# estimating the innovation function from patent numbers: gmm on count panel data[*]

Bruno Crépon[†] and Emmanuel Duguet[‡]

Cahiers Eco & Maths 9611
EUREQua

[*]Revised, February 1996. We thank the editor and an anonymous referee for helpful suggestions and comments. We also thank the participants at the CEME, CREST and NBER weekly seminars. A previous version of this paper was presented at the Econometric Society 7th World Meeting (Tokyo, August 22-29 1995).

[†]Direction de la Prévision, Bureau de la Politique Economique. Visiting National Bureau of Economic Research at Cambridge while writing this paper.

[‡]CEME CNRS URA 924, Université de Paris-I-Panthéon-Sorbonne and CREST, Laboratoire d'Economie Industrielle.

## Abstract

The purpose of this paper is to estimate the patent equation, an empirical counterpart to the 'knowledge production function'. Innovation output is measured through the number of European patent applications and the input by research capital, in a panel of French manufacturing firms. Estimating the innovation function raises specific issues related to count data. Using the framework of models with multiplicative errors, we explore and test for various specifications: correlated fixed effects, serial correlation and weak exogeneity. We also present a first extension to lagged dependent variables.

Keywords: count data, generalized method of moments, innovation, panel data, patents, research and development, serial correlation, weak exogeneity.

JEL Classification: C23, C25, L60, O31, O32

## Résumé

Cet article examine les problèmes d'estimation associés à l'équation de brevet, contrepartie empirique de la fonction d'innovation. L'extrant de l'innovation est mesuré ici par le nombre de dépôts de brevets européens et l'intrant par le capital de recherche et développement, sur un panel d'entreprises françaises de l'industrie manufacturière. L'estimation de cette fonction d'innovation implique le recours à l'économétrie des données de comptage sur panel. A l'aide du modèle à erreur multiplicative nous estimons diverses spécifications: effets fixes, autocorrélation des résidus et exogénéité faible. Nous effectuons également un premier examen de l'introduction du nombre de brevets passés dans la relation.

Mots-clef : brevets, données de comptage, exogénéité faible, innovation, méthode des moments généralisés, panel, recherche et développement, résidus autocorrélés.

Classement JEL : C23, C25, L60, O31, O32

# 1 Introduction

Numerous ¯rms devote large amounts of resources to research investment. In order to improve their products, to launch new products or to lower their unit production cost ¯rms may invest in innovation activities. The innovation process can then be decomposed into several steps allowing for some feedback loops. Investment in research and development, both casual and formal, endows ¯rms with new technical knowledge which can be transformed into an economic value. This knowledge is used to improve on products and production processes, thus it generates growth that can be used for future innovation funding. But measures of the economical value of knowledge are not available, only some of its inputs and outputs are (Griliches, 1994, [8]). So that in most of the cases, we obtain information on knowledge through the research and development expenditures input and the patent numbers output.

Patent equations could be used to evaluate the rate of technical progress since there is an equivalence between the duration of the search leading to an innovation and the number of innovations achieved during a given period of time. The more the number of innovations for a given time period the less the average research lag, the more the rate of technical progress. Then, one could think to use patent numbers to evaluate the speed of technical advance. In fact, this would be true if patents were an unbiased measure of the number of innovations. In practise, we have good reasons to believe that patents embed measurement errors (Levin et al., 1987, [12]). Among them, all innovations are not patented and all patents do not have the same value. This suggest that the use of panel data could improve on our estimates, by allowing for ¯rm-level ¯xed e®ects. The basic relationship we consider links the number of patent to research capital. A growth rate model would thus explain the growth of patent numbers by the research accumulation rate. But research is not the only input in the innovation process. Important discoveries may enable ¯rms to improve the productivity of their innovation process itself. Thus, in this work, past patents are used to reveal e±ciency shifts in the knowledge production function.

Starting from previous works on panel count data by Hausman et al. (1984, [11]), we introduce some extensions about the robust estimation issue of multiplicative error models. Four points will be considered: ¯xed e®ects, exogeneity, serial correlation and lagged dependent variables. We will make a systematic use of the Generalized Method of Moments throughout all the paper (Hansen, 1982,[10]), so that our estimates do not rely on stringent distributional assumptions.

Section 2 presents the basic econometric models for panel count data and their limitations. The robust estimation methods covering various cases are developed in section 3. Section 4 presents the data and section 5 gives the estimation results. The conclusion is given with some comments in the last section.

# 2 Panel count data models

Count data have several salient aspects that require speci¯c econometric methods. They include numerous zero counts and are integer valued. This implies that convenient distributional assumptions are to be made when estimating the patent equation. Among them, the Poisson distribution is the simplest. But this distribution has limitations discussed in this section. Answering these criticisms leads to the multinomial regression by Hausman et al. (1984, [11]), closely linked to GMM applied to count panel data.

## 2.1 The basic Poisson model

Let $n_{i,t}$ be the endogenous count data variable for individual $i$ at time $t$, with mean $\lambda_{i,t} > 0$, a Poisson homogenous model is defined by:

$$n_{i,t}/x_{i,1},\ldots,x_{i,T} \overset{iid}{\sim} P(\lambda_{i,t}) \qquad \forall i,t \tag{1}$$

This implies that the probability to observe $n_{i,t}$ patents given the right hand variables $(x_{i,1},\ldots,x_{i,T}) \equiv \underline{x}_{i,T}$ is equal to:

$$\Pr\left(n_{i,t}/\underline{x}_{i,T}\right) = \frac{\exp(-\lambda_{i,t})\,\lambda_{i,t}^{n_{i,t}}}{n_{i,t}!} \tag{2}$$

The only parameter of a Poisson distribution is its mean, denoted $\lambda_{i,t}$, which depends on explanatory variables $x_{i,t}$ where $i$ denotes individuals and $t$ time. To ensure the positivity of this parameter, a natural feature of counts expectations, the mean is set under the exponential form: $\lambda_{i,t} = \exp(x_{i,t}b)$, where $b$ is the coefficient to be estimated. Thus the conditional expectation of our counts, given our exogeneous variables equals: $E\left(n_{i,t}/\underline{x}_{i,T}\right) = \exp(x_{i,t}b)$. In fact, one can show that the conditional variance of our count data equals its conditional expectation. This property is closely related to the fact that the Poisson distribution does not account for a residual in the relationship between $n_{i,t}$ and $\underline{x}_{i,T}$.

As noticed by several authors, the Poisson model does not account explicitly for heterogeneity in the relationship between the expected counts and the right-hand variables. Extensions have been made. The parametric case is examined by Hausman et al. (1984, [11]). The authors do not assume anymore that the relationship between expected counts and the right hand variables is homogeneous, but that a residual $\varepsilon_{i,t}$ enters this relationship, so that now: $E\left(n_{i,t}/\underline{x}_{i,T},\underline{\varepsilon}_{i,T}\right) = \exp(x_{i,t}b + \varepsilon_{i,t}) = \exp(x_{i,t}b)\,u_{i,t}$. When the distribution of the residual $u_{i,t} = \exp(\varepsilon_{i,t})$ is gamma, the distribution for $n_{i,t}$ given $\underline{x}_{i,T}$ has a closed form, it is negative binomial and estimation proceeds by maximum likelihood. An interesting feature is that this heterogeneous Poisson model implies that there is a conditional overdispersion in the count variable since the conditional variance is here always higher than the corresponding mean. The semi parametric case, where no distributional assumption is made about $u_{i,t}$, has been treated by Gouriéroux et al. (1984, [7]). They show that the relationship between the conditional expectation and the conditional variance of $n_{i,t}$ does not depend on the distribution chosen for the residual (i.e., heterogeneity). Thus, estimation can be carried out by pseudo maximum likelihood.

The weakness of the simple Poisson model lies in the three following restrictions: first, it does not allow for individual effects possibly correlated with the right-hand variables; second, it assumes strict exogeneity of the right-hand variables and, last, it does not allow for serial correlation in the residual.

## 2.2 Robust regressions with individual effects

The basic Poisson specification does not allow for individual effects given the exogeneous variables. The $\underline{x}_{i,T}$ are assumed to summarize all individual deviations. And it is clear that the existence of fixed effects at the individual or sectoral level are likely to exist in innovation relationships. We can think there exist elements that affect the return on research investment like different operating skills, appropriability conditions, demand pull or technological opportunities. These variables should remain relatively constant for each firm and end up into an individual effect. It remains that one can choose between a random and a fixed effect formulation.

### 2.2.1 The multinomial regression

The individual effect problem can be partly solved by assuming that there exist a random individual effect $\alpha_i$ entering the mean of the Poisson distribution. In this case serial correlation in the residuals arises when this effect is omitted. It modifies the Poisson model such that:

$$n_{i,t}/\underline{x}_{i,T}, u_i \overset{iid}{\sim} P(\lambda_{i,t} \times u_i) \qquad \forall i,t \tag{3}$$

where $u_i = \exp(\alpha_i) > 0$ is the multiplicative individual effect. A way to solve for this problem consists in assuming a distribution for $u_i$ and integrate it out. This is a part of the work by Hausman et al. (1984, [11]), henceforth referred as HHG. They show that if $u_i$ is gamma distributed and independent from the explanative variables $\underline{x}_{i,T}$, the count variable $n_{i,t}$ has a negative binomial distribution and estimation of the random effect Poisson model proceeds by maximum likelihood. However, this estimator may be biased if the distribution of the individual effect is not gamma[1]. Therefore the authors introduce the fixed effect Poisson model (FEP).

The FEP model has two advantages over the REP model: first, it does not assume that the heterogeneity term $u_i$ is gamma distributed and, second, the fixed effect is not required to be independent from the right-hand variables $\underline{x}_{i,T}$. Moreover it is possible to obtain consistent estimates of the FEP model by using the conditional maximum likelihood framework of Andersen (1970, [1]). The important result is that the observed counts $n_{i,t}$, given the right-hand variables $\underline{x}_{i,T}$ and the individual sums of counts $n_{i.} = \sum_{s=1}^{T} n_{i,s}$, have a multinomial distribution:

$$n_{i,t}/\underline{x}_{i,T}, n_{i.} \overset{iid}{\sim} M(n_{i.}; p_{i,1}, \ldots, p_{i,T}) \tag{4}$$

where $p_{i,s}$ is the theoretical share of year $s$ in firm $i$'s total patenting over years 1 to $T$: $p_{i,s} = \lambda_{i,s}/\sum_{t=1}^{T} \lambda_{i,t}$, that is the share obtained by replacing patent numbers by their expectations. The probability of observing the counts $(c_1, \ldots, c_T)$ for firm $i$ is thus given

[1] The negative binomial distribution has two parameters $(b, \theta)$, the second of which intervenes in the conditional variance only. When this parameter $\theta$ is fixed, the negative binomial distribution belongs to the linear exponential family and the pseudo maximum likelihood estimator for $b$ is consistent. The problem comes from the fact that, on the one hand, the maximum likelihood approach estimates $b$ and $\theta$ simultaneously and, on the other hand, the negative binomial distribution does not belong to the quadratic exponential family. For a more comprehensive treatment, see Gouriéroux et al. (1984, [7]).

by:

$$\Pr\left[n_{i,1} = c_1, \ldots, n_{i,T} = c_T / \underline{x}_{i,T}, u_i, n_{i.}\right] = \frac{\left(\sum_{s=1}^{T} c_s\right)!}{\prod_{s=1}^{T} c_s!} \prod_{s=1} p_{i,s}^{c_s} \tag{5}$$

Standard estimation proceeds by maximum likelihood, that is through the maximization of:

$$L(b) = \sum_{i=1} \sum_{s=1} c_s \ln p_{i,s} \tag{6}$$

The estimator is consistent and asymptotically normal:

$$\sqrt{N}\left(\hat{b} - b\right) \xrightarrow[N \to +\infty]{d} N\left(0, A^{-1}\right)$$

with $A \equiv E\left(-\frac{\partial^2 L(b)}{\partial b \partial b'}\right)$. In fact, the multinomial regression is robust to weaker assumptions.

### 2.2.2 The multinomial regression as GMM estimation

Wooldridge (1990, [14]) has given a re-interpretation of the HHG model. He showed that the multinomial regression is in fact robust. The multinomial model implies that the conditional mean of counts $n_{i,t}$ given the explanative variables and the individual sum of counts equals:

$$E\left[n_{i,t} / \underline{x}_{i,T}, n_{i.}\right] = n_{i.} \times p_{i,t} \tag{7}$$

Then, as the multinomial distribution belongs to the exponential family, we can apply the result by Gouriéroux et al. (1984, [7]) and conclude that the multinomial estimator is consistent. As with the Poisson model however, the robust (i.e., the pseudo maximum likelihood) covariance matrix has to be computed. We now have:

$$\sqrt{N}\left(\tilde{b} - b\right) \xrightarrow[N \to +\infty]{d} N\left(0, A^{-1} B A^{-1}\right) \tag{8}$$

with $A$ defined previously and $B \equiv E\left(\frac{\partial L(b)}{\partial b}\frac{\partial L(b)}{\partial b'}\right)$.

But Wooldridge also showed that the consistency of the multinomial regression holds under another set of conditional mean restrictions. If the following restrictions hold:

$$E\left[n_{i,t} / \underline{x}_{i,T}, u_i\right] = \lambda_{i,t} u_i \tag{9}$$

then the multinomial estimator remains a consistent estimator of $b$ and the covariance matrix is similar to (8). This new set of conditional mean restrictions (9) is different from the previous one (7). They are both implied by the FEP model but none is a consequence of the other. The second one is, however, a more direct consequence of the FEP model. One can show that when (9) holds:

$$E\left[n_{i,t} - n_{i.} p_{i,t} / \underline{x}_{i,T}\right] = 0 \tag{10}$$

which is a key condition for the robustness of the multinomial regression[2]. In fact, relation (10) can be interpreted as a way to eliminate the individual fixed effects from the semi parametric model (9) in order to obtain a set of orthogonality conditions, here:

$$E\left[\underline{x}_{i,T} \times (n_{i,t} - n_{i\cdot}\, p_{i,t})\right] = 0 \tag{11}$$

Thus this set of conditions can be used in the generalized method of moments (GMM) framework. As the estimator is based on deviations of the counts from an expression of their theoretical means, the corresponding GMM estimators will be henceforth referred as `within' estimators[3].

In this different approach to the estimation problem, we seek to minimize the distance between our orthogonality conditions and the nul vector according to a metric $\Omega$. That is, we estimate the parameter $\beta$ by $\hat{\beta}$:

$$\hat{\beta} = \arg\min_{\beta} \bar{h}' \Omega \bar{h}$$

where $\bar{h}$ is the sample counterpart of equation (11) expectation, a column vector with as many elements as instruments[4], $\bar{h} = 1/N \sum_{i=1}^{N} h_i$, with:

$$h_i \equiv h\left(\underline{n}_{i,T}, \underline{x}_{i,T}, \beta\right) \equiv \underline{x}'_{i,T} \otimes \begin{pmatrix} n_{i,1} - n_{i\cdot}\, p_{i,1} \\ \vdots \\ n_{i,t} - n_{i\cdot}\, p_{i,t} \\ \vdots \\ n_{i,T} - n_{i\cdot}\, p_{i,T} \end{pmatrix}$$

The estimation proceeds in two steps. First, we set $\Omega = Id$ to get a first estimate $\hat{\beta}$ of $\beta$ that enables us to compute a consistent estimate of the optimal metric given by $\Omega^* = V(h)^{-1} = E(hh')^{-1}$. We can take the sample counterpart of it as a consistent estimate: $\hat{\Omega}^* = \left(1/N \sum_{i=1}^{N} \hat{h}_i \hat{h}'_i\right)^{-1}$ with $\hat{h}_i = h\left(\underline{n}_{i,T}, \underline{x}_{i,T}, \hat{\beta}\right)$. Second, we estimate $\beta$ again with $\Omega = \hat{\Omega}^*$, which gives us the estimates $\hat{\beta}^*$ presented in the tables. The asymptotic distribution of $\hat{\beta}^*$ is given by:

$$\sqrt{N}\left(\hat{\beta}^* - \beta\right) \xrightarrow[N \to +\infty]{d} N(0, \Sigma^*) \tag{12}$$

with the consistent estimate of $\Sigma^*$:

$$\hat{\Sigma}^* = \left[\left(\frac{1}{N}\sum_{i=1} \frac{\partial h'_i}{\partial \beta}\left(\underline{n}_{i,T}, \underline{x}_{i,T}, \hat{\beta}^*\right)\right) \hat{\Omega}^* \left(\frac{1}{N}\sum_{i=1} \frac{\partial h_i}{\partial \beta'}\left(\underline{n}_{i,T}, \underline{x}_{i,T}, \hat{\beta}^*\right)\right)\right]^{-1}$$

---

[2] The key condition is that the objective we maximize converges uniformly in probability to a function that reaches its maximum at the true value of the parameter $\beta_0$, say. The first order condition for the multinomial regression is: $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} (n_{i,t} - n_{i\cdot}\, p_{i,t})\, \nabla_\beta \exp(x_{i,t}\beta)/\exp(x_{i,t}\beta) = 0$ and by the law of large numbers the left hand side converges to $E_{\beta_0}\left[\sum_{t=1}^{T} (n_{i,t} - n_{i\cdot}\, p_{i,t})\, \nabla_\beta \exp(x_{i,t}\beta)/\exp(x_{i,t}\beta)\right]$.

This requires only (10) to be true at the true value of the parameter and therefore the multinomial regression to be consistent.

[3] Notice it is not the deviation from the empirical mean with count data.

[4] To simplify the exposition, we consider the case with one explanative variable. The extension is straightforward, by stacking.

In fact, we can extend a little this is standard way of estimation. For any function G of explanatory variables $\underline{x}_{i,T}$, the following orthogonality conditions hold:

$$E\left[(n_{i,t} - n_{i\cdot}\, p_{i,t}) \times G\left(\underline{x}_{i,T}\right)\right] = 0 \qquad (13)$$

Moreover, as (13) is true for any function $G\left(\underline{x}_{i,T}\right)$ a simple manipulation allows for re-writing these conditions as:

$$E\left[\left(n_{i,t} - n_{i,t+1}\frac{\lambda_{i,t}}{\lambda_{i,t+1}}\right) \times G\left(\underline{x}_{i,T}\right)\right] = 0 \qquad (14)$$

which is the quasi-differentiation proposed by Chamberlain (1992, [3]) for panel data models with multiplicative errors. We will henceforth refer to this estimator equivalently as `first difference' or `correlated effect'.

This raises the issue of the optimal choice of instruments, namely of function G: This also falls into the framework studied by Chamberlain (1992, [4]) of efficiency bounds for semi parametric regressions. It is possible to derive the theoretical semi parametric efficiency bound and to exhibit the optimal choice of instruments G to use. However this function of observations has the form of theoretic conditional expectation and must also be estimated. Newey (1990, [13]) achieves this through a non parametric step by using series estimators and the k-nearest neighbors method to determine the optimal instruments. We will not developp this approach in this study. Instead, we will just try to get an idea about the magnitude of possible efficiency gains by comparing the results achieved through two sets of instruments. The first instrument set is the simplest: a constant term[5] and the explanative variables for all time periods. The second instrument set includes both the former one and the independent cross products from this first set variables. With one explanatory variable and $T$ time periods (i.e., years), the first set of instruments includes $T + 1$ instruments which support $(T + 1)(T - 1)$ orthogonality conditions[6]. The second instrument set includes the $T(T + 1)/2$ cross-products instruments: this gives $T(T + 1)(T - 1)/2$ additional orthogonality conditions. In this application, the number of time periods is $T = 6$ so that the numbers of orthogonality conditions are respectively 35 and 105 + 35 = 140.

## 3 Robust estimation with panel count data

The FEP model assumes that the counts, given the explanatory variables and the fixed effect, are independently and identically Poisson distributed. Thus nor does it relax the conditional variance-to-mean ratios neither does it allow for serial correlation. In their paper, HHG (1984, [11]) proposed a negative binomial fixed effect model that relaxes this assumption, they allow for a firm-specific conditional variance-to-mean ratio given the right hand variables and the fixed effect. This has interesting economic interpretations: on the one hand, uncertainty in the production of innovation is more or less important across firms for a given mean and; on the other hand, the propensity to patent differs among

[5] Or, alternatively, a full set of time dummies.

[6] There are $T + 1$ instruments and $T - 1$ quasi differences.

firms in the sample. However, their modeling implies a variance-to-mean ratio greater than one, that is overdispersion. Thus, in this setting, uncertainty in the production of innovations should increase monotonically with the number of discoveries. Looking for models with more flexibility is thus interesting. We can keep the same basic framework by relaxing both the Poisson and overdispersion assumptions.

## 3.1 Weak exogeneity

The previous quasi differentiation can be extended easily to the weak exogeneity case. Under this assumption the relationship (9) is no more valid but the different conditioning with $\underline{x}_{i;t} \equiv (x_{i;1};\ldots;x_{i;t})$ depending on the date t at which the expectation is taken:

$$E\left[n_{i;t}/\underline{x}_{i;t};u_i\right] = \lambda_{i;t}\, u_i \tag{15}$$

In fact, the expectation is conditional on both $u_i$ and $\underline{x}_{i;t}$ so that we have both weak exogeneity and correlated fixed effects. Here also estimation stems from the elimination of fixed effects which provides a set of orthogonality conditions available to the GMM approach. We use the chamberlinian differentiation again. The key idea here is that the fixed effect can be expressed as:

$$u_i = E\left[u_i \mid \underline{x}_{i;t};u_i\right] = E\left[E\left(\frac{n_{i;t+a}}{\lambda_{i;t+a}} \middle| \underline{x}_{i;t+a};u_i\right) \middle| \underline{x}_{i;t};u_i\right] = E\left(\frac{n_{i;t+a}}{\lambda_{i;t+a}} \middle| \underline{x}_{i;t};u_i\right)$$

$$\forall\, a = 0;\ldots;T - t; \quad \forall\, t = 1;\ldots;T$$

Thus as the fixed effect is time independent by definition, it is possible to exclude a part of the $x'$s lags out of the instruments set. We use this result for a = 1; derive the corresponding expression for $u_i$ and substitute it in (15) to get the following expected quasi difference :

$$E\left[\rho_{it}/\underline{x}_{i;t}\right] = 0 \quad \text{with} \quad \rho_{it} = n_{i;t} - n_{i;t+1}\frac{\lambda_{i;t}}{\lambda_{i;t+1}} \tag{16}$$

From this expression, GMM estimation is implemented by using the corresponding orthogonality conditions of the form:

$$E\left\{\rho_{it} \times G\left[\underline{x}_{i;t}\right]\right\} = 0;\ \forall\, t = 1;\ldots;T - 1 \tag{17}$$

Once more, the issue of the optimal instruments sets arises. Chamberlain (1992, [4]) derives the corresponding expression of the semi parametric efficiency bound and that of the optimal instrument to use. They are more complicated than in the 'within' case since now the instrument set changes with the date t. As for the strict exogeneity case, a first non parametric step would be needed. So, we use here the same two sets of instruments as before and refer to these estimators as 'weak exogeneity with correlated effects'.

## 3.2 Restricted serial correlation

The previous GMM estimators require the simple conditional mean restrictions (9) and (15) only. For that reason, they are consistent even when the three restrictions of the Poisson model are relaxed: they allow for fixed effects correlated with the right hand variables, for any relationship between the mean and variance of the count distribution and for any pattern of serial correlation. In the previous sections we explained how it is possible to move away from parametric models that embed restrictions we would like to get rid of. On the one hand, semi parametric estimation is less demanding than maximum likelihood by only keeping the basic interesting properties of the parametric models. But, on the other hand, we lose efficiency if some constraints are wrongly excluded. Thus we would like to impose more structure on the previous GMM estimators, especially on the existence of restrictions in the correlations of the residuals, these constraints being tested.

The canonical situation for linear models is the case in which the time varying parts of the perturbations are uncorrelated. The interesting features associated with it is that past values of the dependent variable can be used as instruments. We introduce similar models for count panel data, in two different ways. The first method consist in adding past values of the count variable in the list of conditioning variables. The second method to restrict serial correlation imposes less restrictive assumptions since it uses conditional covariance constraints only.

### 3.2.1 Lagged dependent variable as an instrument

We keep the weak exogeneity with correlated effects basic framework. Depending on the $x'$s are strictly or weakly exogenous, this gives the respective moment conditions:

$$E\left(n_{i,t}/\underline{x}_{i,T}, \underline{n}_{i,t-a}, u_i\right) = \lambda_{i,t} u_i \tag{18}$$

$$E\left(n_{i,t}/\underline{x}_{i,t}, \underline{n}_{i,t-a}, u_i\right) = \lambda_{i,t} u_i \tag{19}$$

We will instrument by all convenient past values of patents so that in our applications $a = 1$. Notice here that both the random and fixed effects models of HHG imply such restrictions. In these cases the orthogonality conditions are obtained in the same way as the previous estimates under weak exogeneity with correlated effects. Strict exogeneity with correlated effects imply the restrictions:

$$E\left(\rho_{i,t}/\underline{x}_{i,T}, \underline{n}_{i,t-1}\right) = 0, \ \forall t = 1, \dots, T-1$$

while weak exogeneity with correlated effects relies on

$$E\left(\rho_{i,t}/\underline{x}_{i,t}, \underline{n}_{i,t-1}\right) = 0, \ \forall t = 1, \dots, T-1$$

### 3.2.2 Covariance restrictions

**Strict exogeneity.** Consider first strict exogeneity with correlated effects; if $\left(n_{i,t}\right)$ and $\left(n_{i,s}\right)$ are independent for all $s \neq t$ then for any function $g$ we have $\mathrm{Cov}\left(n_{i,t}, g(n_{i,s}) \mid \underline{x}_{i,T}, u_i\right) = 0$; although the reverse is not necessarily true. Then using these covariances lead to less restrictive conditions. Strict heterogeneity implies $T(T-1)/2-1$ orthogonality conditions since $u_i$ is to be eliminated. The absence of serial correlation can thus be represented by

$$E\left(n_{i,t}n_{i,s}/\underline{x}_{i,T}, u_i\right) = E\left(n_{i,t}/\underline{x}_{i,T}, u_i\right) E\left(n_{i,s}/\underline{x}_{i,T}, u_i\right) = \lambda_{i,t}\lambda_{i,s}u_i^2, \ \forall s \neq t$$

$$\Rightarrow E\left[\left(\frac{n_{i,t}}{\lambda_{i,t}}\frac{n_{i,s}}{\lambda_{i,s}} - \frac{n_{i,t'}}{\lambda_{i,t'}}\frac{n_{i,s'}}{\lambda_{i,s'}}\right)/\underline{x}_{i,T}\right] = 0, \ \forall s \neq t, \ \forall s' \neq t' \tag{20}$$

A first set of orthogonality conditions is obtained by setting $s' = s$ and $t' = t + 1$ in the previous relationship (20), after some simplifications we get:

$$E\left(n_{i,s}\rho_{i,t}/\underline{x}_{i,T}\right) = 0, \ \forall s < t = 2, \ldots, T-1 \tag{21}$$

This makes $(T-1)(T-2)/2$ constraints[7]. Let $s' = t+1$ and $t' = t+2$ in the relationship (20) and we have the additional independent constraints:

$$E\left[n_{i,t+1}\left(n_{i,t} - n_{i,t+2}\frac{\lambda_{i,t}}{\lambda_{i,t+2}}\right)/\underline{x}_{i,T}\right] = 0, \ \forall t = 1, \ldots, T-2 \tag{22}$$

**Weak exogeneity.** The case of weak exogeneity is more restrictive since this time the conditioning is made on the past and current values of the right-hand variables only. This suppresses a number of orthogonality conditions. More precisely the second set of conditions (22) cannot be employed anymore. We have: $\mathrm{Cov}\left(n_{i,t}, n_{i,s}/\underline{x}_{i,t}, u_i\right) = 0, \ \forall s < t = 2, \ldots, T$. Then:

$$E\left(\frac{n_{i,t}}{\lambda_{i,t}}n_{i,s}/\underline{x}_{i,t}, u_i\right) = E\left(\frac{n_{i,t}}{\lambda_{i,t}}/\underline{x}_{i,t}, u_i\right) E\left(n_{i,s}/\underline{x}_{i,t}, u_i\right) = u_i E\left(n_{i,s}/\underline{x}_{i,t}, u_i\right)$$

As the individual effect $u_i$ is time independent by definition, we can write:

$$u_i = E\left(\frac{n_{i,t+1}}{\lambda_{i,t+1}}/\underline{x}_{i,t+1}, u_i\right)$$

thus:

$$\begin{aligned}
E\left(\frac{n_{i,t}}{\lambda_{i,t}}n_{i,s}/\underline{x}_{i,t}, u_i\right) &= E\left(u_i n_{i,s}/\underline{x}_{i,t}, u_i\right) \\
&= E\left[E\left(\frac{n_{i,t+1}}{\lambda_{i,t+1}}n_{i,s}/\underline{x}_{i,t+1}, u_i\right)/\underline{x}_{i,t}, u_i\right] \\
&= E\left(\frac{n_{i,t+1}}{\lambda_{i,t+1}}n_{i,s}/\underline{x}_{i,t}, u_i\right) \\
\Rightarrow \quad & E\left(n_{i,s}\rho_{i,t}/\underline{x}_{i,t}\right) = 0, \ \forall s < t = 2, \ldots, T-1
\end{aligned}$$

As in the previous sections we use the $x'$s and a constant term as instruments. Then, we add the cross products of the $x'$s to form the second instrument set. The corresponding overidentification tests are also provided.

[7] Thus, there remains $T(T-1)/2 - 1 - (T-1)(T-2)/2 = T-2$ independent constraints to exhibit.

## 3.3 Lagged dependent variables

It is possible to embed lagged dependent variables in the previous models both in the weak exogeneity and in the correlated effects cases. For weak exogeneity with correlated effects the orthogonality conditions can be written:

$$E\left(n_{i,t}/\underline{x}_{i,t},\underline{n}_{i,t-1},u_i\right) = h\left(n_{i,t-1},\gamma\right)\lambda_{i,t}u_i \tag{23}$$

where $h(.,.) > 0$ is any given function describing the way by which past patents are to affect current innovations and $\gamma$ is the associated vector parameter summarizing the link between $n_{i,t}$ and its past. To be consistent with the previous models we add the constraint: $h(n_{i,t-1},0) = 1$, without loss of generality provided $\underline{x}_{i,t}$ includes a constant term. It ensures that when $\gamma = 0$ the model reduces to the ones of the previous sections[8]. This model can easily be extended to allow for more lags.

There are two innovation sources in this new setting: on the first hand, endogenous shifts in the patent equation with $h(n_{i,t-1},\gamma)$ that is the weight of past discoveries in current innovation and, on the other hand, innovations stemming from research investment $\lambda_{i,t}$. Both innovation sources combine in $h(n_{i,t-1},\gamma)\lambda_{i,t}$ to produce innovations $n_{i,t}$, together with a fixed effect $u_i$. The correlated effects with strict exogeneity version is achieved through:

$$E\left(n_{i,t}/\underline{x}_{i,T},\underline{n}_{i,t-1},u_i\right) = h\left(n_{i,t-1},\gamma\right)\lambda_{i,t}u_i \tag{24}$$

The way we introduce past patents here is the simplest and does not imply stringent assumptions on function h. First, a dummy $d_{i,t-1}$ indicating if firm i has applied for a patent in the previous year, defined as $d_{i,t} = \mathbb{1}_{n_{i,t-1}\geq 1}$. The second model is a simple extension of the first one; we allow for different coefficients corresponding to different values of the past number of patents, the model is now made of M dummies $d_{i,t}^{(m)} = \mathbb{1}_{\alpha_{m-1}\leq n_{i,t-1}<\alpha_m}$, $m = 1,\ldots,M$, with $\alpha_0 = 1$ and $\alpha_M = +\infty$ by convention. These dummies have a direct interpretation.

To see this, consider the first case:

$$h\left(n_{i,t-1},\gamma\right) = \exp\left(\gamma\mathbb{1}_{n_{i,t-1}\geq 1}\right)$$

The parameter $\gamma$ is approximately equal to the percentage performance advantage for firms that have patented the previous year:

$$\frac{E\left(n_{i,t}/\underline{x}_{i,T},\underline{n}_{i,t-1},u_i,d_{i,t-1}=1\right)}{E\left(n_{i,t}/\underline{x}_{i,T},\underline{n}_{i,t-1},u_i,d_{i,t-1}=0\right)} - 1 = \exp(\gamma) - 1 \simeq \gamma$$

with $\gamma$ close to zero. The extension to several dummies is straightforward.

However, the orthogonality conditions like (23) or (24) are only a part of the specification of autoregressive models. As is well-known from studies of the linear model, the specification of the initial conditions is also important. Here, we do not observe the

[8]Other specifications are available, for example: $E\left(n_{i,t}/\underline{x}_{i,T},\underline{n}_{i,t-1},u_i\right) = \left(h\left(n_{i,t-1},\gamma\right)+\lambda_{i,t}\right)u_i$, with $h(n_{i,t-1},0) = 0$. In this case, past patenting 'adds' to current research. A direct generalization is to allow for a cross product so that the impact of past patents depends on the amount of research capital.

number of past patents for the first period so that these numbers are to be treated as additional firm specific effects. For instance, in the correlated effects model, we have:

$$E\left[n_{i,1}/\underline{x}_{i,T},\underline{n}_{i,0},u_i\right] = h\left(n_{i,0},\gamma\right)\lambda_{i,0}u_i = \lambda_{i,0}\tilde{u}_i \tag{25}$$

with $\tilde{u}_i = h\left(n_{i,0},\gamma\right)u_i$. Defining the quasi difference $\rho_{i,t}$ in the same way as (16) :

$$\rho_{i,t} = n_{i,t} - n_{i,t+1}\frac{h\left(n_{i,t-1},\gamma\right)\lambda_{i,t}}{h\left(n_{i,t},\gamma\right)\lambda_{i,t+1}} \tag{26}$$

we get the following set of orthogonality conditions after the fixed effects are eliminated:

$$E\left[\rho_{i,t}/\underline{x}_{i,t},\underline{n}_{i,t-1}\right] = 0,\ \forall t = 2,\ldots,T-1$$

in the weak exogeneity with correlated effects case, and:

$$E\left[\rho_{i,t}/\underline{x}_{i,T},\underline{n}_{i,t-1}\right] = 0,\ \forall t = 2,\ldots,T-1$$

in the strong exogeneity one. Notice that here the 'residual' can be used from date $t = 2$ only, since it is not possible to form a valid residual for the first period. In the followings, we use the notation $u_i$ instead of $\tilde{u}_i$.

The previous set of orthogonality conditions make an explicit use of restricted serial correlation. In the context of linear panel data we know it is possible to estimate autoregressive models without assumptions on serial correlation. One just needs to assume that enough past values of the explanatory variables are used as instruments. Here, there is an equivalence although estimation practise appears to be more troublesome. Let the model:

$$E\left[n_{i,t}/\underline{x}_{i,t},n_{i,t-1},u_i\right] = h\left(n_{i,t-1},\gamma\right)\lambda_{i,t}u_i \tag{27}$$

instead of (23) in which instruments where all past values of patents. Using the fact that the initial value $n_{i,0}$ is embedded in the fixed effect[9], we can write down:

$$\begin{aligned}
E\left(\frac{n_{i,t}}{h\left(n_{i,t-1},\gamma\right)\lambda_{i,t}}\ \Big/\ \underline{x}_{i,t},n_{i,t-1},u_i\right) &= E\left(\frac{n_{i,t}}{h\left(n_{i,t-1},\gamma\right)\lambda_{i,t}}\ \Big/\ \underline{x}_{i,t},u_i\right)\\
&= E\left[u_i/\underline{x}_{i,t},u_i\right]\\
= E\left(E\left(\frac{n_{i,t+1}}{h\left(n_{i,t},\gamma\right)\lambda_{i,t+1}}\ \Big/\ \underline{x}_{i,t+1},u_i\right)\Big/\ \underline{x}_{i,t},u_i\right)&\\
= E\left(\frac{n_{i,t+1}}{h\left(n_{i,t},\gamma\right)\lambda_{i,t+1}}\ \Big/\ \underline{x}_{i,t},u_i\right)&\\
\Rightarrow E\left(\frac{n_{i,t}}{h\left(n_{i,t-1},\gamma\right)\lambda_{i,t}} - \frac{n_{i,t+1}}{h\left(n_{i,t},\gamma\right)\lambda_{i,t+1}}\ \Big/\ \underline{x}_{i,t}\right) = 0&
\end{aligned}$$

which could be used to form orthogonality conditions. Unfortunately, it does not seem possible to transform the previous relationships in a way that give them the shape of the

[9] This implies that the information on $n_{i,t}$ is summarized by $\left[\underline{x}_{i,t},u_i\right]$ instead of $\left[\underline{x}_{i,t},\underline{n}_{i,t-1},u_i\right]$.

previous `residual' like in (26): In fact, the residual obtained here is the one proposed by Wooldridge (1991, [15]).

A practical problem appears: this way to write down the orthogonality conditions leads to spurious maxima. More precisely, if there is a right-hand variable that takes on the same sign for all the observations, then it is possible to make the previous `residual' arbitrarily close to zero by increasing the value of the corresponding coefficient. In this case, it is necessary to impose concretely the compactness of the parameters set. This is the reason why we do not use this elimination device in this paper but conditioning on the past endogenous variables.

# 4 The sample

The data we use are similar to that of previous works ([5], [6]). Our sample is made of 698 firms in French manufacturing other 6 years 1984-1989 (i.e., 4188 data points). It includes informations on the number of European patent applications as well as inflation corrected research and development expenditures. It is build from two sources: the R&D survey and the European patent (EPAT) data base.

## 4.1 Research and development capital

In order to build our research capital, we use the answers to the R&D survey. This survey has been carried out in France since the early 1970's and gives various informations about research expenditures for firms satisfying the Frascati criteria[10]. This allows us to start our research sample in year 1974. We then compute a research capital $k_{i;t}$ by the perpetual inventory method. That is, research capital for firm i at the end of year t is obtained from the formula: $k_{i;t} = (1 - \delta)\, k_{i;t-1} + r_{i;t}$; where $\delta$ is the annual depreciation rate and $r_{i;t}$ the inflation corrected total research expenditures, including the ones purchased from outside the firm[11]. The research capital is computed assuming an annual obsolescence rate of 15% like in Hall and Mairesse (1995, [9]) on similar data. For the firms that we observe since 1974 we assumed a pre-sample annual growth rate of 5% so that the starting value of their capital is $k_0 = r_0/(0.05 + \delta) = 5 \times r_0$. For the other firms, the initial research capital is assumed to be zero.

## 4.2 Patent numbers

Patent data come from the European PATent (EPAT) data base which records patents since 1978. However, these data did not include the French national identification codes[12] of firms so that a specific work had to be done at INSEE (Bussy et al., 1995, [2]). The matched sample covers the years 1980-1989 but we use it from 1984 only. The reason

[10] Mainly, at least one employee working full time on research.

[11] The deflation index for R&D is computed from the decomposition of internal research expenditures between materials, wages and investment. We weighted the corresponding macroeconomic prices by the corresponding average shares in the R&D survey.

[12] The so-called SIREN code (SIREN: Service Informatique de RÉpertoire des ENtreprises) allows to match French firm-level data coming from all public surveys.

is that the European registration progressively developed at the expense of the French patent data base (FPAT) to reach a steady state around 1984, year in which about 70% of French patents are registered in EPAT[13]. This specificity of our dependent variables, the number of patents $n_{i,t}$, could appear as problematic. In fact, panel data allows us to circumvent this problem. If we model the number of patents registered in EPAT $n_{i,t}$ as a proportion $\phi_{i,t}$ of the total number of patents $z_{i,t}$, say, we have: $E(n_{i,t}/z_{i,t}, \phi_{i,t}) = \phi_{i,t} z_{i,t}$. Provided that the choice of the registration mode is independent from both the number of innovations $z_{i,t}$ and the explanatory variables $x_{i,t}$, only is the intercept of our model changed: $\ln E(n_{i,t}/z_{i,t}, \phi_{i,t}) = \ln \phi_{i,t} + \ln z_{i,t}$. We can account for this heterogeneity in the intercept by allowing both the intercept to change with time $\psi_t$ and an individual fixed $u_i$ effect in the equation. In the multiplicative framework we use this leads to $\phi_{i,t} = u_i \times \psi_t$, so that:

$$E(n_{i,t}/z_{i,t}, \phi_{i,t}) = \phi_{i,t} E\left(z_{i,t}/\underline{x}_{i,T}, u_i, \psi_t\right) = u_i \psi_t \lambda_{i,t}$$

$$\Leftrightarrow \ln E\left(n_{i,t}/\underline{x}_{i,T}, \phi_{i,t}, u_i, \psi_t\right) = \ln u_i + \ln \psi_t + x_{i,t} b$$

Another problem often encountered with patent panel data is the fact that firms do not patent all their innovations. It can be solved exactly as above, so that the fixed effect summarizes both incomplete patenting of innovations and the link between European and total patenting.

We have included a full set of time dummies in all the regressions so that the average probability to register in EPAT as well as the patenting probability are allowed to change with time.

# 5 The results

Several issues will be discussed in this section: the choice of the instruments, the level versus first difference estimates, exogeneity, restricted serial correlation and lagged dependent variables.

## 5.1 Instruments selection

All our GMM estimators can be interpreted as the outcome of zero conditional expectation conditions. These orthogonality conditions are valid conditional to a given set of instruments: the products of the `residual' and any function of the instruments are available for this purpose. We present two GMM estimators stemming from two different instrument sets. The first one is the conveniently lagged x variables and an intercept, henceforth referred as $GMM_1$. The second set of instruments includes the first set plus all the independent cross products of the previous $x'$s. We will refer to it as $GMM_2$. The $GMM_2$ estimators are introduced because they should be closer to the optimal GMM estimators than $GMM_1$. Therefore we expect an efficiency gain when we move from $GMM_1$ to $GMM_2$. We also test that the additional constraints from $GMM_2$ are satisfied.

[13] FPAT is not accessible to us and cannot be matched at the firm level since the SIREN codes are not a available for this data base.

The first result is that the accuracy gains are very important when passing from $GMM_1$ to $GMM_2$, the use of the latter instrument set leads to sharply lower standard errors. They are nearly always divided by more than three, as can be seen from the estimation tables.

The second result is that the estimates obtained from the two instrument sets are close. This was an expected result since both estimators are consistent. The changes are always minor and within standard errors. In fact, we can test the adequacy of $GMM_1$ and $GMM_2$ in a more formal way. Using the overidentification statistics given in the tables, we test for the validity of the additional constraints associated to $GMM_2$. The results are presented in table 3. In all but two cases we do accept the compatibility of $GMM_1$ and $GMM_2$ at the 5% level. Moreover, when a rejection occurs, we are never far from accepting the null. Actually, the worst case is for the within estimation, where the statistics has a significance probability of 4.3%, quite close to 5%. The overall impression is thus that introducing the cross-products in the instrument set is fruitful and allows for increasing drastically the efficiency of the estimates.

## 5.2 Levels versus correlated effects

An interesting and widespread comparison in panel data is between the level and first difference (or within) estimators. From HHG results we expect that the research elasticity is to decrease when we pass from levels to first differences. We also perform the corresponding specification test through the overidentifying statistics, since the estimators in levels includes the constraints defining the first difference one. Thus we can test for the validity of the additional constraints implied by the estimator in levels.

As expected, the two estimators give very different results (table 4). The levels estimates ranges from 0.75 with $GMM_1$ to 0.82 for $GMM_2$, which increases the accuracy by lowering the standard error from 0.04 to 0.01. This order of magnitude is compatible with the results of HHG and lower than these of Crépon and Duguet (1995, [5]) that obtained constant returns to scale on a cross section[14].

The correlated fixed effects estimators give very close results for the research elasticity, between 0.25 and 0.26. As expected the values are lower than with levels but they remains significant at the 5% level for the GMM estimators. The HHG pseudo maximum likelihood estimate gives a comparable coefficient but its standard error is much higher, so that we accept it is not significant at the 5% level. Thus it appears that this latter estimator is less accurate than the two $GMM_1$ ones[15].

The tests of levels versus first difference are obtained from the overidentification statistics in table 4. The test statistic for $GMM_1$ is equal to 34.23–21.79=12.44 with 35–29=6 degrees of freedom. The critical value at the 5% level for the $\chi(8)$ distribution is 12.59 so that the null is just accepted although we are close to rejection at the 5% level. This can be a bad result if it means that introducing fixed effects in the modeling does not changes the results significantly. If there was an individual opportunity factor other than

[14] The reason for this departure could be that the former study includes a full set of industry dummies when estimating the cross section. We do not include it here since we focus on the first difference estimates.

[15] We do not comment on $GMM_2$ since this instrument set is rejected for the two GMM correlated effects estimators (table 3).

research, the null hypothesis should be rejected. It remains that, on the one hand, the level elasticity strongly departs from the others estimates and, on the other hand, we are close to rejection. An explanation to this could also be a poor power of the test.

## 5.3 Restricted serial correlation

Using the overidentification statistics,we test for the existence of serial correlation among the residuals. These tests are linked to the following issue: a remaining correlation in the residuals once accounted for fixed effects and research capital could indicate an importance of correlated random shocks in the innovation function. Conversely, accepting restricted serial correlation would suggest that innovation, as far as we can observe it, looks more like a steady process within the period under study.

Under strict exogeneity (table 4) the statistic reaches 13.38 while it is 14.84 under weak exogeneity (table 5), both with ten degrees of freedom. Thus we cannot reject the hypothesis of no serial correlation at the 5% level (the probabilities are respectively 0.20 and 0.14). There is no evidence of correlated random shocks in the innovation function. Here again, the estimates remain close among the different estimation methods, from 0.26 to 0.30. We now look at the exogeneity issue.

## 5.4 Strict versus weak exogeneity

The less constrained estimator is the one obtained under weak exogeneity without restricted serial correlation based on $GMM_1$, the most constrained is obtained under strict exogeneity with both restricted serial correlation and $GMM_2$: To see this, notice that the number of orthogonality conditions used by these estimators are respectively 14 and 224. Using these additional constraints allows for reducing the standard error of research elasticity from 0.22 to 0.004 which is a strong gain. The size of the coefficients are slightly affected by these assumptions since they go from 0.26 to 0.32. Using $GMM_2$ instead of $GMM_1$ improves the accuracy of weak exogeneity estimates. The tests of weak versus strict exogeneity are presented in table 7. All the tests do not allow to reject the strict exogeneity assumption at the 5% level, that is the simple correlated effects model. What seems to matter here is clearly the instrument set used and $GMM_2$ is accepted for the two weak exogeneity estimates.

## 5.5 Models with past dependence

The main issue we want to address is to know wether past innovation induces shifts in technological opportunities or not. The sign of the shift is also interesting: on the one hand, new discoveries can lead to explore fertile research areas and therefore to an increase of technological opportunities; on the other hand, if the technological basis is fixed, firms will experience an innovation shortage since the more they have discovered in the past the less remains to be discovered. In this latter case, innovation is associated to a decrease in technological opportunities.

We examine two different functional forms for the effect of past innovation on research productivity. The first one is a dummy variable indicating that firm has applied for at least one patent in the previous year, the second model is a set of three dummies representing different number of patent applications in the previous years: between 1 and 5, 6 and 10 and more than 10. Table 6 gives the estimates obtained with restricted serial correlation[16]. In all the cases, we find that one cannot reject there is a significant effect associated with past patents.

For the different specifications we also perform the tests for instrument selection and exogeneity. We always accept the strict exogeneity constraints as well as the $GMM_2$ instrument set.

Consider the first model (table 6). The $GMM_1$ estimates always conclude that the research capital elasticity is not significant at the 5% level. Only the $GMM_2$ estimates do accept its significance. The coefficient estimated in this first model is stable among estimation methods, around 0.10. The effect of past patent is significant in all specifications at the 5% level, around 0.25. This implies that the average gain of past patenting is an extra $\gamma = 25\%$ current patenting.

The second specification, with three dummies, is displayed in the same table. Here, we want to allow for different effects of past innovation depending on the size of this activity measured by its past output. The first result is that the effect of past research is strongly weakened with this specification also. The second one is that the first patent class $\{1 \le n_{i,t-1} \le 5\}$ is predominant. Clearly, the innovation function does not appear to be the same for small and big innovators. Negative coefficients even appear for research under weak exogeneity. But this is likely to be an artefact. More precisely, two arguments can explain our result. The first one is about the value of patents, which would be bigger the bigger the firm. In this case, the evolution of patents numbers for the large firms would be flat even though their research capital is increasing, but the total value of their patents, unobservable, would be increasing. In a first difference model, this kind of measurement error on innovation output would lead to a research elasticity near 0 or even negative if research capital increases strongly enough while the number of patents is flat or decreases. The second argument is the exhaustion of innovation resources, more likely to concern advanced firms than small innovators. The evolution of patents for the biggest firm would then be flat even though research capital is steadily increasing. Clearly, more specification work has to be done on modeling the innovation process.

# 6 Conclusion

In this work we propose different estimations of the innovation function in which the output of the innovation process is measured by the number of patents. We implement new estimation methods based on Wooldridge and Chamberlain works, that allow for less demanding specifications that the ones proposed originally by Hausman, Hall and Griliches. The main methodological idea is to specify models via conditional expectations in which enters a multiplicative individual effect. The elimination of the firm-level effect leads to a set of orthogonality conditions that can be used in the GMM framework to provide robust estimates. This enables us to provide estimation of the correlated effect models,

[16] As we saw, it is possible to define an estimator without serial correlation but its implementation is tricky given the spurious maxima problem it implies.

including in the weak exogeneity case. Further restrictions can be imposed to get more e±ciency and can be tested, like constraining observed counts to be uncorrelated through time conditional on explanatory variables and the ¯xed e®ect. Other speci¯cations are be examined embedding the e®ect of past innovation.

This new set of speci¯cations is extensively explored. One of the most interesting ¯nding is that we do not reject the strict exogeneity version of the correlated e®ect model against its weak exogeneity counterpart. On the other hand, we nearly reject the model in levels at 5%. This is associated to an important decrease in the research capital coe±cient from 0.7 to 0.3 which is compatible with previous studies. We also examine the existence of restricted serial correlation and the impact of past patenting. However, there is one inconsistency between the di®erent results: on the one hand, we accept the existence of restricted serial correlation and, on the other hand, we also accept that past patenting a®ects the ability to innovate in the future. Thus, more work remains to be done in order to choose between competing speci¯cations.

Two points are on our research agenda. First, to study the di®erences in innovation productivity by cohort of research, taking ¯rms by subsample of research beginning date. This would let both the way research is accumulated and the impact of past innovation successes be a function of the age of research. Second, to use an unbalanced panel that would include younger ¯rms, that is with a higher patenting growth rate.

# References


[1] Andersen E (1970) Asymptotic properties of conditional maximum likelihood estimators. Journal of the Royal Statistical Society 32:283-301.

[2] Bussy J-C, Carpentier C, Kabla I (1995) Industrial property in France: the European patent. Economie et Statistique : special issue on patents.

[3] Chamberlain G (1992a) Comment: sequential moment restrictions in panel data. Journal of Business, Economics and Statistics 10.

[4] Chamberlain G (1992b) E±ciency bounds for semi parametric regression. Econometrica 60:567-596.

[5] Cr¶epon B, Duguet E(1995) Research and development, competition and innovation: pseudo maximum likelihood and simulated maximum likelihood applied to count data models with heterogeneity. Eco & Maths working paper 9508, CEME, University of Paris 1. Forthcoming in the Journal of Econometrics.

[6] Cr¶epon B, Duguet E, Kabla I (1996) Testing schumpeterian conjectures from various innovation measures in France. Forthcoming in Kleinknecht ed., Determinants of innovation: the message from new indicators, London: Macmillan Press 1996.

[7] Gouri¶eroux C, Monfort A, Trognon A (1984) Pseudo maximum likelihood methods: application to Poisson models. Econometrica 52:701-720.

[8] Griliches Z (1994) Productivity, R&D and the data constraint. American Economic Review 84:1-23.

[9] Hall B, Mairesse J (1995) Exploring the relationship between R&D and productivity in French manufacturing ¯rms. Journal of Econometrics 65:263-293.

[10] Hansen L (1982) Large sample properties of generalized method of moments estimators. Econometrica 50:1029-1054.

[11] Hausman J, Hall B, Griliches Z (1984) Econometric models for count data with an application to the patent-R&D relationship. Econometrica 52:909-938.

[12] Levin R, Klevorick A, Nelson R, Winter S (1987) Appropriating the returns from industrial research and development. Brookings Papers on Economic Activity 3:783-831.

[13] Newey W (1990) E±cient instrumental variables estimation of nonlinear models. Econometrica 58:809-837.

[14] Wooldridge J (1990) Distribution free estimators of some nonlinear panel data models. MIT working paper,Department of Economics 564.

[15] Wooldridge J (1991) Multiplicative panel data without the strict exogeneity assumption. MIT working paper,Department of Economics 574.

Table 1: Patent numbers distribution

| Number of patents | Percentage |
|---|---|
| 0 | 44.1 |
| 1 to 5 | 32.7 |
| 6 to 15 | 11.5 |
| 16 to 35 | 5.4 |
| 36 and more | 6.3 |
| Mean | 11.6 |
| Standard error | 49.0 |

Number of European patent applications over 1984-1989, the sample is 4188 points from 698 ¯rms.

Table 2: Research capital and patent numbers

| Year | n | ln k | n=k |
|---|---|---|---|
| 1984 | 1.47<br>(6.74) | 9.67<br>(1.90) | 0.11<br>(0.45) |
| 1985 | 1.65<br>(7.34) | 9.88<br>(1.79) | 0.13<br>(0.49) |
| 1986 | 1.81<br>(8.00) | 10.02<br>(1.73) | 0.14<br>(0.54) |
| 1987 | 1.99<br>(8.63) | 10.14<br>(1.70) | 0.15<br>(0.58) |
| 1988 | 2.28<br>(10.40) | 10.23<br>(1.68) | 0.17<br>(0.68) |
| 1989 | 2.39<br>(9.85) | 10.31<br>(1.67) | 0.18<br>(0.66) |
| Mean | 1.93<br>(8.17) | 10.04<br>(1.73) | 0.15<br>(0.54) |

Sample averages over ¯rms.

n : number of European patent applications.

k : research capital.

Table 3: Instrument set selection

| Model | Statistic[a] | Degrees of freedom | Probability |
|---|---|---|---|
| **Strict exogeneity** | | | |
| Without FE (Levels) | 149.46 | 126 | 0.0715 |
| Correlated FE (Within) | 131.07 | 105 | 0.0433 |
| Correlated FE (First difference) | 130.55 | 105 | 0.0462 |
| Correlated FE and restricted SC | 196.22 | 185 | 0.2721 |
| Correlated FE, restricted SC and 1 lagged dummy | 127.45 | 130 | 0.5469 |
| Correlated FE, restricted SC and 3 lagged dummies | 169.24 | 164 | 0.3733 |
| **Weak exogeneity** | | | |
| Correlated FE (First difference) | 43.29 | 35 | 0.1586 |
| Correlated FE and restricted SC | 111.60 | 95 | 0.1174 |
| Correlated FE, restricted SC and 1 lagged dummy | 77.20 | 70 | 0.2596 |
| Correlated FE, restricted SC and 3 lagged dummies | 97.91 | 94 | 0.3708 |

FE: fixed effects

SC: serial correlation

a: The null hypothesis is that the additional constraints from the $GMM_2$ instrument set are satisfied. Under the null, the statistics is distributed as a Chi-Square with the degrees of freedom indicated in the next column.

Table 4: Estimates under strict exogeneity

| Instrument set | GMM$_1$ | | GMM$_2$ | |
|---|---|---|---|---|
| Model | $k_{i;t}$<br>elasticity | Overidenti¯cation<br>test | $k_{i;t}$<br>elasticity | Overidenti¯cation<br>test |
| Without ¯xed e®ects<br>(levels) | 0.75<br>(0.04) | 34.23<br>df=35<br>p=0.50 | 0.82<br>(0.01) | 184.19<br>df=161<br>p=0.90 |
| Correlated e®ects<br>(Within) | 0.27<br>(0.10) | 21.43<br>df=29<br>p=0.16 | 0.28<br>(0.02) | 152.50<br>df=134<br>p=0.87 |
| Correlated e®ects<br>(First di®erence) | 0.26<br>(0.10) | 21.79<br>df=29<br>p=0.17 | 0.26<br>(0.02) | 152.34<br>df=134<br>p=0.87 |
| Correlated e®ects<br>(Pseudo maximum<br>likelihood, HHG [11]) | 0.26<br>(0.16) | { | { | { |
| Correlated e®ects and<br>restricted serial correlation | 0.30<br>(0.09) | 35.17<br>df=39<br>p=0.35 | 0.29<br>(0.004) | 231.39<br>df=224<br>p=0.65 |

GMM$_1$: $\underline{k}_{i;t}$ plus 6 time dummies.

GMM$_2$: GMM$_1$ and its independent cross products.

df: degrees of freedom for the overidenti¯cation test.

p: signi¯cance level of the overidenti¯cation test.

Table 5: Estimates under weak exogeneity

| Instrument set | GMM$_1$ | | GMM$_2$ | |
|---|---|---|---|---|
| Model | $k_{i;t}$<br>elasticity | Overidenti¯cation<br>test | $k_{i;t}$<br>elasticity | Overidenti¯cation<br>test |
| Correlated e®ects | 0.32<br>(0.22) | 10.79<br>df=14<br>p=0.30 | 0.33<br>(0.14) | 54.08<br>df=49<br>p=0.71 |
| Correlated e®ects and<br>restricted serial correlation | 0.31<br>(0.18) | 25.63<br>df=24<br>p=0.62 | 0.31<br>(0.03) | 137.23<br>df=119<br>p=0.88 |

GMM$_1$: $\underline{k}_{i;t}$ plus 6 time dummies.

GMM$_2$: GMM$_1$ and its independent cross products.

df: degrees of freedom for the overidenti¯cation test.

p: signi¯cance level of the overidenti¯cation test.

Table 6: Estimates with lagged patents

| Model | Strict exogeneity | | Weak exogeneity | |
|---|---|---|---|---|
| Instrument set | $GMM_1$ | $GMM_2$ | $GMM_1$ | $GMM_2$ |
| **With 1 lagged dummy** | | | | |
| $\mathbb{1}_{n_{i,t-1} \geq 1}$ | 0.25 | 0.24 | 0.24 | 0.30 |
| | (0.10) | (0.03) | (0.11) | (0.07) |
| $\ln k_{i,t}$ | 0.11 | 0.11 | 0.08 | 0.08 |
| | (0.13) | (0.01) | (0.20) | (0.04) |
| Statistic[a] | 51.18 | 178.63 | 41.27 | 118.47 |
| Degrees of freedom | 42 | 172 | 32 | 102 |
| Probability | 0.84 | 0.65 | 0.87 | 0.87 |
| **With 3 lagged dummies** | | | | |
| $\mathbb{1}_{1 \leq n_{i,t-1} \leq 5}$ | 0.15 | 0.17 | 0.19 | 0.24 |
| | (0.09) | (0.01) | (0.10) | (0.04) |
| $\mathbb{1}_{6 \leq n_{i,t-1} \leq 10}$ | 0.04 | 0.07 | 0.09 | 0.11 |
| | (0.11) | (0.01) | (0.13) | (0.04) |
| $\mathbb{1}_{n_{i,t-1} \geq 11}$ | 0.07 | 0.10 | 0.03 | 0.05 |
| | (0.12) | (0.01) | (0.15) | (0.04) |
| $\ln k_{i,t}$ | 0.14 | 0.14 | −0.06 | −0.05 |
| | (0.11) | (0.003) | (0.16) | (0.01) |
| Statistic[a] | 67.15 | 236.39 | 60.58 | 158.44 |
| Degrees of freedom | 60 | 224 | 50 | 144 |
| Probability | 0.75 | 0.73 | 0.85 | 0.81 |

All estimates are obtained under correlated fixed effects and restricted serial correlation.

a: It is the standard overidentification specification test. The null is that the orthogonality conditions are satisfied. Under the null, the statistic is distributed as a Chi-Square with the degrees of freedom indicated in the next line.

Table 7: Strict versus weak exogeneity

| Instrument set | $GMM_1$ | | | $GMM_2$ | | |
|---|---|---|---|---|---|---|
| Model | Statistic[a] | Degrees of freedom | Probability | Statistic[a] | Degrees of freedom | Probability |
| Correlated FE | | | | | | |
| First difference | 11.00 | 15 | 0.7526 | 98.26 | 85 | 0.1541 |
| Correlated FE and restricted SC | | | | | | |
| Without lagged dummy | 9.54 | 15 | 0.8476 | 94.16 | 105 | 0.7670 |
| With 1 lagged dummy | 9.91 | 10 | 0.4484 | 60.16 | 70 | 0.7931 |
| With 3 lagged dummies | 6.57 | 10 | 0.7653 | 77.90 | 80 | 0.5456 |

FE: fixed effects

SC: serial correlation

a: The null hypothesis is that the additional constraints from the strict exogeneity hypothesis are satisfied. Under the null, the statistic is distributed as a Chi-Square with the degrees of freedom indicated in the next column.