Econometrics 2
Lecture 6: Instrumental Variables and Moment Conditions
GMM Objective Functions

AUEB | Spring Semester 2026

Taught by: Prof. S. Arvanitis
Notes transcribed & digitized by: T. Kourtalis

October 31, 2026

1. System of Moment Conditions

In the framework of the linear model with instrumental variables, we are led to the q-dimensional vector of moments ($q \geq p$):

$$f(\beta) = \mathbb{E}\left[\frac{1}{n} Z_n'(Y_n - X_n\beta)\right] = \cdots = \frac{1}{n}\mathbb{E}[Z_n' X_n](\beta_0 - \beta)$$

This utilizes the Law of Iterated Expectations, since the errors of Z_n will be orthogonal to the columns of $X_n \implies \mathbb{E}(Z_n'\epsilon_n) = 0$.

Using the identification condition $rank(X_n) = p$, the previous equation implies that we could find β_0 if we could solve the system:

$$\mathbb{E}\left(\frac{1}{n} Z_n' X_n\right)(\beta_0 - \beta) = 0_{q\times 1}$$

This is a system of **Moment Conditions**. Obviously, the left side of our system is unobservable since it depends on the true parameter β_0.

2. Sample Approximation and Objective Functions

What we could do is try to solve the sample approximation:

$$\frac{1}{n} Z_n'(Y_n - X_n\beta) = 0_{q\times 1}$$

(We try to solve for this equation when $q \geq p$).

When we talk about length and distance, we are talking about the same thing here. We will try to express this as something that results from an optimization problem. In this case, we will try to construct an objective function that will represent some notion of distance of $\frac{1}{n}\mathbb{E}(Z_n' X_n)(\beta_0 - \beta)$ from 0.

There are different ways to do this:

Example 1: Maximum Absolute Distance

We could use the following distance. If $x, y \in \mathbb{R}^q$, then as a distance between them we could use $\max_{j=1,\dots,q} |x_j - y_j|$. So, we can define:

$$\mu_n^*(\beta) = \max_{j=1,\dots,q} \left| \left(\frac{1}{n} \mathbb{E}(Z_n' X_n)(\beta_0 - \beta) \right)_j \right|$$

Because of the previous properties, we have that $\arg\min_{\beta \in \Theta} \mu_n^*(\beta) = \beta_0$.

Correspondingly with what we have done in our first example, under certain assumptions the unobservable μ_n^* can be approximated by the observable counterpart:

$$\mu_n(\beta) = \max_{j=1,\dots,q} \left| \left(\frac{1}{n} Z_n'(Y_n - X_n\beta) \right)_j \right|$$

So the corresponding estimator $\hat{\beta}_n$ will be defined from $\hat{\beta}_n \in \arg\min_{\beta \in \Theta} \mu_n(\beta)$.

Example 2: Mahalanobis Type Distance

We will use distances of the Mahalanobis type. Let matrix W be **symmetric and positive definite**; the distance between x, y with respect to W is $\sqrt{(x-y)'W(x-y)}$. Here, W acts as a *weighting matrix*. For example, when $W = I_{n \times n}$, we have the usual Euclidean distance.

Given such a W, we can construct:

$$\mu_n^*(\beta, W) = \left[\left(\frac{1}{n} \mathbb{E}(Z_n' X_n)(\beta_0 - \beta) \right)' W \left(\frac{1}{n} \mathbb{E}(Z_n' X_n)(\beta_0 - \beta) \right) \right]^{1/2}$$

Correspondingly, under assumptions, the unobservable μ_n^* can be approximated by the observable. *(Note: Minimizing a positive distance is equivalent to minimizing its square. We drop the square root for computational convenience in derivatives.)*

$$\mu_n(\beta, W) = \left(\frac{1}{n} Z_n'(Y_n - X_n\beta) \right)' W \left(\frac{1}{n} Z_n'(Y_n - X_n\beta) \right) = \frac{1}{n^2}(Y_n - X_n\beta)'(Z_n W Z_n')(Y_n - X_n\beta)$$

3. The Instrumental Variables Estimator (IVE)

The usual version of the Instrumental Variables Estimator (IVE) is:

$$\hat{\beta}_n \in \arg\min_{\beta \in \Theta} \frac{1}{n^2}(Y_n - X_n\beta)'(Z_n W Z_n')(Y_n - X_n\beta)$$

Questions arise of the form:

- Does $\hat{\beta}_n$ depend on, and what are its properties based on, the choice of the constant matrix W?
- If yes, how can we choose the optimal W depending on the sample?

If we study the optimization problem a bit more, let's initially assume that $\Theta = \mathbb{R}^p$. We have:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{n^2}(Y_n - X_n\beta)'(Z_n W Z_n')(Y_n - X_n\beta)$$

This objective function is at least twice continuously differentiable. Since $\Theta = \mathbb{R}^p$, we can use first and second-order conditions.

4. Parenthesis: Matrix Derivatives

Let's say I have a function $G(\beta) : \mathbb{R}^p \to \mathbb{R}^n$. For example, $G(\beta) = Y_n - X_n\beta$. And an inner matrix $A_{n\times n}$ (e.g., $A = Z_n W Z_n'$).

We are interested in differentiating with respect to β the quadratic form:

$$G'(\beta)AG(\beta)$$

The first derivative with respect to β is:

$$\frac{\partial G'(\beta)AG(\beta)}{\partial \beta} = 2\frac{\partial G'(\beta)}{\partial \beta}AG(\beta)$$

Additionally, if G is linear with respect to β (which implies $\frac{\partial^2 G'(\beta)}{\partial \beta \partial \beta'} = 0_{p\times p}$), then the second derivative is:

$$\frac{\partial^2 G'(\beta)AG(\beta)}{\partial \beta \partial \beta'} = 2\frac{\partial G'(\beta)}{\partial \beta}A\frac{\partial G(\beta)}{\partial \beta'}$$

5. First-Order Conditions and Optimization

Building on our matrix derivative rules, we want to find the parameter that minimizes our objective function.

1st Order Conditions: We set the first derivative of the objective function $\mu_n(\beta, W)$ with respect to β equal to zero:

$$\frac{\partial \mu_n(\beta, W)}{\partial \beta} = 0_{p\times 1}$$

Applying our chain rule:

$$\iff \frac{1}{n^2} \cdot 2\frac{\partial (Y_n - X_n\beta)'}{\partial \beta}(Z_n W Z_n')(Y_n - X_n\beta) = 0_{p\times 1}$$

Let's evaluate the inner derivative:

$$\frac{\partial (Y_n - X_n\beta)'}{\partial \beta} = \frac{\partial (Y_n' - \beta' X_n')}{\partial \beta} = -\frac{\partial (\beta' X_n')}{\partial \beta} = -X_n'$$

Substituting this back into our first-order condition:

$$\frac{2}{n^2}(-X_n')(Z_n W Z_n')(Y_n - X_n\beta) = 0_{p\times 1}$$

By simplifying and rearranging the terms, we get:

$$X_n'(Z_n W Z_n')X_n\beta = X_n'(Z_n W Z_n')Y_n$$

(Note: This structure corresponds to the normal equations).

6. Matrix Invertibility and the IV Estimator

To solve for β, we must ensure the matrix on the left-hand side is invertible.

Since we assume:

- $rank(Z_n) = q$
- $rank(W) = q$
- $rank(X_n) = p$
- $rank(X_n' Z_n) = p$ *(Relevance Condition of Instruments)*

The matrix $(X_n' Z_n) W (Z_n' X_n)$ has rank p and is therefore invertible. (It is also positive definite).

Critical Point: Solving for the estimator gives us:

$$\hat{\beta}_n = (X_n' Z_n W Z_n' X_n)^{-1} X_n' Z_n W Z_n' Y_n$$

2nd Order Conditions: Checking the second derivative confirms we have a minimum:

$$\frac{\partial^2 \mu_n(\beta)}{\partial \beta \partial \beta'} = \frac{2}{n^2} X_n' Z_n W Z_n' X_n$$

Since the resulting matrix is positive definite, our critical point is indeed a minimum.

7. Properties of the IV Estimator

Therefore, in the parameter space $\Theta = \mathbb{R}^p$, with a specific choice of weighting matrix W, the Instrumental Variables Estimator (IVE) is:

$$\hat{\beta}_n = (X_n' Z_n W Z_n' X_n)^{-1} X_n' Z_n W Z_n' Y_n$$

We know the true data-generating process is $Y_n = X_n \beta_0 + \epsilon_n$. Substituting Y_n into our estimator:

$$\hat{\beta}_n = (X_n' Z_n W Z_n' X_n)^{-1} X_n' Z_n W Z_n' (X_n \beta_0 + \epsilon_n)$$
$$\implies \hat{\beta}_n = \beta_0 + (X_n' Z_n W Z_n' X_n)^{-1} X_n' Z_n W Z_n' \epsilon_n$$

Is it Unbiased?

We don't know yet, we shall see. Taking the expectation:

$$\mathbb{E}(\hat{\beta}_n) = \beta_0 + \mathbb{E}\left[(X_n' Z_n W Z_n' X_n)^{-1} X_n' Z_n W Z_n' \epsilon_n\right]$$

(Note: Evaluating this expectation to $0_{p\times 1}$ is not straightforward due to the non-linear way the random variables interact inside the inverse matrix).