Innovation and Productivity in SMEs. Empirical Evidence for Italy

discussed by Alessandro Sembenelli (Università di Torino and Collegio Carlo Alberto)

What is this paper about?

- Very important research area which desperately needs high quality research (especially in this country)
- Application of the so-called “CDM three-equation model” to a sample of small and medium size firms (SMEs)
  - R&D Equation
  - Innovation Equation (Knowledge Production Function)
  - Productivity Equation
- Specification of the model to accomodate some SMEs known features

The structure of the model

- First step (R&D equation)
  - Selection (Tobit type II) model
- Second step (Innovation equation)
  - Bivariate probit model (product and process innovation)
- Third step (Productivity equation)
  - Standard linear model

Data

- Data come from three consecutive waves of the Mediocredito-Capitalia-Unicredit “Survey on Manufacturing Firms”.
- Unbalanced panel of 7,375 firms of which only 361 are present in all three waves.
- Since data are “essentially cross-sectional” no effort is made to control for unobserved firm heterogeneity (compariosn with Parisi et al (2006)).

Plan of the discussion

- Specification issues
- What do we learn?
- Suggestions

Specification Issues (1)

- Only on the sub-sample of R&D performing firms (Pooled OLS)

$$RD = z\beta + e$$

- On all firms (Pooled Bivariate probit)

$$PROD = x_1\gamma_1 + \delta_1 \underline{RD} + u_1$$

$$PROC = x_2\gamma_2 + \delta_2 \underline{RD} + u_2$$

- On all firms (Pooled OLS)

$$y = w\pi + \underline{PROD}\eta_1 + \underline{PROC}\ \eta_2 + v$$

Specification issues (2)

- R&D Equation

$$RD=z\beta+e$$

  - Unobserved heterogeneity (e.g. managerial quality) is unlikley to be orthogonal to some (all?) of the regressors in z, including size, firm geographical market and, obviously, the amount of received subsidies.
  - Does it make sense to predict R&D effort also for those firms that have declared zero R&D Investment? Robustness check?

Specification Issues (3)

- Innovation Equations

$$PROD = x_1\gamma_1 + \delta_1 \underline{RD} + u_1$$

$$PROC = x_2\gamma_2 + \delta_2 \underline{RD} + u_2$$

  - Unobserved heterogeneity problem (again). How can Investment be treated as exogenous?
  - Unobserved heterogenity also invalidates the exogeneity of predicted R&D unless we assume that error terms in the innovation equation are orthogonal to the regressors in the R&D equation.
  - Exclusion restriction for Investment is not necessary. More generally a discussion of exclusion restrictions is needed.

Specification issues (4)

- Productivity Equation

$$y = w\pi + \underline{PROD}\eta_1 + \underline{PROC}\,\eta_2 + v$$

  - Unobserved eterogeneity (again and again)
  - Omitted variables (capital?)
  - Errors in variables

Summary

- Heroic identification assumptions. Causal interpretation? I doubt it
- Therefore, no straightforward policy implications, but
- Interesting (partly novel) facts

What do we learn?

- Large firms invest more in R&D compared to small-medium sized firms (especially in LT industries)
- Firms facing international competition invest more in R&D (especially in HT industries)
- Predicted R&D intensity has a strong positive correlation with product innovation (as in Parisi et al, 2006)
- Investment has (possibly) a strong positive correlation with process innovation (as in Parisi et al, 2006)
- Process innovation is more positively correlated with labor productivity than product innovation (as in Parisi et al, 2006) and this is more the case in HT industries

Suggestions

- Exploit also the longitudinal dimension to address some of the endogeneity issues (two consecutive observations can be enough)
- Since the focus is on SMEs try to understand better why, ceteris paribus, small firms innovate more. More cooperation?
- Small technical problems need to be fixed.
  - Capital must be included in the productivity equations.
  - Standard errors have to be corrected to allow for correlation in the score vectors and to take into account the extra variation induced by estimated variables.
  - Same number of observations in the “non-parametric selectivity test” equations. Typo or someting else?