Microeconometrics

Based on the textbooks
VERBEEK: *A Guide to Modern Econometrics*
and CAMERON AND TRIVEDI: *Microeconometrics*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

November 7, 2013

Outline

Strict exogeneity violated

In many empirical situations, the strict exogeneity condition (A2) cannot hold. Consider four main cases:

- If, with time-series data $i = t$, some regressors are lagged dependent variables, y_t depends on ε_t, so X and ε cannot be independent;
- If some regressors are observed with an error u_t that is independent of the pure regression error v, the overall error has a component that is correlated with X;
- If some regressors depend on missing or unobserved information, there will be dependence between X and ε: narrow-sense *endogeneity*;
- If some regressors $x_{i,j}$ logically depend on the current regressand y_i, X and ε will be dependent: *feedback*.

Endogeneity impairs OLS consistency

Considering the limit of the OLS estimate,

$$\text{plim}\hat{\beta} = \beta + \text{plim}(N^{-1}X'X)^{-1}N^{-1}X'\varepsilon,$$

it is seen that (A1), (A4), (A5) together with

A8 x_i and ε_i are independent for all i;

A11 ε_i is $IID(0, \sigma^2)$;

suffices for OLS consistency and asymptotic normality, as LLN and CLT can be applied. For outright endogeneity, however, even this weaker set of consistency conditions will be violated. OLS will become inconsistent.

Even weaker conditions

It can be shown that with the assumptions

A7 $\mathrm{E}(x_i \varepsilon_i) = 0 \quad \forall i$;

A12 ε_i is serially uncorrelated and $\mathrm{E}(\varepsilon_i) = 0$;

OLS remains consistent, assuming some technical regularity conditions. This weak set of conditions is important for time-series data, it admits time-changing variance and conditional heteroskedasticity. The classical time-series concept of *predeterminedness* is stronger: regressors are uncorrelated with current and future errors. With endogeneity, even these conditions are violated.

Bias due to an unobserved omitted variable

Presume the true relationship is

$$y_i = x_{1i}'\beta_1 + x_{2i}\beta_2 + \gamma u_i + \nu_i,$$

with observed regressors x_1 (a vector) and x_2 (a scalar) and unobserved u. Further, assume that $\mathrm{corr}(x_2, u) \neq 0$. Presume the estimated relationship is

$$y_i = x_{1i}'\beta_1 + x_{2i}\beta_2 + \varepsilon_i = x_i'\beta + \varepsilon_i.$$

True $\varepsilon_i = \gamma u_i + \nu_i$, with $\mathrm{cov}(\varepsilon_i, x_{2i}) \neq 0$. OLS imposes $\mathrm{cov}(x, e) = 0$ and $\gamma = 0$, which creates a bias:

$$b = \beta + (X'X)^{-1}X'(\gamma u_i + \nu_i) = \beta + (X'X)^{-1}X'(\gamma u_i) + (X'X)^{-1}X'\nu_i.$$

Biased but not incorrect

Note that OLS in the regression omitting u and in the generating model that includes u measure different objects:

- The coefficient β_2 in the regression omitting u is the marginal reaction of y to changes in x_2 within the sample, changing u implicitly along with x_2. [Talented individuals tend to be interested in extending their education time];
- The coefficient β_2 in the virtual regression that includes u is the marginal reaction of y to changes in x_2 while keeping u constant. This version may be of interest in causal interpretations and policy implementations [By how much would a person's wage increase if she without increasing her innate skills and talents enjoyed longer education].

Reverse causality

Consider the simple regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.$$

If the researcher ignores the true relationship

$$x_i = y_i + z_i,$$

$\mathrm{cov}(x, \varepsilon) \neq 0$ and the OLS estimate b will be biased.

Assume z is observed and $\mathrm{cov}(z, \varepsilon) = 0$. Then, regressing y on z yields an unbiased estimate, albeit for a possibly less interesting parameter.

Indirect least squares

Substituting the equation for x into the one for y yields

$$y_i = \beta_0 + \beta_1(y_i + z_i) + \varepsilon_i$$

or

$$y_i = \frac{\beta_0}{1-\beta_1} + \frac{\beta_1}{1-\beta_1} z_i + \varepsilon_i^* = \gamma_0 + \gamma_1 z_i + \varepsilon_i,$$

a standard regression satisfying Gauss-Markov. Estimating the coefficients γ_0, γ_1 by OLS as g_0, g_1 and retrieving the parameters of interest from

$$\frac{b_0}{1-b_1} = g_0, \quad \frac{b_1}{1-b_1} = g_1,$$

implies consistent estimates for β_0, β_1: *indirect least squares*. The method fails when the analytical equation system becomes intractable.

Structural and reduced form

In the presence of feedback or reverse causality, it is often possible to transform the 'bad' equation with endogenous regressors into a 'good' equation with y depending on exogenous variables only. This **reduced form** does not satisfy the researcher's need, as she may be interested in the parameters of the original **structural form**.

Appropriate estimation with endogenous regressors

There have been several suggestions in the literature:

- *Indirect least squares*: estimate the reduced form and determine estimates for the structural parameters by solving equations analytically: in most applications infeasible;
- *Two-stage least squares*: estimate reduced forms for all endogenous variables, replace endogenous regressors by the fitted values, i.e. linear combinations of exogenous variables, finally regress y on the fitted values: the dominant method for some time, now seen as a special case of IV;
- *Instrumental variables* (IV): needs exogenous variables ('instruments') that correlate with endogenous regressors;
- *Generalized method of moments* (GMM): generalizes IV by viewing it as a method of moments. Arbitrary restrictions can be easily imposed by a general scheme.

Instrumental variables: the idea

For the regressors $x_1, \ldots, x_K$ with potential endogeneity problems, there exist K exogenous variables $z_1, \ldots, z_K$, such that

$$\mathrm{cov}(Z, X) \neq 0, \quad \mathrm{cov}(Z, \varepsilon) = 0.$$

Then, the estimator

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'y$$

has good properties, such as consistency, asymptotic normality, and even unbiasedness (for appropriate assumptions on Z).

IV as a method of moments

OLS can be derived from population conditions $\mathrm{E}(x_j \varepsilon) = 0$ for all regressors and the corresponding sample condition $X'e = 0$. Similarly, the population **exclusion restrictions**

$$\mathrm{E}(z_j \varepsilon) = 0$$

for all instruments yield the sample conditions

$$\frac{1}{N} \sum_{i=1}^{N} (y_i - x_i' \hat{\beta}_{IV}) z_{ji} = 0,$$

which yields the **instrumental variables estimator** $\hat{\beta}_{IV}$ as

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'y = \left(\sum_{i=1}^{N} z_i x_i' \right)^{-1} \sum_{i=1}^{N} z_i y_i,$$

for the K–vectors z_i and x_i.

Asymptotic properties of IV

It can be shown that, under valid exclusion restrictions, with the condition for *valid instruments*

$$\text{plim} \frac{1}{N} Z'X = \Sigma_{zx},$$

a finite and nonsingular matrix, assuming (A11), and

$$\text{plim} \frac{1}{N} Z'Z = \Sigma_{zz},$$

a finite and nonsingular matrix, it follows that

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 (\Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx})^{-1}).$$

Estimating the variance matrix

The usual suggestion is to estimate the variance of the IV estimate by

$$\hat{\mathrm{V}}\{\hat{\beta}_{IV}\} = \hat{\sigma}^2 \{X'Z(Z'Z)^{-1}Z'X\}^{-1},$$

with

$$\hat{\sigma}^2 = \frac{1}{N-K} \sum_{i=1}^{N} (y_i - x_i'\hat{\beta}_{IV})^2.$$

There are also robust variants, in the sense of White-Eicker.

Testing for the IV assumptions

- The moment conditions or exclusion restrictions cannot be tested statistically. They are identifying. Only if there are more restrictions than parameters (generalized IV, GMM), the over-identifying restrictions can be tested;
- The endogeneity of regressors can be tested on the basis of a *Hausman test*: if exogeneity holds, OLS and IV approach the same limit, as OLS is valid, consistent, and linear efficient. If exogeneity does not hold, the limits will differ. A variant of the Hausman test statistic is the t–statistic in an OLS regression of y on all regressors $x_1, \ldots, x_K$ and on the residual from an auxiliary regression of the doubtful regressor on reliable variables (regressors and instruments). Insignificance supports exogeneity.

The case of many instruments

Suppose there exist $R > K$ instruments and $R > K$ exclusion restrictions $\mathrm{E}(z_j \varepsilon) = 0$, all of which are valid in population. In the sample, the corresponding sample moment conditions

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\hat{\beta}_{IV})z_{ji} = 0, \quad j = 1, \ldots, R,$$

form an over-determined equation system and do not have an exact solution. The way out is to minimize a quadratic form

$$Q_N(\beta) = \left\{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\beta)z_i\right\}' W_N \left\{\frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\beta)z_i\right\}.$$

The positive definite weighting matrix W_N may depend on N.

Under- and over-identification

Consider the three cases:

- $R < K$: R equations (exclusion restrictions) do not uniquely determine K variables (coefficients). The solution is a subspace, not a value. As $N \to \infty$ the subspace will contain the true value, but this cannot be called consistency: the case of *under-identification*;
- $R = K$: $R = K$ equations uniquely determine K coefficients. This is the narrow-sense IV estimator, which is consistent: the case of *exact identification*;
- $R > K$: there is no solution. If the R exclusions hold in population, the quadratic form Q_N will converge to 0, the **generalized instrumental-variables estimator** (GIVE) is consistent: the case of *over-identification*.

The formal representation of GIVE

Evaluate the derivative of the expression

$$Q_N(\beta) = \left\{ \frac{1}{N} Z'(y - X\beta) \right\}' W_N \left\{ \frac{1}{N} Z'(y - X\beta) \right\}$$

with respect to β and equate it to 0. This yields the solution

$$\hat{\beta}_{IV} = (X'ZW_NZ'X)^{-1} X'ZW_NZ'y.$$

For $R = K$ and for non-singular $X'Z$, the narrow-sense IV formula is retrieved.

Choosing the weighting matrix

It can be shown that $W_N = (Z'Z)^{-1}$ minimizes the variance of (wide-sense) GIVE. The resulting estimator

$$\hat{\beta}_{IV} = \{X'Z(Z'Z)^{-1}Z'X\}^{-1}X'Z(Z'Z)^{-1}Z'y.$$

is often referred to as (narrow-sense) GIVE. Others call it the **two-stage least squares estimator**, as it can be obtained in two steps:

1. Regress X on all instruments Z. Truly exogenous regressors are their own instruments. Keep the fitted values $\hat{X}$, which are 'exogenous';
2. Regress y on $\hat{X}$. Statistics in this regression (standard errors, R^2) will be incorrect.

The variance of GIVE

Assuming the usual model $y = X\beta + \varepsilon$ and $\mathrm{V}(\varepsilon) = \sigma^2 I_N$, it is easily shown that

$$\mathrm{V}(\hat{\beta}_{IV}|Z, X) = \sigma^2 \{X'Z(Z'Z)^{-1}Z'X\}^{-1}.$$

The unknown σ^2 can be estimated by

$$\hat{\sigma}^2 = \frac{1}{N-K} \sum_{i=1}^{N} \hat{\varepsilon}_i^2,$$

with $\hat{\varepsilon}$ denoting the GIVE residuals.

Testing over-identifying restrictions

If there are $R > K$ exclusion restrictions, the $R - K$ over-identifying restrictions may be invalid. If they are valid (the null hypothesis), the **Sargan statistic** is asymptotically distributed $\chi^2(R - K)$. It is uncertain which restrictions are rejected, however.

The Sargan statistic can be obtained as an NR^2 version of a LM test: the GIVE residuals $\hat{\varepsilon}$ are regressed by OLS on all R instruments, keep the corresponding R^2.

Weak instruments

If the instruments z_j are only loosely related to the endogenous regressors, the precision of the IV/GIVE estimation will be poor. Tests, particularly the Hausman test, will also be unreliable.

STOCK AND WATSON recommend a rule of thumb: run reduced-form regressions of the x_j on the good (guaranteed exogenous) regressors and on the instruments. If the F–statistic on joint exclusion of the instruments is less than 10, the instruments can be regarded as *weak*.

Generalized method of moments: the concept

The **generalized method of moments** (GMM) generalizes IV/GIVE to nonlinear and implicit functions. Assume an economic model is characterized by $R > K$ moment conditions

$$\mathrm{E}\{f(w, z, \theta)\} = 0,$$

with $f : \mathbb{R}^{M_1+M_2+K} \to \mathbb{R}^R$. w contains observed variables, z are instruments, and θ is the K—vector of parameters. Under regularity conditions, the sample counterpart

$$g_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} f(w_i, z_i, \theta)$$

can be taken as close to 0 as possible, which then defines the GMM estimator $\hat{\theta}_{GMM}$.

GMM and the weighting matrix

Formally, looking for the value that takes the sample moment conditions closest to an R–vector of 0s is implemented by minimizing (numerically) the quadratic form

$$Q_N(\theta) = g_N(\theta)' W_N g_N(\theta).$$

The minimizer $\hat{\theta}_{GMM}$ may depend on the choice of the weighting matrix. One can show that the asymptotically optimal weighting matrix is

$$W^{opt} = [\mathrm{E}\{f(w, z, \theta) f(w, z, \theta)'\}]^{-1}.$$

Obviously, this matrix is not available as θ is unknown.

Estimating the optimal weighting matrix

Replacing the expectation operator by the sample mean and θ by a consistent preliminary estimate $\hat{\theta}_0$ yields the estimated optimal weighting matrix

$$W_N^{opt} = \left[\frac{1}{N}\sum_{i=1}^{N} f(w_i, z_i, \hat{\theta}_0) f(w_i, z_i, \hat{\theta}_0)'\right]^{-1}$$

A popular suggestion for $\hat{\theta}_0$ is GMM with the weighting matrix I_R,. This 'two-step GMM' method can be iterated by updating W_N with the most recent availabe θ estimate. The iterated estimator is asymptotically equivalent to the two-step GMM but may improve small-sample properties.

Asymptotic properties of GMM

Efficient GMM estimators (two-step and iterated) have the property that

$$\sqrt{N}(\hat{\theta}_{GMM} - \theta) \xrightarrow{d} \mathcal{N}(0, V)$$

with

$$V = (DW^{opt}D')^{-1},$$

where

$$D = \mathrm{E}\left\{ \frac{\partial f(w, z, \theta)}{\partial \theta'} \right\}$$

is the $K \times R$ matrix of derivatives. Unless f has a simple analytic derivative, D must be determined numerically.

Conditions for GMM properties

The asymptotic properties require several technical conditions, such as:

- The population moment conditions hold in the true generating model;
- The function $f(., \theta)$ does not yield the same value for different values of θ;
- The weight matrix converges to a finite and positive definite limit;
- Derivatives of $f(., \theta)$ w.r.t. θ exist and their sample averages converge to D;
- The limit of the sample variance matrix of f exists and is positive definite.

Testing over-identifying restrictions in GMM

In analogy to GIVE, the Sargan test statistic

$$\xi = N g_N(\hat{\theta}_{GMM})' W^{opt} g_N(\hat{\theta}_{GMM})$$

is used to test over-identifying restrictions. Under the null, ξ is asymptotically distributed as $\chi^2(R-K)$.

The history of GMM

GMM is mainly due to Lars Peter Hansen who contributed on GMM in 1982. The idea has spread quickly in the following years and decades, as

- Economic theory models routinely contain moment conditions. Confidence in the correctness of these models may have increased in the recent decades;
- Semiparametric methods avoid specification of statistical distributions and may require the specification of functional forms. This preference may correspond to the needs of many economists.

GMM has gained so much ground relative to estimation methods based on maximizing the likelihood (ML) that the expression 'GMM revolution' may be appropriate.