Recent Advances in the Field of Trade Theory and Policy Analysis Using Micro-Level Data

July 2012
Bangkok, Thailand

Cosimo Beverelli (World Trade Organization)
a) Endogeneity
b) Instrumental variables
c) Instrumental variables in practice
d) Example with firm-level analysis
a) Endogeneity

- Definition and sources of endogeneity
- Inconsistency of OLS
- Example with omitted variable bias
Definition and sources of endogeneity

- A regressor is endogenous when it is correlated with the error term.
- Leading examples of endogeneity:
  a) Reverse causality
  b) Omitted variable bias
  c) Measurement error bias
  d) Sample selection bias

- In case a), there is a two-way causal effect between $y$ and $x$. Since $x$ depends on $y$, $x$ is correlated with the error term (endogenous) and $x$.
- In case b), the omitted variable is included in the error term. If $x$ is correlated with the omitted variable, it is correlated with the error term (endogenous).
- In case c), under the classical errors-in-variables (CEV) assumption (measurement error uncorrelated with unobserved variable but correlated with the observed-with-error one), the observed-with-error variable is correlated with the error term (endogenous).
Inconsistency of OLS

• In the model:

\[ y = X\beta + u \quad (1) \]

The OLS estimator of \( \beta \) is consistent is the true model is (1) and if \( \text{plim}(N^{-1}X'u) = 0 \)

• Then:

\[ \text{plim}(\hat{\beta}) = \beta + \text{plim}(N^{-1}X'X)^{-1}\text{plim}(N^{-1}X'u) = \beta \]

• If, however, \( \text{plim}(N^{-1}X'u) \neq 0 \) (endogeneity), OLS estimator of \( \beta \) is inconsistent

• The direction of the bias depends on whether correlation between \( X \) and \( u \) is positive (upward bias, \( \hat{\beta} > \beta \)) or negative (\( \hat{\beta} < \beta \))
Example with omitted variable bias

- True model is:
  \[ y = x' \beta + z\alpha + \nu \]

- Estimated model is:
  \[ y = x' \beta + (z\alpha + \nu) = x' \beta + \varepsilon \]

- From OLS estimation:
  \[ \text{plim}(\hat{\beta}) = \beta + \delta \alpha \]

Where \( \delta = \text{plim}[(N^{-1}X'X)^{-1}(N^{-1}X'z)] \)

- If \( \delta \neq 0 \) (the omitted variable is correlated with the included regressors), the basic OLS assumption that the error term and the regressors are uncorrelated is violated, and the OLS estimator of \( \beta \) will be inconsistent (omitted variable bias)
The direction of the omitted variable bias can be established, knowing what variable is being omitted, how it is correlated with the included regressor and how it may affect the LHS variable.

If correlation between the omitted variable and the included regressor is positive ($\delta > 0$) and the effect of the omitted variable on $y$ ($\alpha$) is supposedly positive, $\delta \alpha > 0$ and the bias is positive:

- $\hat{\beta}$ is overestimated

The same is true if both $\delta$ and $\alpha$ are negative.

If $\delta$ and $\alpha$ have opposite signs, the bias is negative:

- $\hat{\beta}$ is underestimated
Example with omitted variable bias (ct’d)

• Standard textbook example: returns to schooling
• We want to estimate the effect of schooling on earnings
• We omit the variable “ability”, on which we do not have information...
• ...But ability is positively correlated with schooling
• OLS regression will yield inconsistent parameter estimates
• Since ability should positively affect earnings, the omitted variable bias is positive
• OLS of earnings on schooling will overstate the effect of education on earnings
b) Instrumental variables

• Visual representation
• Definition of an instrument
• Examples of instrumental variables
• Example with market demand and supply
• Instrumental variables in multiple regression
• Identification issues
• The instrumental variable (IV) estimator
• IV estimator as two-stages least squares (2SLS)
• OLS is consistent if

• OLS is inconsistent if

• We need a method to generate exogenous variation in $x$
• Randomized experiment is the first best...
• ...In the absence of randomized experiment, we can use an instrument $z$ that has the property that changes in $z$ are associated with changes in $x$ but do not lead to changes in $y$
Definition of an instrument

- A variable $z$ is called an instrument (or instrumental variable) for regressor $x$ in the scalar regression model $y = x\beta + u$ if:
  
  a) $z$ is uncorrelated with $u$
  
  b) $z$ is correlated with $x$

- Condition a) excludes $z$ from directly affecting $y$
  
  - If this was not the case, $z$ would be in the error term in a regression of $y$ on $x$, therefore $z$ would be correlated with the error term

- The instrument should not affect $y$ directly, but only indirectly, through its effect on $x$
Examples of instrumental variables

• In the returns to schooling example, good candidates for \( z \) (uncorrelated with ability – and not directly affecting earnings – and correlated with schooling) are proximity to college and month of birth

Examples from gravity equations with reverse causality (see [here](#) for details)

• In a gravity estimation of the effect of trading time on trade, an instrument is needed that is correlated with trading time and that affects trade only indirectly, through its impact on trading time
  • Number of administrative formalities (documents)
  • Trading times using times in neighboring countries

• In a gravity estimation of the effect of contract enforcement on trade, an instrument is needed that is correlated with contract enforcement and that affects trade only indirectly, through its impact on contract enforcement
  • Settlers’ mortality
Example with market demand and supply

• The IV method was originally developed to estimate demand elasticity for agricultural goods, for example milk:

\[ \ln(Q_t) = \beta_0 + \beta_1 \ln(P_t) + u_t \]

• OLS regression of \( \ln(Q_t) \) on \( \ln(P_t) \) suffers from endogeneity bias
  • Price and quantity are simultaneously determined by the interaction of demand and supply

(a) Demand and supply in three time periods
Example with market demand and supply (ct’d)

- The interaction between demand and supply could reasonably produce something not useful for the purpose of estimating the price elasticity of demand.
Example with market demand and supply (ct’d)

• But, what if only supply shifts?

• The instrument $Z$ is a variable that affects supply but not demand

• The IV method estimates the elasticity of the demand curve by isolating shifts in price and quantity that arise from shifts in supply
Example with market demand and supply (ct’d)

• An ideal candidate for $Z$ is rainfall in dairy-producing regions:

a) We can *reasonably* assume that rainfall in dairy-producing regions does not directly affect demand for milk (exogeneity condition)

b) We can *reasonably* assume that insufficient rainfall lowers food available to cows and milk production as a consequence (relevance condition)
Consider the general regression model:

$$y = x'\beta + u$$

Where $x$ is $K \times 1$

- Some components of $x$ are endogenous (endogenous regressors): $x_1$
- Some components of $x$ are exogenous (exogenous regressors): $x_2$
- Partition $x$ as $[x'_1 x'_2]'$
- Instruments are needed for the endogenous regressors (in $x_1$), while exogenous regressors (in $x_2$) can be instruments for themselves
- Assume we have an $R \times 1$ vector of instruments $z$ that satisfies the conditions for being a good instrument
- We can then use $z = [z'_1 x'_2]'$ as an instrument for $x = [x'_1 x'_2]'$
Identification issues

• Identification requires $R \geq K$ (number of instruments must be at least equal to the number of endogenous regressors)

• If $R = K$, the model is just-identified
  • For instance, there are two endogenous variables and two instruments

• If $R > K$, the model is overidentified
  • For instance, there is one endogenous variable and two instruments

• Overidentification is desirable because only if the model is overidentified one can test for instruments’ exogeneity and excludability
  • This is Hansen’s J test (see below)
The instrumental variable (IV) estimator

• For the general model

\[ y = X\beta + u \]

where \( X \) contains endogenous regressors, construct the matrix of instruments \( Z \)

• For \( Z \) to be valid, it must be that:

a) \[ \operatorname{plim}(N^{-1}Z'X) = \Sigma_{ZX}, \] a finite matrix of full rank

b) \[ \operatorname{plim}(N^{-1}Z'u) = 0 \]

• Premultiply by \( Z' \), apply GLS to obtain:

\[ \hat{\beta}_{IV} = (X'P_ZX)^{-1}X'P_Zy \] (2)

where \( P_Z = Z(Z'Z)^{-1}Z' \) is \( Z \)'s projection matrix

• In the just-identified case, \( \hat{\beta}_{IV} = (Z'X)^{-1}Z'y \) (3)
The instrumental variable (IV) estimator (ct’d)

• The IV estimator is consistent
• Take the just-identified case

\[ \hat{\beta}_{IV} = (Z'X)^{-1}Z'(X\beta + u) = \beta + (N^{-1}Z'X)^{-1}(N^{-1}Z'u) \]

• Under assumptions a) and b) in previous slide, the IV estimator is consistent

• The asymptotic VCV matrix is given in Cameron and Trivedi (2005), respectively in expression 4.55 (p. 102) for the estimator in (2) and in expression 4.52 (p. 101) for the estimator in (3)
IV estimator as two-stages least squares (2SLS)

• The IV estimator in (2) can be seen as the result of a double application of least squares:

1. Regress each of the variables in the $X$ matrix on $Z$, and obtain a matrix of fitted values $\hat{X}$:

   $\hat{X} = P_Z X$

2. Regress $y$ on $\hat{X}$ to obtain:

   $\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y = (X'P_ZX)^{-1}X'P_Zy = \hat{\beta}_{IV}$
IV estimator as two-stages least squares (2SLS) (ct’d)

• Intuitively, the first stage “cleanses” the endogeneity from the variables we are worried about. By using predicted values based on genuinely exogenous variables only, we obtain the exogenous part of their variation.

• Consider the example of milk demand:
  1. The predicted value of OLS regression of milk price on rainfall is the milk price that isolates changes in price itself due to the supply side of the economy (partially, at least).
  2. OLS regression of milk quantity on the predicted milk price is the regression counterpart of using shifts in the supply curve to identify the demand curve.

• In practice, avoid doing two stages manually:
  • You will get incorrect standard errors (too small), and you might mistakenly exclude exogenous variables from the main model.

• IV estimator can also be derived using GMM (one-step, not optimal GMM).
c) Instrumental variables in practice

- Overview
- Weak instruments
- Specification tests
  - Testing for endogeneity
  - Testing for overidentifying restrictions
- Summing up
Overview

• IV estimators are less efficient than the OLS estimator
• They are biased in finite samples, even if asymptotically consistent
• This finite sample bias is there even in relatively large samples
• Most importantly, in the presence of weak instruments, IV estimation estimator can actually produce worse (less consistent) results than simple OLS even in large samples
• So the first step in testing must be to ensure that the instruments are strongly enough correlated with the potentially endogenous variables

• Specification tests for endogeneity and overidentifying restrictions exist, but they have limitations
  • In particular, the test of overidentifying restrictions cannot be carried out in a just-identified regression
• If instruments are weak, these tests can produce misleading results
Weak instruments

- The weak-instruments problem arises when the correlations between the endogenous regressors and the excluded instruments are non-zero but small.
- The weak-instruments problem can arise even when the correlations between $X$ and $Z$ are significant at conventional levels (5% or 1%) and the researcher is using a large sample.
- Under weak instruments, even mild endogeneity of the instrument can lead to IV parameter estimates that are much more inconsistent than OLS.
- Example with one endogenous regressor ($y = x\beta + u$), one instrument ($z$) and iid errors:

\[
\frac{\text{plim } \hat{\beta}_{IV} - \beta}{\text{plim } \hat{\beta}_{OLS} - \beta} = \frac{\text{corr}(z,u)}{\text{corr}(x,u)} \times \frac{1}{\text{corr}(z,x)}
\]

- Thus with an invalid instrument $\text{corr}(z,u) \neq 0$ and low $\text{corr}(z,x)$ the IV estimator can be even more inconsistent than OLS.
Weak instruments (ct’d)

- Informal rules of thumb exist to detect weak instruments problems
  - Partial $R^2$
  - Partial $F$ statistics ($F$ test of the excluded instruments in the corresponding first-stage regression) – Staiger and Stock’s rule of thumb is that is should be greater than 10

- More formal criteria: Stock-Yogo weak instruments tests ($H_0$: Instruments are weak)
  - Comparison on bias of OLS and bias of IV
  - Wald test

- Anderson-Rubin and Stock-Wright tests
  - Null hypothesis that the coefficients of the endogenous regressors in the structural equation are jointly equal to zero (so that overidentifying restrictions are valid)
  - These tests are robust to the presence of weak instruments
Low precision

• Although IV estimation can lead to consistent estimation when OLS is inconsistent, it also leads to a loss in precision
• Example with one endogenous regressor, one instrument and iid errors

\[ \text{Var}(\hat{\beta}_{IV}) = \frac{\text{Var}(\hat{\beta}_{OLS})}{r^2_{xz}} \]

• The IV estimator has a larger variance unless \( \text{corr}(x, z) = 1 \)
• If the squared sample correlation coefficient between \( z \) and \( x \) is 0.1, IV standard errors are 10 times those of OLS
• Therefore, weak instruments exacerbate the loss in precision
Testing for endogeneity

• Endogeneity test: is there evidence that correlation between the potentially endogenous variables and the error term is strong enough to result in substantively biased OLS estimates?

• We can test for the endogeneity of suspect independent variables using a Hausman test

• Consider the model

$$y = x'_1 \beta_1 + x'_2 \beta_2 + u$$  \hspace{1cm} (4)

where \(x_1\) is potentially endogenous and \(x_2\) is exogenous

• The Hausman test of endogeneity can be calculated by testing \(\gamma = 0\) in the augmented OLS regression

$$y = x'_1 \beta_1 + x'_2 \beta_2 + \widehat{x}'_1 \gamma + u$$  \hspace{1cm} (5)

or (equivalently) in the augmented OLS regression

$$y = x'_1 \beta_1 + x'_2 \beta_2 + \widehat{v}'_1 \gamma + u$$  \hspace{1cm} (6)
Testing for endogeneity (ct’d)

• In equation (5), $\hat{x}_1$ is the predicted value of endogenous regressors $\hat{x}_1$ from an OLS regression of $x_1$ on the instruments $z$

• In equation (6), $\hat{v}_1$ is the residual from an OLS regression of $x_1$ on the instruments $z$

• Intuitively, if the error term $u$ in equation (4) is uncorrelated with $x_1$ and $x_2$, then $\gamma = 0$

• If, instead, the error term $u$ in equation (4) is correlated with $x_1$, this will be picked up by significance of additional transformations of $x_1$, such as $\hat{x}_1$ (equation 5) or $\hat{v}_1$ (equation 6)

• Rejection of the null hypothesis $H_0: \gamma = 0$ indicates endogeneity
Testing for overidentifying restrictions

- The instruments must be exogenous for the IV estimator to be consistent. For overidentified models, a test of instruments’ exogeneity is possible.
- This is Hansen’s J test of overidentifying restrictions.
- Derivations are based on GMM theory and can be found here (pp. 16-18).

- Intuitively, in the model $y = x' \beta + u$, instruments $z$ are valid if $E[u|z] = 0$ or if $E[zu] = 0$.
- A test of $H_0: E[zu] = 0$ is naturally based on departures of $N^{-1} \sum_i z_i \hat{u}_i$ from zero.
- In the just-identified case, IV solves $N^{-1} \sum_i z_i \hat{u}_i = 0$ so this test is not useful.
- In the overidentified case, the Hansen’s J test is:
  $$J = \hat{u}' Z \hat{S}^{-1} Z' \hat{u}$$
  Where $\hat{u}$ comes from optimal GMM estimation and $\hat{S}$ is a weighting matrix.
Testing for overidentifying restrictions (ct’d)

• This is an extension of the Sargan test that is robust to heteroskedasticity and clustering
• Large J leads to rejection of the null hypothesis that the instruments satisfy orthogonality conditions
• This may be due to:
  • Endogeneity: instruments are correlated with the main equation errors because there is feedback running from the dependent variable to the instruments; and/or
  • Non-excludability: the instruments should appear in the main regression, and the test is effectively picking up an omitted variables problem
• Hansen’s J should not reject the null for instruments to be exogenous and excludable
• An important limitation of the J test is that it requires that the investigator believes that at least some instruments are valid
• You can also test for subsets of overidentifying restrictions
Summing up

• Always test for instrument relevance first
• If instruments are weak, the cure (IV) can be much worse than the disease (inconsistency of OLS)
• Use the Hausman test to assess the extent to which endogeneity is really a problem
• If at all possible, ensure the model is overidentified, and test exogeneity and excludability via Hansen’s J
• If the model passes all of these tests, then it should provide a reasonable guide for causal inference
d) Example with firm-level analysis

- Topalova and Khandelwal (2011)
The problem

- The authors are interested in the effect of trade reform on firm-level productivity in a sample of Indian firms
- Endogeneity concerns for the productivity effect of trade policy:
  - Governments may reduce tariffs only after domestic firms have improved productivity, which would result in a spurious relationship between trade and productivity
  - Selective protection of industries (tariffs may be adjusted in response to industry productivity levels)
  - If policy decisions on tariff changes across industries were indeed based on expected future productivity or on industry lobbying, isolating the impact of the tariff changes would be difficult. Simply comparing productivity in liberalized industries to productivity in non-liberalized industries would possibly give a spurious correlation between total factor productivity (TFP) growth and trade policies
The solution

• Since 1991, over a short period of time, India drastically reduced tariffs and narrowed the dispersion in tariffs across sectors

• Since the reform was rapid, comprehensive, and externally imposed (IMF), it is reasonable to assume that the changes in the level of protectionism were unrelated to firm- and industry-level productivity

• However, at the time the government announced the export-import policy in the Ninth Plan (1997-2002), the sweeping reforms outlined in the previous plan had been undertaken and pressure for further reforms from external sources had abated

• More difficult to isolate the causal impact of tariff changes
The solution (ct’d)

• The authors address the concern of possible endogeneity of trade policy in 3 ways:

1. Examining the extent to which tariffs moved together
   • Tariff movements were uniform until 1997 and less uniform afterwards, indicating a more pronounced problem of endogenous trade protection in the second period

2. Testing whether protection correlates with industry characteristics (employment, output, average wage, concentration etc.)
   • No statistical correlation (indication of exogeneity)

...
3. Investigating whether policymakers adjusted tariffs in response to industry's productivity levels
   • The correlation between future trade protection and current productivity is indistinguishable from zero for the 1989-96 period
   • The pattern, however, is quite different for the 1997-2001 period. Here, the coefficient on current productivity is negative and significant, suggesting that trade policy may have been adjusted to reflect industries’ relative performance
The solution (ct’d)

• These tests lead to conclude that trade policy was *not* endogenously determined *during the first period*

• The 1991 liberalization episode in India is good to examine the causal effects of trade reform on firm-level productivity
Results

• The main result: 10% reduction in tariffs will lead to about 0.5% increase in firm TFP. Decreasing trade protection in the form of lower tariffs raises productivity at the firm level

• There are two forces driving this finding
  1. Increases in competition resulting from lower output tariffs caused firms to increase their efficiency
  2. The trade reform lowered the tariffs on inputs, which lead to an increase in the number and volume of imported inputs from abroad

• The larger impact appears to have come from increased access to foreign inputs. Thus, India’s break from import substitution policies not only exposed these firms to competitive pressures, but more importantly, relaxed the technological constraint on production
Results (ct’d)

• Melitz (2003) has shown that trade liberalization may result in a reallocation from low- to high-productivity firms which would increase average productivity because of selection.

• Re-estimating the equation only for the set of companies in operation in 1996, the positive impact of tariff reductions on productivity levels is virtually unchanged.

• This constitutes some mild evidence against the selection channel.

• While the exit of less efficient companies might contribute to productivity improvements, it does not drive the results within this sample.