# Comments on "Research, Innovation and Productivity in Four European Countries"

## by R. Griffith, E. Huergo, J. Mairesse, and B. Peters

Manuel Arellano

Madrid, June 25, 2004

## Introduction

- The combination of the substantive questions, econometric modelling, and the CIS data makes this an exciting line of research.
- I was left with the desire to know more about the CIS questionnaires and data.
- The contribution of the paper is mainly empirical: Exploring variants of the Crépon-Duguet-Mairesse model with a new wave of the CIS data for four countries.
- Main results seem to be in line with previous findings.
- More work is needed to understand the sources of cross-country heterogeneity.
- I will provide a quick summary of the model, together with some comments focusing on the empirical strategy.

**A Model for innovators**

- Determinants of R&D
  - *R& D participation:* Probit for being an R&D performer

$$\Pr(s = 1 \mid x) \equiv p(x) = \Phi(xb_0)$$
$$s^* = xb_0 + u_0$$

  - *R& D intensity:* Letting $r$ be log R&D expenditure per employee for performing firms

$$E(r \mid x, s = 1) = xb_1 + \gamma\lambda(xb_0) = xb_1 + g[p(x)]$$
$$r^* = xb_1 + u_1$$

- Production of knowledge:
  - *Process innovation* probit ($pc = 1$ if firm is a process innovator). Letting $x = (x_A, x_B)$:

$$pc1(a_2 s^* + x_A b_2 + u_2 > 0)$$

    $x_A$={demand pull, cost push}. $x_B$=external instruments. I abstract from a subset of $x$ that appear in all equations.

  - *Product innovation.* Let *spd*=1 if product innovator; *zinno*= logit-share of innovative sales for *spd*=1 firms:

$$\begin{aligned} spd &= 1(a_3 s^* + x_A b_3 + u_3 > 0) \\ zinno &= a_4 s^* + x_A b_4 + u_4 \end{aligned}$$

    Other versions use $r^*$ in place of $s^*$.

- Productivity equation. Let $prod$ be log output per worker:

$$prod = a_5 zinno + u_5$$

  Other versions use $s^*$ or $r^*$ instead of $zinno$.

**Comments about the model**

- No discussion of the selection problem arising from the fact that only innovating firms are used. There are two aspects:
  (a) Effect of R&D on innovation: some non-innovating firms may be R&D performers.
  (b) Effect of innovation on productivity: there may be different effects at the non-innovation margin.
- Not clear why using $s^*$ or $r^*$ in the innovation equations as opposed to $s$ or $r$.
- The variable $s^*$ is a propensity to perform R&D, but one would expect production of innovations to depend on actual performance rather than propensities.
- The construction of the potential outcome $r^*$ is particularly unsatisfactory:
  (a) It lacks interpretation,
  (b) it relies on distributional assumptions, and
  (c) no determinant of selection is excluded from the potential outcome equation.

**Using predicted R&D versus Instrumenting R&D**

*A) Linear Models*

- Let $f(z) = E(x|z)$ and consider

$$\text{Model I:} \quad y = bx + u \quad E(u|z) = 0$$
$$\text{Model II:} \quad y = af(z) + v \quad E(v|z) = 0.$$

- Model I is the IV model: $b$ measures the effect of $x$ on $y$ holding $u$ constant in a situation where $x$ and $u$ are potentially correlated. Model II is a regression model that relates $y$ to expected $x$.

- They have different economic interpretations: in Model I agents respond to $x$ whereas in Model II they respond to $f(z)$. But replacing $x = f(z) + e$ in the first equation

$$\begin{aligned} a &= b \\ v &= b * e + u. \end{aligned}$$

- The implication is that whether we use $f(z)$ as a regressor or as an instrument we are estimating the same coefficient.

- The fact that $a = b$ is a peculiarity of the linear case, which does not hold more generally, i.e. it would not hold for:

$$\text{Model I'}: \quad y = g(x, b) + u \quad E(u|z) = 0$$
$$\text{Model II'}: \quad y = g[f(z), a] + v \quad E(v|z) = 0$$

*B) Binary Choice*

- Suppose a probit case such that $u$ and $e$ are normal

$$y = 1\left(bx + u > 0\right)$$
$$y = 1\left(bf\left(z\right) + v > 0\right)$$

- Here derivative effects on the probability are different:

$$\frac{\partial}{\partial f\left(z\right)}\Pr\left(y=1\right) = \frac{b}{\left(1+b^2\sigma_e^2\right)^{1/2}}\phi\left[\frac{bf\left(z\right)}{\left(1+b^2\sigma_e^2\right)^{1/2}}\right]$$
$$\frac{\partial}{\partial x}\Pr\left(y=1\right) = b\phi\left(bx\right).$$

*C) Predicted Latent R& D*

- So far we dealt with predicted R&D. If we have a discrete choice or selection model, we could consider using
  - "predicted propensity to perform R&D"
  - or "predicted potential R&D intensity"

  as explanatory variables, where $x^*$ denotes de propensity or potential outcome, and $f^*\left(z\right)$ its conditional mean.

- In those cases, the model that substitutes $f^*\left(z\right)$ could be equivalent to a hypothetical model that instruments $x^*$.
- But they are different economic models to those using $x$ or $f\left(z\right)$ because they are modelling the response to different explanatory variables.

**Comments about the results**

- We would expect the effect of R&D on productivity to be roughly the product of the effect of R&D on innovation and the effect of innovation on productivity,
- but ONLY if demand pull and cost push variables have been included in the productivity equation;
- i.e. combining the equations for $prod$ and $zinno$ we obtain the semi-reduced form equation:
$$prod = (a_5 a_4)\, s^* + x_A\,(b_4 a_5) + (a_5 u_4 + u_5)$$
- If there is measurement error in sales, the product innovation share coefficient in the labor productivity equation may be downward biased.
- More generally, if $zinno$ is a very noisy measure of product innovation, it may be easier to identify $(a_5 a_4)$ than $a_5$.
- The importance of this is that for policy evaluation we might just be interested in the productivity of R&D.
- Understanding the production of innovations seems an interesting but separate concern.

**Other comments about the econometrics**

- From the policy evaluation perspective of modern labor econometrics some concerns are:
  (a) *Functional form restrictions and common support:* In a matching exercise, assuming that $s$ is exogenous given $x$, one would calculate an average return of R&D performance on productivity as:

$$\int \left[ E\left( prod \mid p\left( x\right) ,s=1\right) -E\left( prod \mid p\left( x\right) ,s=0\right) \right] dG\left( p\right)$$

  where $E\left( prod \mid p\left( x\right) ,s=j\right)$ is a non-parametric simple regression and $p\left( x\right)$ is a propensity score.

  (b) *Instrument validity:* Why are quality improvement, cooperation, international market, and sources of information valid instruments for the innovation and productivity equations?

  (c) *Heterogeneity of returns:* If returns of R&D performance are homogeneous we would expect the regression of productivity on the propensity score to be linear. Testing for nonlinearity can be regarded as a test of heterogeneity.