_0) + \boldsymbol{G}'(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \simeq 0 \\
&\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \simeq [-\boldsymbol{G}'(\boldsymbol{\theta}_0)]^{-1}\bar{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta}_0),
\end{aligned}
$$

where

$$
\boldsymbol{G}(\boldsymbol{\theta}_0) = \begin{bmatrix} -1 & 0 \\ \frac{2}{n}\sum_{i=1}^{n}(y_i - \mu) & -1 \end{bmatrix}
$$

# The Method of Moments

## Example: The Normal Distribution

- Therefore we have,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, [\boldsymbol{G}'(\boldsymbol{\theta}_0)]^{-1}\boldsymbol{\Phi}[\boldsymbol{G}(\boldsymbol{\theta}_0)]^{-1})$$

- Thus with an estimator of the covariance equal to:

$$\begin{aligned}\text{Est.Asy.Var}[\hat{\boldsymbol{\theta}}] =& \frac{1}{n}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}^{-1}\\ & \begin{bmatrix}\frac{1}{n}\sum_i^n m_1(y_i,\hat{\boldsymbol{\theta}})^2 & \frac{1}{n}\sum_i^n m_1(y_i,\theta)m_2(y_i,\hat{\boldsymbol{\theta}})\\ \frac{1}{n}\sum_i^n m_1(y_i,\theta)m_2(y_i,\hat{\boldsymbol{\theta}}) & \frac{1}{n}\sum_i^n m_2(y_i,\hat{\boldsymbol{\theta}})^2\end{bmatrix}\\ & \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}^{-1}\end{aligned}$$

# The Method of Moments

## Example: The Gamma Distribution

- Now let's go again back to our example of the gamma distribution.

$$\bar{m}_1(\hat{\boldsymbol{\theta}}, \boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{P}/\hat{\lambda}$$

$$\bar{m}_2(\hat{\boldsymbol{\theta}}, \boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} 1/y_i - \hat{\lambda}/(\hat{P}-1)$$

Thus,

$$\boldsymbol{G}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} -1/\hat{\lambda} & \hat{P}/\hat{\lambda}^2 \\ \hat{\lambda}/(\hat{P}-1)^2 & -1/(\hat{P}-1) \end{bmatrix}$$

# The Method of Moments

## Example: Linear regression model

- In the previous case the optimal weighting matrix is only a function of the data.
- Now let's have a look into the linear regression model:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i$$

- The lack of contemporaneous correlation, gives us a set of moment equations:

$$E[m_{i,k}] = E[x_{i,k}\epsilon_i] = 0$$

- We have $K$ equations and $K$ unknowns.

# The Method of Moments

## Example: Linear regression model

$$\bar{\boldsymbol{m}}(\hat{\boldsymbol{\beta}}, \boldsymbol{x}, \boldsymbol{y}) = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} x_{i,1} \hat{\epsilon}_i \\ \vdots \\ \frac{1}{n} \sum_{i=1}^{n} x_{i,K} \hat{\epsilon}_i \end{bmatrix} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^{n} x_{i,1} (y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^{n} x_{i,K} (y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}) \end{bmatrix}$$

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

$$\hat{\boldsymbol{\beta}} = \Big[ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \Big]^{-1} \Big[ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i y_i \Big]$$

# The Method of Moments

## Example: Linear regression model

$$\boldsymbol{G}(\hat{\theta}) = \begin{bmatrix} -\frac{1}{n}\sum_{i=1}^{n} x_{i,1}x_{i,1} & \cdots & -\frac{1}{n}\sum_{i=1}^{n} x_{i,1}x_{i,K} \\ \vdots & \cdots & \vdots \\ -\frac{1}{n}\sum_{i=1}^{n} x_{i,K}x_{i,1} & \cdots & -\frac{1}{n}\sum_{i=1}^{n} x_{i,K}x_{i,K} \end{bmatrix} = -\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'$$

$$F = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\hat{\epsilon}_i^2$$

- Note that this is the heteroskedasticity consistent variance estimator of White.

# Generalized Method of Moments

- Following our discussion using the example from the gamma distribution, what do we do when we have more moments than parameters?

# Generalized Method of Moments

- Suppose now that the model involves $K$ parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)'$ and that the theory provides a set of $L \geq K$ moment conditions:

$$E[m_l(\boldsymbol{\theta}_0, y_i)] = 0$$

- Denote the corresponding sample mean as:

$$\bar{m}_l(\boldsymbol{\theta}_0, \boldsymbol{y}) = \frac{1}{n} \sum_{i=1}^{n} m(\boldsymbol{\theta}_0, y_i)$$

# Generalized Method of Moments

- We aim at finding $\hat{\boldsymbol{\theta}}$ that solves the following system of $L$ equations and $K$ unknowns:

$$\bar{\boldsymbol{m}}(\hat{\boldsymbol{\theta}}, \boldsymbol{y}) = 0$$

- As long as the equations are independent, the system will not have a unique solution.

- It will be necessary to reconcile the different sets of estimates that can be produced.

- We can use as the criterion a weighted sum of squares:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ \bar{\boldsymbol{m}}'(\boldsymbol{\theta}, \boldsymbol{y}) W \bar{\boldsymbol{m}}(\boldsymbol{\theta}, \boldsymbol{y}),$$

  where $W$ is any positive definite matrix that may depend on the data but is not a function of $\boldsymbol{\theta}$

# Generalized Method of Moments

## Identification

- We have a set of moment condition that hold in the population:

$$E[\boldsymbol{m}(\boldsymbol{y}, \boldsymbol{\theta_0})] = 0 \tag{7}$$

- Let $\hat{\boldsymbol{\theta}}$ be a a vector of parameter such that:

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

- We say that the coefficient vector is identified if $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$

- Conditions for identification:
  1. Number of moment conditions larger or equal to number of parameters.
  2. The matrix of derivatives, $\bar{\boldsymbol{G}}(\boldsymbol{\theta}_0)$, will have full rank (i.e. rank of K). Question: Is it a problem if two moments are linearly dependent?
  3. If $\boldsymbol{m}(\boldsymbol{y}, \boldsymbol{\theta})$ is continuous, the parameter vector that satisfies the population moments conditions is unique.

# The Method of Moments

## Asymptotic Properties

- From the application of the central limit theorem we have the same asymptotic distribution of mean as before:

$$\sqrt{N}\bar{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta}_0) = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{m}(\boldsymbol{y_i}, \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Phi}),$$

  where $\boldsymbol{\Phi} = E[\boldsymbol{m}(\boldsymbol{y}, \boldsymbol{\theta}_0)\boldsymbol{m}(\boldsymbol{y}, \boldsymbol{\theta}_0)']$ is the asymptotic variance covariance matrix of the moment conditions but now is of dimension $L \times L$ (instead of $K \times K$)

- Let's denote $\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)$ the gradient of the moment conditions:

$$\boldsymbol{\Gamma}(\boldsymbol{\theta}_0) = \frac{\partial \bar{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_0}$$

# The Generalized Method of Moments

## Asymptotic Properties

- An appropriate estimator of the asymptotic covariance of the moment conditions $\bar{\boldsymbol{m}} = [\bar{m}_1, \ldots, \bar{m}_l]$ can be computed using:

$$\boldsymbol{F}_{jk} = \frac{1}{n}\sum_{i=1}^{n} m_j(y_i, \hat{\boldsymbol{\theta}}) m_k(y_i, \hat{\boldsymbol{\theta}}))$$

- The estimator provides the asymptotic covariance matrix of the moments.

$$\boldsymbol{F} \xrightarrow{p} \boldsymbol{\Phi},$$

# The Generalized Method of Moments

## Asymptotic Properties

- Let $\bar{\boldsymbol{G}}(\hat{\boldsymbol{\theta}})$ denote the $L \times K$ matrix whose $l$th row is the vector of partial derivatives,

$$\bar{\boldsymbol{G}}_l(\hat{\boldsymbol{\theta}})' = \frac{\partial \bar{m}_l(y, \hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}$$

- Assuming that the functions in the moment conditions are continuous and functionally independent:

$$\bar{\boldsymbol{G}}(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \boldsymbol{\Gamma}(\boldsymbol{\theta}_0)$$

# The Generalized Method of Moments

## Asymptotic Properties

- The first-order conditions for the GMM estimator are:

$$2\bar{\boldsymbol{G}}'(\hat{\boldsymbol{\theta}})W\bar{\boldsymbol{m}}(\hat{\boldsymbol{\theta}},\boldsymbol{y}) = 0 \tag{8}$$

- We apply the mean-value theorem for a point in the parameter space $\bar{\boldsymbol{\theta}}$:

$$\bar{\boldsymbol{m}}(\hat{\boldsymbol{\theta}}) = \bar{\boldsymbol{m}}(\boldsymbol{\theta_0}) + \bar{\boldsymbol{G}}'(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) \tag{9}$$

- Insert equation (9) in (8) to obtain:

$$\bar{\boldsymbol{G}}'(\bar{\boldsymbol{\theta}})W\bar{\boldsymbol{m}}(\boldsymbol{\theta_0}) + \bar{\boldsymbol{G}}'(\bar{\boldsymbol{\theta}})W\bar{\boldsymbol{G}}(\bar{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) = 0$$

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0}) = -[\bar{\boldsymbol{G}}'(\bar{\boldsymbol{\theta}})W\bar{\boldsymbol{G}}(\bar{\boldsymbol{\theta}})]^{-1}\bar{\boldsymbol{G}}'(\bar{\boldsymbol{\theta}})W\sqrt{n}\bar{\boldsymbol{m}}(\boldsymbol{\theta_0})$$

# The Generalized Method of Moments

## Asymptotic Properties

- By CLT we have:

$$\sqrt{n}\bar{\boldsymbol{m}}_0 \xrightarrow{d} N(\mathbf{0}, \mathbf{\Phi})$$

and $\bar{\boldsymbol{G}}(\bar{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{\Gamma}(\boldsymbol{\theta}_0)$, therefore,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, [\mathbf{\Gamma}'(\boldsymbol{\theta}_0)W\mathbf{\Gamma}(\boldsymbol{\theta}_0)]^{-1}\mathbf{\Gamma}'(\boldsymbol{\theta}_0)W\mathbf{\Phi}W\mathbf{\Gamma}(\boldsymbol{\theta}_0)[\mathbf{\Gamma}'(\boldsymbol{\theta}_0)W\mathbf{\Gamma}(\boldsymbol{\theta}_0)]^{-1})$$

- Then the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ may be estimated with:

$$\text{Est.Asy.Var}[\hat{\boldsymbol{\theta}}] = \frac{1}{N}[\bar{\boldsymbol{G}}'(\hat{\boldsymbol{\theta}})W\bar{\boldsymbol{G}}(\hat{\boldsymbol{\theta}})]^{-1}\bar{\boldsymbol{G}}'(\hat{\boldsymbol{\theta}})W\boldsymbol{F}W\bar{\boldsymbol{G}}(\hat{\boldsymbol{\theta}})[\bar{\boldsymbol{G}}'(\hat{\boldsymbol{\theta}})W\bar{\boldsymbol{G}}(\hat{\boldsymbol{\theta}})]^{-1}$$

# The Generalized Method of Moments

## Asymptotic Properties

- When using the identity matrix we get the White estimator:

$$\text{Asy.Var}[\boldsymbol{\theta}_0] = \frac{1}{N}[\boldsymbol{\Gamma}'(\boldsymbol{\theta}_0)\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)]^{-1}\boldsymbol{\Gamma}'(\boldsymbol{\theta}_0)\boldsymbol{\Phi}\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)[\boldsymbol{\Gamma}'(\boldsymbol{\theta}_0)\boldsymbol{\Gamma}(\boldsymbol{\theta}_0)]^{-1}$$

- If we define the weighting function as the inverse of the variance-covariance matrix of the moment condition (the optimal weighting matrix) we obtain:

$$\text{Asy.Var}[\boldsymbol{\theta}_0] = \frac{1}{N}[\boldsymbol{\Gamma}'(\boldsymbol{\theta}_0)\boldsymbol{\Phi}^{-1}\boldsymbol{\Gamma}[(\boldsymbol{\theta}_0)]^{-1}$$

# The Generalized Method of Moments

## 2-step Estimation

1. Use $\boldsymbol{W} = \boldsymbol{I}$ to obtain a consistent estimator of $\boldsymbol{\theta}_0$. Then obtain an estimate of $\boldsymbol{\Phi}$ using the variance covariance matrix $\hat{\boldsymbol{F}}$ of $\bar{\boldsymbol{m}}(\boldsymbol{y}, \hat{\boldsymbol{\theta}})$.

2. Setting $\boldsymbol{W} = \boldsymbol{F}^{-1}$, compute a new estimation of $\boldsymbol{\theta}_0$ using a weighting matrix “close” to the optimal.

# Testing the Validity of the Moment Restrictions

- If the parameters are overidentified by the moment equations, then these equations imply substantive restrictions.
- As such, if the hypothesis of the model that led to the moment equations in the first place is incorrect, at least some of the sample moment restrictions will be systematically violated.
- When the optimal weighting matrix is used:

$$nq = [\sqrt{n}\bar{\boldsymbol{m}}(\boldsymbol{y},\boldsymbol{\theta})'] \{\text{Est.Asy.Var}[\sqrt{n}\bar{\boldsymbol{m}}(\boldsymbol{y},\boldsymbol{\theta})]\}^{-1} [\sqrt{n}\bar{\boldsymbol{m}}(\boldsymbol{y},\boldsymbol{\theta})]$$

- Under the null that the restrictions are true,

$$nq \xrightarrow{d} \chi^2[L-K],$$

where q is the value of the objective function.

# The Method of Moments

## Examples: Gamma distribution

- For the Gamma distribution case that we saw before, we have 4 moment conditions and 2 parameters to estimate:

$$\bar{\boldsymbol{m}} = \begin{bmatrix} \frac{1}{n}\sum_i^n y_i - \frac{\hat{P}}{\hat{\lambda}} \\ \frac{1}{n}\sum_i^n y_i^2 - \frac{\hat{P}(\hat{P}+1)}{\hat{\lambda}^2} \\ \frac{1}{n}\sum_i^n \ln(y_i) - \Psi(\hat{P}) + \ln(\hat{\lambda}) \\ \frac{1}{n}\sum_i^n \frac{1}{y_i} - \frac{\hat{\lambda}}{\hat{P}-1} \end{bmatrix}$$

# The Method of Moments

## Examples: Gamma distribution

$$\bar{\boldsymbol{G}}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} -\frac{1}{\hat{\lambda}} & \frac{\hat{P}}{\hat{\lambda}^2} \\ -\frac{2\hat{P}+1}{\hat{\lambda}^2} & \frac{\hat{P}(\hat{P}+1)}{\hat{\lambda}^4} \\ \frac{\hat{\lambda}}{(\hat{P}-1)^2} & -\frac{1}{\hat{P}-1} \\ -\psi'(\hat{P})\ln(\hat{\lambda}) & -\psi(\hat{P})\frac{1}{\hat{\lambda}} \end{bmatrix}$$

$$\hat{\boldsymbol{\Phi}} =$$

$$\begin{bmatrix} var(y_i) & cov(y_i, y_i^2) & cov(y_i, 1/y_i) & cov(y_i, ln(y_i)) \\ cov(y_i^2, y_i) & var(y_i^2) & cov(y_i^2, 1/y_i) & cov(y_i^2, ln(y_i)) \\ cov(1/y_i, y_i) & cov(1/y_i, y_i^2) & var(1/y_i) & cov(1/y_i, ln(y_i)) \\ cov(ln(y_i), y_i) & cov(ln(y_i), y_i^2) & cov(ln(y_i), 1/y_i) & var(ln(y_i)) \end{bmatrix}$$

# The Generalized Method of Moments

## Example: IV with more instruments than exogenous regressors

- In the previous case the optimal weighting matrix is only a function of the data.
- Now let's have a look into the linear regression model:

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \epsilon_i$$

- Now let's imagine the exogeneity assumption does not hold but we have access to $L \geq K$ variables correlated with $X$ but not with $\epsilon$.

$$E[m_{i,l}] = E[z_{i,l}\epsilon_i] = 0$$

- We have $L$ equations and $K$ unknowns.

# The Generalized Method of Moments

## Example: IV with more instruments than exogenous regressors

$$
\begin{aligned}
\bar{\boldsymbol{m}} &= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} z_{i,1}\hat{\epsilon}_i \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} z_{i,L}\hat{\epsilon}_i \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} z_{i,1}(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} z_{i,L}(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}) \end{bmatrix} \\
&= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}) = 0
\end{aligned}
$$

- This is a system of $L$ equations and $K$ unknowns.

# The Generalized Method of Moments

## Example: Estimate life-cycle model using consumption data

-

# Method of Simulated Moments

- GMM requires the sample moment restrictions to have a closed form as a function of the underlying parameters.
- Sometimes a close form solution is not available.
- Mcfadden (1989) and Pollard and Pakes (1989) propose a simulation based algorithm to compute moment conditions.
  - Much more computer intensive.

# Method of Simulated Moments

- Suppose we have the following moment condition

$$E[m(y_i, \boldsymbol{\theta})] = 0$$

- However and in contrast to the previous section, we do not have a close form solution to compute $m(y_i, \boldsymbol{\theta})$.
  - ▶ This could be because we do not have an analytic mapping between data moments and the parameters that we want to estimate.
  - ▶ Presence of unobserved heterogeneity.
- Given a function $g$ such that $m(y, \boldsymbol{\theta}) = \int g(y, \zeta, \boldsymbol{\theta}) P(\zeta) d\zeta$, the simulated method of moments simulates a large number of auxiliary data $\boldsymbol{\zeta}^{(s)}$ so that we are able to produce an estimate of the moment conditions

$$\hat{m}_k(y_i, \boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^{S} g_k(y_i, \zeta_i^s, \boldsymbol{\theta}),$$

# Method of Simulated Moments

- Then the objective then is to find:

$$\hat{\theta} = \arg\min_{\theta} \bar{\boldsymbol{m}}'(\boldsymbol{\theta}, \boldsymbol{y}) W \bar{\boldsymbol{m}}(\boldsymbol{\theta}, \boldsymbol{y}),$$

where

$$\bar{\boldsymbol{m}}(\boldsymbol{y}, \boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^{n} \hat{m}_1(y_i, \boldsymbol{\theta}) \\ \vdots \\ \frac{1}{N} \sum_{i=1}^{n} \hat{m}_l(y_i, \boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^{n} \frac{1}{S} \sum_{s=1}^{S} g_1(y_i, \zeta_i^s, \boldsymbol{\theta}) \\ \vdots \\ \frac{1}{N} \sum_{i=1}^{n} \frac{1}{S} \sum_{s=1}^{S} g_l(y_i, \zeta_i^s, \boldsymbol{\theta}) \end{bmatrix}$$

# Method of Simulated Moments

- With the optimal weighting matrix we obtain:

$$Est.Asy.Var[\hat{\boldsymbol{\theta}}] = \frac{1}{N}(1 + \frac{1}{S})[\bar{\boldsymbol{G}}'(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Phi}}^{-1}\bar{\boldsymbol{G}}(\hat{\boldsymbol{\theta}})]^{-1}$$

- When S is large, the variance convergences to the GMM case.
- $\bar{\boldsymbol{G}}'(\hat{\boldsymbol{\theta}})$ needs generally to be computed numerically.

# Method of Simulated Moments

## Example: Gamma Distribution

- Imagine that we ignored the statistical properties of the gamma distribution that we used to construct moment conditions.
- We could estimate $(P, \lambda)$ by matching the sample mean of $y, y^2$, and $ln(y)$ by constructing:

$$\bar{\boldsymbol{m}}(\hat{P}, \hat{\lambda}) = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{S} \sum_{s=1}^{S} y_i - y_{i,s}(\hat{P}, \hat{\lambda}, \zeta_i^s) \\ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{S} \sum_{s=1}^{S} y_i^2 - y_{i,s}(\hat{P}, \hat{\lambda}, \zeta_i^s)^2 \\ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{S} \sum_{s=1}^{S} \ln(y_i) - \ln\left(y_{i,s}(\hat{P}, \hat{\lambda}, \zeta_i^s)\right) \end{bmatrix}$$

- $y_{i,s}(\hat{P}, \hat{\lambda}, \zeta_i^s)$ is sampled from a gamma distribution with $\hat{P}$ and $\hat{\lambda}$.
- Don't forget to set the seed each time you try a new set of parameters to fix the sequence of $\zeta_i^s$

# Method of Simulated Moments

## Example: Life-cycle model

- A social planner maximizes:

$$\max_{c_t, k_{t+1}} E_0 \sum_{t=0}^{\infty} \beta^t (1+\epsilon_t) \frac{c^{1-\sigma}}{1-\sigma}$$
$$\text{s.t. } c_t + k_{t+1} = z_t f(k_t) + (1-\delta) k_t,$$

  with $\epsilon \sim N(0, \sigma_\epsilon)$ known.

- The euler equation becomes:

$$(1+\epsilon_t) c_t^{\sigma} = \beta E[z_{t+1}(f_k(k_{t+1}) + 1 - \delta)(1+\epsilon_{t+1}) c_{t+1}^{\sigma}]$$

- Draw $\{\epsilon_t^{(s)}\}_{t=1}^{T}$.

- Given $\epsilon$ we can estimate the model using the MSM.

# Method of Simulated Moments

## Example: Estimate life-cycle model using asset data

-