ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 49/50 — 1998

Reputation and Competence in Publicly Funded Science: Estimating the Effects on Research Group Productivity

Ashish ARORA, Paul A. DAVID, Alfonso GAMBARDELLA *

ABSTRACT. – This paper estimates the "production function" for scientific research publications in the field of biotechnology. It utilises an exceptionally rich and comprehensive data set pertaining to the universe of research groups that applied to a 1989-1993 research programme in biotechnology and bio-instrumentation, sponsored by the Italian National Research Council, CNR. A structural model of the resource allocation process in scientific research guides the selection of instruments in the econometric analysis, and controls for selectivity bias effects on estimates based on the performance of funded research units. The average elasticity of research output with respect to the research budget is estimated to be 0.6; but, for a small fraction of groups led by highly prestigious PIs this elasticity approaches 1. These estimates imply, conditional on the distribution of observed productivity, that a more unequal distribution of research funds would increase research output in the short-run. Past research publication performance is found to have an important effect on expected levels of grant funding, and hence on the unit's current productivity in terms of (quality adjusted) publications. The results show that the productivity of aggregate resource expenditures supporting scientific research is critically dependent on the institutional mechanisms and criteria employed in the allocation of such resources.

Réputation et compétences dans la recherche scientifique publique : Évaluation des effets sur la productivité des équipes de recherche

RÉSUMÉ. – Cet article évalue la « fonction de production » des publications de la recherche scientifique dans le domaine de la biotechnologie. L'article utilise un « data base » très vaste et exceptionnellement riche. Il concerne l'univers des équipes de recherche qui ont fait une demande de programmes de recherche de 1989 à 1993 en « biotechnologie et bio-instrumentation », financés par le Conseil National de recherche (CNR) italien. Un modèle structural du processus de distribution des ressources pour la recherche scientifique guide la sélection des instruments dans l'analyse économétrique et contrôle par effet de « selectivity bias » sur les évaluations fondées sur les « performances » des équipes de recherche. L'évaluation de l'élasticité moyenne de résultat (« output ») de la recherche a été 0,6, mais pour un petit nombre d'équipes de recherche dirigées par PIs prestigieux arrive à 1. Ces évaluations impliquent qu'une distribution unique des fonds pour la recherche augmenterait l'« output » de la recherche sur une courte période. Les publications passées ont un effet important sur les niveaux attendus de financement obtenu et, donc, sur la productivité courante des équipes de recherche en termes des publications (« quality adjusted »). Les résultats montrent que la productivité des frais pour la recherche scientifique, dépend fortement de mécanismes institutionnels.

* A. ARORA: Heinz School, Carnegie Mellon University; P. A. DAVID: Oxford University; A. GAMBARDELLA: University of Urbino. We thank the Director of the CNR Programme "Biotechnology and Bio-instrumentation", Professor Antonio De Flora, for his co-operation. Previous versions of this paper were presented to seminars at Carnegie Mellon University, the London School of Economics (STICERD), the NBER Summer Workhops in Industrial Organization (Cambridge), and the University of Pittsburgh. We have benefited particularly from the comments provided by Zvi Griliches, Bronwyn Hall, and Steve Nickell, and from Jacques Mairesse's encouragement in preparing this version of the paper for publication. We also acknowledge the financial support we received for various stages of this research from the International Centre for Economic Research (ICER) in Torino, the University of Urbino (MURST 60%), the EC Human Capital Mobility Programme (Contract N.ERBCHRXCT920002), the Renaissance Trust (UK), and MERIT, University of Limburg. Fabrizio Cesaroni, Marco Cioppi, and Aldo Geuna provided competent research assistance.

1 Introduction

The complex and multi-dimensional links between technological progress and scientific research have been recognised for a long time by economists as well as by science administrators and business managers. (See, e.g., DAVID [1993] for an overview.) Moreover, several recent quantitative studies have shown that there is a significant correlation between scientific research and technical change in industry (NELSON [1986], JAFFE [1989], MANSFIELD [1991], NARIN and OLIVASTRO [1992]).

Given this recognition, it is surprising how little attention economists have paid to the determinants of the productivity in scientific research. We know little about how increases in inputs affect the output of the research process, or how shifting marginal research expenditures across research groups with different characteristics would affect total research output. At a time when public support for scientific research is being questioned, and national research budgets are being subjected to retrenchment and restructuring in many countries, empirically-grounded answers to these and related questions are especially important for the sensible conduct of science policy.

Our approach to this problem is to develop a structural model of the process by which research units apply for and receive funds, and estimate the corresponding production function of scientific output, gauged in terms of (journal quality-weighted) publications. In implementing this approach we use an exceptionally complete and comprehensive data set from a pioneering Italian research programme in "Biotechnology and Bio-instrumentation" that was in effect during 1989-1993 under the auspices of the Centro Nazionale delle Ricerche (CNR), the Italian equivalent of the U.S. National Science Foundation (NSF).

We begin by discussing the general institutional context within which public resource allocation for scientific research takes place, and the implications for our modeling strategy. Section 3 describes the specific features of the CNR biotechnology data set, and motivates the development of the formal model in section 4. Section 5 presents the empirical estimates of the parameters of the model. Section 6 examines how changing the distribution of resources would affect the average productivity of research budgets in the short run, taking the characteristics of the population of research units as given. We also compute the estimated direct and indirect effects of past performance, in order to assess the potential way in which budget allocations affecting a unit's publication rate would impinge upon the reputational standing of its principal investigators (PI's), and so affect their expected future levels of public finding support from similarly organised public programmes. Section 7 summarises the findings and concludes the paper.

2 The Institutional Context

2.1. The Institutions of Scientific Research

In this paper we must move beyond the view of science as the pursuit of solitary researchers linked in “invisible colleges”. Research in the natural and life sciences has become a collaborative enterprise, carried on by very visible teams that are organised around increasingly expensive physical facilities and instruments. Yet, even the recent careful econometric studies by LEVIN and STEPHAN [1991], and STEPHAN and LEVIN [1992] continue the traditional individualistic focus of sociologists and historians of science, by investigating the life cycle productivity of individual academic scientists in the U.S. Their research seeks to estimate the effects of ageing through analysis of panel data, using fixed effects type procedures to control for unobserved differences. Such an approach could be justified as warranted by the emphasis that American public research programmes have placed upon grants to individual investigators, and the comparatively high degree of inter-institutional mobility that characterises university researchers’ careers in the U.S. Such a rationale is not uniformly appropriate across research areas, however, and it is considerably less apposite when applied to the western European institutional context. Moreover, although establishing individual life cycle profiles in productivity allows one to assess the implications for aggregate research productivity of changes in the demographic structure of the scientific community, the immediate policy relevance of such relationships is not so obvious. Policy options for manipulating the demographic composition, essentially those affecting age-specific rates of entry and exit from particular areas of scientific research, are likely to involve indirect and lagged effects that are both costly and difficult to control. By contrast, policy instruments targeting resource allocation and reward are likely to impinge more immediately on research productivity.

The mechanism for resource allocation in the world of open, academic science is different from that of the private sector in which business corporations set R&D budgets and manage the activities of employed scientists; and the situation of the non-profit research organization, whether that of a free-standing institute or department or research unit within a university, differs also from that of the individual scientist [1]. Scientific research groups obtain the bulk of their resources from public programmes in which government agencies offer research grants and contracts to competing applicants. Resources are allocated to selected groups according to the nature of the programme objectives and the scientific reputation that the team or of the unit has established in that area over an extended period of time – a reputation that often is linked with the unit’s leadership by one or a few senior scientists.

1. DAVID [1994], DASGUPTA and DAVID [1994] examine the efficiency implications of the institutional structures and reward systems characterising academic science. See also DASGUPTA and DAVID [1987].

If some of these group characteristics are only observed and taken into account by the funding agency (and possibly by the groups themselves), but cannot be seen by the econometrician studying the outcomes, and if some among those same characteristics also affect the production of observable scientific results by the group, then the research budgets allocated to the groups cannot be treated as an exogenous factor in the research production function. The "endogeneity" of this input implies that, in the absence of appropriate controls, there will be a bias in the estimated elasticity of research outputs with respect to the associated budget allocations. Therefore, in fitting a cross-section "production function", one needs a model of how resources are allocated among the groups, in order to choose meaningful econometric instruments and properly interpret the estimation results.

A model of the resource allocation process is also helpful in sorting out the different ways in which past performance, by affecting the scientific competence and professional reputation of the researchers associated with a particular unit, will be related to future performance. In addition to a direct competence-based effect, past performance may have two "indirect effects" on research output. First, units with better past records are more likely to be successful in getting research grants. Second, knowing this, they will invest in applying for larger grants. Both effects imply higher expected research budgets for these units, which (stochastically) raises their publication output rates. The model presented in this paper enables us to identify and separately estimate the direct and indirect effects of past performance upon group productivity.

The indirect effects of past performance deserve attention because they may underlie what has been referred to by ROBERT K. MERTON [1968], as the "Matthew Effect in Science" [2]. It is widely observed in studies of scientific productivity that a small fraction of the individuals accounts for the preponderant part of the body of published work (LOTKA [1926], PRICE [1963, 1976] ALLISON *et al.* [1976]). While differences in talent and ability may be part of the explanation for the pronounced left-skew that characterises distributions of individual research productivities in many specific fields of science, something further must be working to produce the phenomenon (also observed) of temporally increasing skewness of such distributions over the life of given cohorts of scientists. Institutional resource allocation mechanisms that would tend to differentially channel funding towards those who already have established a "track record" of research successes, is a likely candidate for this role in creating a dynamic "cumulative advantage" process. In other words, productive disparities also reflect the outcome of stochastic processes that cumulate advantage by

2. By allusion to the passages of the New Testament according to St. Matthew: "For unto everyone that hath shall be given, and he shall have abundance; but... from him that hath not shall be taken away even that which he hath". (Matthew 13:12 and 25:39.) In his original formulation of the Matthew Effect, Merton emphasised the disproportionately greater credit received for their contributions by scientists who had obtained a measure of eminence, but in subsequent work the original formulation was generalised by proposing that self-reinforcing processes affected productivity as well as recognition in science. See DAVID [1994: pp. 77-80] for further discussion.

amplifying the effects of initial heterogeneities in the productivity-related attributes of individuals, or research groups [3].

In this paper, we estimate a static model using cross-section data, and therefore cannot directly identify how initial advantage may cumulate over time. Nonetheless, a model such as ours would be a first step towards specifying the appropriate dynamic structure. In other areas of empirical economies, researchers typically have collected and analyzed cross-section data before proceeding to work with panel data, and we see our present analysis as a similar first step in the empirical research programme of the "new economics of science". But, more immediately, there are other issues of intrinsic economic interest and potential policy relevance that can be addressed directly through the analysis of cross-section data of the form that we have at our disposal, such as the impact of increased funding on research output in a given field of scientific inquiry.

2.2. **Biotech Research in Italy**

Biotechnology and Bio-instrumentation (henceforth B&B) was the first major public research programme in biotechnology in Italy, and virtually every Italian research group in molecular biology and genetic engineering, more than 800 in all, applied to it. We have been able to obtain information on the characteristics of both the units that were selected for funding, and those that were rejected. Thus, not only can we relate inputs to scientific output, but we can also correct for selection effects.

Unlike the US or UK, Italy is not considered a scientific powerhouse. Between 1989-1991, US based scientists accounted for about 40% of the publications in biomedical research, while Italian based scientists accounted for a little over 2.7%. By comparison, the corresponding figures for Germany, France and the UK are 6.2%, 5.2% and 7.8% respectively [4]. One may be tempted by this to suppose that a study of the Italian academic research sector is therefore of limited value for understanding the economics of modern scientific research in biotechnology, let alone as a basis for broader generalisations. We believe such a view to be mistaken. For one thing, certain structural features of national (and international) research processes are preserved when scale changes. An instance of this, which we will show below in greater detail, is that the distribution of publications for Italian scientists looks no different from that observed in other countries – it is highly skewed, with a small fraction of the researchers accounting for a large fraction of the total publications generated by the population as a whole [5].

---

3. ALLISON *et al.* [1982] surveys the sociological literature in cumulative advantage processes. For economists' view on this subject, see the discussions in DAVID [1994], STEPHAN [1996], and ARORA and GAMBARDELLA [1997].

4. See *The European Report on Science & Technology Indicators* [1994].

5. See figure 2b below. Recall that our sample is close to the relevant universe. The equivalent sample for the US would not be simply the leading research labs but any research lab working in molecular biology or genetic engineering in the country.

Moreover, the relatively small size of the Italian molecular biology sector conveys at least one significant advantage for the purposes of this study. It allows us to neglect the constraint on aggregate output that arises from the fixed number of scientific journals. Put differently, since US scientists author a large fraction of the publications in international journals, their aggregate publication output is constrained by the growth rate of the number of existing journals.

Estimates of the marginal product of research inputs based on US data may be subject to a downward bias on this account. This problem is much less severe in the case of Italy.

Another source of potential doubts about the value of relying upon empirical findings about the scientific research process based upon the Italian CNR experience is the casual supposition that non-scientific, political considerations may intrude into the details of the funding agency's decision process to a degree that is not present in the peer review processes through which resources are allocated in, say, Britain or the US. We believe that in the particular CNR programme we have studied here such a suspicion is unwarranted, and the processing of proposals for funding was conducted in a way that conformed with the norms of scientific peer review. Moreover, our measures of scientific "output" are publications weighted by the quality of the publishing journal, using the so-called "impact factors" of journals based upon the computations of the *Science Citation Index.* Hence, the productivity of the Italian biotechnology research units is being evaluated by the same standards that are used to measure the publication outputs of any other international scientific community [6]. Furthermore, this is not a matter that is left for conjecture; we shall explicitly model the selection process, as well as the setting of the research budgets that the selected units receive, and thereby allow the data themselves to reveal the extent to which the programme from which the data used here have been drawn was broadly conformable with our priors based on information about corresponding institutional arrangments and procedures in the Anglo-Saxon world.

3 Data Description

3.1. Variables Used in the Analysis

B&B is a five-year programme (1989-1993) for research in molecular biology and genetic engineering issued by the Italian CNR in 1987. The programme was divided into seven sub-programmes. The first six

6. We may point out again that the small relative size of the Italian research community is a virtue in this respect, since their proportionate contributions to the aggregate citations of journal article, like their proportionate contributions to the international journals' contents, are so small that they cannot be thought capable of influencing the relative impact factors of those journals.

were concerned with various sub-disciplines of molecular biology and genetic engineering. Sub-Programme 7, Bio-Instrumentation, focused on development and experimentation of scientific instrument prototypes. A total of 858 research laboratories applied to B&B, with universities and CNR in-house laboratories accounting for about 62% and 15% respectively. The rest of the applications were from other non-profit research institutions such as foundations and hospital research labs, as well as some commercial firms. The research groups that applied to the programme are well-defined units of scientific production. They are teams of scientists, researchers, technicians and other personnel within established institutions, which are stable over time. They were not formed to carry out just this project. Of the original 858 units, CNR selected 360 for funding. Due to missing data, our final sample is composed of 797 units, of which 347 were selected for funding.

Table 1 defines all the variables used in our empirical analysis. We collected most of our data from the application forms. B&B had an explicit goal of encouraging industrial "transferability" of research. Applicants had to indicate whether the project had potential industrial uses, and if so, who the potential users were. We summarised this information in a dummy variable, *TRANSF*, which takes the value 1 if the applicant declared that his project had potential practical uses, and indicated the name of one or more firms that could use those results. These are projects that signal concrete opportunities of application as the units were able to indicate precise names of industry users. CNR programmes typically have as one of their policy objectives the encouragement of research in the less advanced regions of the country, and particularly in the "Mezzogiorno" or the southern part of Italy. The dummy variable, *DSOUTH*, takes the value 1 for units located in the "Mezzogiorno". The variable, *DPRO7*, is a dummy variable for bio-instrumentation related proposals. We also created dummies *DCNR* and *DUNI* for CNR and university labs. Other characteristics of the units are:

TABLE 1

Definition of Variables Used in the Analysis.

I	Index for selection – dummy equal to 1 if the unit was granted a positive budget
B_A	Total 1989-1991 budget asked by the units (in millions Italian Lire)
B_G	Total 1989-1991 budget granted (in millions Lire)
PUB	Quality-adjusted number of publications of the units that acknowledged contribution of this Programme
DPRO7	Dummy for units in sub-Programme 7 (Bio-Instrumentation)
DCNR	Dummy for CNR laboratories
DUNI	Dummy for university laboratories
DSOUTH	Dummy for units located in the South
TRANSF	Dummy for industrial "transferability" of the project
SIZE	Size of the group (number of people)
K	Quality-adjusted 1983-1987 publications of the PI listed in the application form
COLLAB	Number of research collaborations with foreign non-profit institutions listed in the application form
NUIST	Number of units from the same institution of the applicant that applied to the programme (e.g. University of Rome, CNR of Naples)
PROV_POP	Total 1987 population in the province of the unit (thousands)
AGEPI	Age of the PI (years)

the size of the research unit (*SIZE*) measured by the number of researchers and technicians in the unit; *AGEPI*, the age of the PI, *NUIST*, the number of units from the same institution (e.g. University of Rome, CNR of Naples) that applied to the programme; and *PROV_POP*, the population of the province wherein the unit is located.

CNR supplied us with the list of units that were selected, and the total budget granted to each in each of the five years of the programme. From the CNR we also obtained data on the total number of publications produced by the selected units. These are all the publications available in 1994 that explicitly acknowledged the financial support of this programme. As B&B actually ended in late 1994, the CNR warned us that some of its results were yet to be published, and these publications referred mostly to activities conducted in the first three years of the programme (1989-1991). Accordingly, we define our budget requested and budget granted variables as the amounts pertaining to the first three years of the programme. Since annual budgets tended to be constant over time, this involved little more than simply scaling the variables by three fifths.

To weight publications so as to measure output in comparable units of "quality", we employ the 1987 impact factor (as computed by the *Science Citation Index*, SCI) for the respective journals in which the units' publications appeared [7]. Using the number of citations to the papers is an alternative way to weight for quality. However, because the papers produced in the programme were all relatively recent at the time of data collection, citation measures were likely to be biased. Instead, we define $PUB = \Sigma_j (s_{ij} w_j)$, where s_{ij} is the number of publications of the i-th unit in the j-th journal, and w_j is a linear function of the impact factor of the j-th journal. We experimented with a variety of specifications for w_j. Here we report results using $w_j = 0.5 + IF_{ij}$ [8].

As a measure of past performance of the group, we used the 1983-1987 publications of the PI listed in the application form (K) [9]. These were adjusted for quality in the same way as the publication output. Even though these publications are "older", and the citations are therefore more complete, we have chosen to be consistent with the quality measure used for the units' publication output [10]. We used the research collaborations (*COLLAB*) of the unit with foreign non-profit institutions as another measure of quality. One can think of the past publications of the PI as a measure of the quality of

---

7. The IF is the ratio between the number of citations of the journal by other journals, and its number of citations to other journals. A high value of the IF thus indicates a journal that is cited more frequently than it cites. In our data, IF for journals ranged from close to 0 to about 15. *Nature* for instance had a 1987 IF of 14.77. Articles in books and working papers have a nominal IF of 0.
8. We experimented with $1 + IF_{ij}$ as a weight, as well as simply IF_{ij}. These correspond to giving a high weight to working papers and journal articles to giving no weight to these. After some casual search, we settled on the specification noted above. In all cases, the results were substantially unchanged.
9. Along with other information, the units had to list all relevant publications (to this programme or in related areas) of the PI and of other members of the unit in the previous five years (1983-1987).
10. Note that the research output is of the entire research unit, rather than just the PI.

the unit, while *COLLAB* measures the quality of the project, although the two overlap. As with other self reported variables, we assume that research collaborations are exogenously given, at least in the short run.

3.2. **Correlations & Reduced Form Regressions**

Table 2 presents descriptive statistics for out data. Figures 1 and 2 show the distributions of budget asked (B_A), budget granted (B_G), publication output (*PUB*), and past publications of the PI (K). All four distributions are skewed, especially the two distributions of publications. The log-normal distribution appears to be a good approximation to these distributions, and is the one specified in our structural model.

Tables 3 and 4 show reduced form regressions of B_A, B_G and *PUB*, for all 797 units as well as for the 347 units that were selected for funding. We also present the publication equations with and without the budget among the regressors. Table 3 shows that both expected budget asked and budget granted are positively related to transferability, past publications, foreign collaborations and the size of the unit. However, while budget asked by research units from the South was 116 million Lire more than the average, the expected budget given to them conditional on selection is actually lower than the corresponding average. As we show below, this behaviour can be motivated by a higher than average probability of selection, as well as a lower per unit effort cost of preparing larger research proposals.

As Table 4 shows, past publications has a large impact on publication output even after we add budget granted amongst the regressors. Together with the results from Table 3, this suggests and past performance may have

TABLE 2

Descriptive Statistics (797 observations).

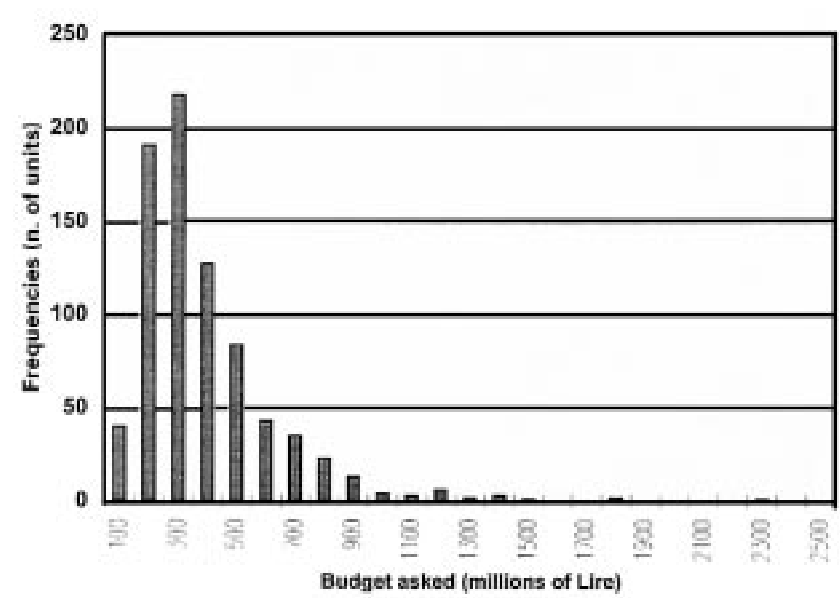
	Mean	Std Dev	Minimum	Maximum
I	0.435	0.496	0	1
B_A (*)	350.244	275.743	25	4224
B_G (*) (+)	105.035	62.147	3	519
PUB (+)	17.468	22.442	0	199.19
(non quality-adj. number)	(6.761)	(6.065)	(0)	(50)
DPRO7	0.107	0.309	0	1
DCNR	0.154	0.361	0	1
DUNI	0.645	0.479	0	1
DSOUTH	0.189	0.392	0	1
TRANSF	0.375	0.484	0	1
PROV_POP (§)	195.260	126.290	100	3477
COLLAB	4.018	3.323	0	25
K	31.672	43.556	0	636.79
(non quality-adj. number)	(13.065)	(17.020)	(0)	(226)
SIZE	12.055	6.221	2	99
NUIST	25.152	18.377	1	63
AGEPI	52.287	8.490	34	85

(*) Millions Italian Lire; (+) 347 observations for $I = 1$; (§) thousands.

FIGURE 1

Budget asked, Budget granted.
(Frequency distributions, 797 observations).

a) Budget Asked – B_A.



b) Budget Granted – B_G.

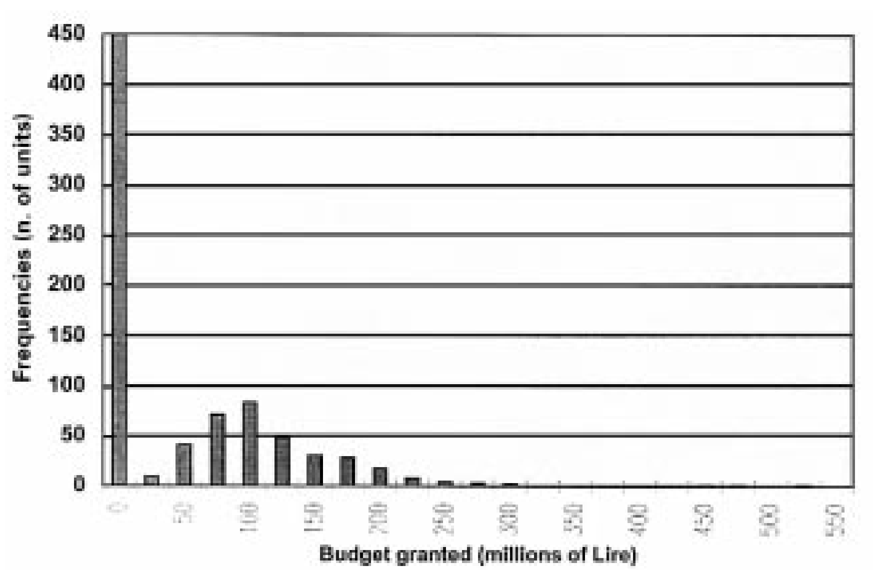
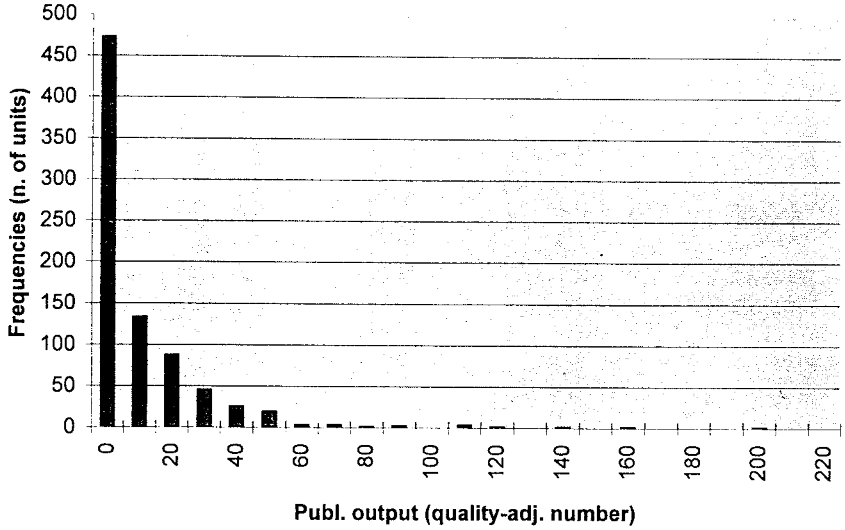


FIGURE 2

Publication Output, Past Publications (Frequency distributions, 797 observations).

a) Publication Output – PUB.



b) Past Publications – K.

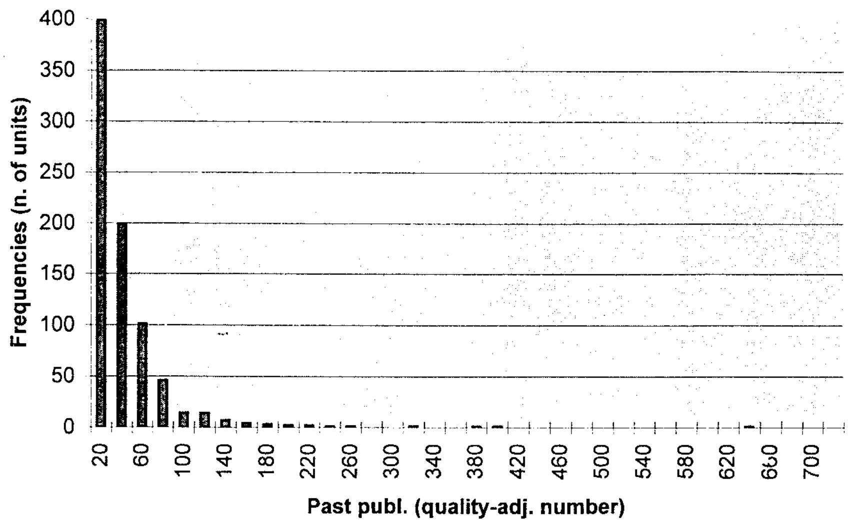


TABLE 3

OLS Estimates: (Budget asked, Budget granted).

	$\ln B_A$	$\ln B_A$ $(I=1)$	$\ln B_G$	$\ln B_G$ $(I=1)$
Const	3.497	3.691	– 3.294	3.742
	(0.530)	(0.722)	(2.009)	(0.782)
DPRO7	0.165	0.141	0.577	0.413
	(0.079)	(0.106)	(0.253)	(0.121)
DCNR	– 0.060	0.033	0.763	0.244
	(0.067)	(0.093)	(0.257)	(0.111)
DUNI	– 0.038	0.045	0.380	0.050
	(0.057)	(0.092)	(0.212)	(0.115)
DSOUTH	0.274	0.231	0.100	– 0.127
	(0.051)	(0.077)	(0.185)	(0.078)
TRANSF	0.122	0.109	0.587	0.095
	(0.040)	(0.054)	(0.159)	(0.061)
ln(SIZE)	0.547	0.543	0.426	0.177
	(0.059)	(0.093)	(0.167)	(0.074)
ln(K)	0.065	0.105	0.646	0.092
	(0.021)	(0.027)	(0.077)	(0.033)
ln(COLLAB)	0.102	0.118	0.359	0.042
	(0.030)	(0.039)	(0.123)	(0.043)
ln(NUIST)	– 0.066	0.006	– 0.074	– 0.017
	(0.026)	(0.037)	(0.086)	(0.042)
ln(PROV_POP)	0.041	– 0.001	0.294	0.031
	(0.027)	(0.039)	(0.096)	(0.046)
ln(AGEPI)	0.101	0.037	– 0.130	– 0.089
	(0.129)	(0.160)	(0.461)	(0.199)
No of obs	797	347	797	347
Adj.R^2	0.242	0.308	0.174	0.089

Heteroskedastic consistent standard errors in parenthesis.

a direct and an indirect effect on publication output. The impact of budget given varies considerably between the full sample and the restricted sample. This problem arises because both, selection and the amount of funding, are potentially correlated with unobserved variables that also affect output. The structural model set out in the following section attempts to address both issues: selection and endogeneity of budget.

4 The Model and the Estimated Equations

4.1. Two Caveats

• **Knowledge *vs.* reputation capital:** As remarkable as the available data set is, it does not permit us to distinguish empirically between a

TABLE 4

OLS Estimates:* PUB *(Publication Output).

	$\ln B_A$	$\ln B_A$ $(I=1)$	$\ln B_G$	$\ln B_G$ $(I=1)$
Const	– 1.688	3.302	– 0.128	2.239
	(1.194)	(1.577)	(0.733)	(1.591)
$\ln(B_G)$ (*)	–	–	0.474	0.284
			(0.014)	(0.100)
DPRO7	– 0.108	– 0.720	– 0.382	– 0.837
	(0.117)	(0.196)	(0.091)	(0.194)
DCNR	0.193	– 0.019	– 0.168	– 0.089
	(0.155)	(0.179)	(0.095)	(0.177)
DUNI	0.149	– 0.041	– 0.031	– 0.055
	(0.140)	(0.179)	(0.079)	(0.174)
DSOUTH	– 0.178	– 0.570	– 0.225	– 0.534
	(0.110)	(0.160)	(0.080)	(0.160)
TRANSF	0.184	– 0.106	– 0.094	– 0.133
	(0.096)	(0.114)	(0.062)	(0.113)
ln(SIZE)	0.163	0.105	– 0.039	0.054
	(0.099)	(0.132)	(0.071)	(0.132)
ln(*K*)	0.498	0.435	0.192	0.409
	(0.052)	(0.065)	(0.035)	(0.065)
ln(COLLAB)	0.270	0.220	0.100	0.208
	(0.165)	(0.078)	(0.044)	(0.077)
ln(NUIST)	– 0.091	– 0.003	– 0.055	0.002
	(0.049)	(0.062)	(0.033)	(0.061)
ln(PROV_POP)	0.156	– 0.075	0.017	– 0.084
	(0.056)	(0.079)	(0.035)	(0.077)
ln(AGEPI)	– 0.086	– 0.519	– 0.024	– 0.493
	(0.273)	(0.365)	(0.167)	(0.356)
No of obs	797	347	797	347
Adj.R^2	0.202	0.239	0.683	0.256

Heteroskedastic consistent standard errors in parenthesis. (*) For $I = 0$, $\ln(B_G)$ set equal to 0.

measure of researchers' knowledge capital or "competence", and a measure of their scientific "reputation" capital. Past performance may be related to future performance both because the two reflect some inherent (fixed) attribute of the researchers that affects their productivity, and because superior performance enhances reputation, and thereby results in access to more generous funding. This open up the possibility that past performance may reflect not just inherent differences among researchers' capabilities, but also small idiosyncratic factors of personality, or extraneous circumstances unrelated to talent but nonetheless affecting early funding success. When initial success is rewarded with greater funding, then this increases the likelihood of future success as well. If funding agencies estimate future productivity based on past publication performance without taking into account past levels of funding, this gives rise to state dependence in the production process. Were one to allow for learning effects, gained through the experience of carrying out funded projects, the state dependent nature of the process would become even more marked [11]. Using cross-section data, as we do, one cannot separate state dependence from unobserved heterogeneity. What we seek to do in this paper is to separate the direct effects from the indirect effects of "knowledge capital" on the likelihood of units receiving funding – a reflection of the working of the external institutional mechanisms rather than the internal production capabilities of the group.

• **Marginal *vs*. total inputs and outputs.** In a given year, a unit may choose to work on more than one project. Unfortunately, we do not observe the inputs (funding) received by the units for other projects. This is potentially a source of bias if the projects are inter-dependent in either inputs or output. The intuition is straight forward. Suppose each unit can work on two projects. One of these is the CNR project and the other is an alternative project (possibly less attractive). The productivity, per unit of research effort, in the latter may fall if the research unit also works on the CNR project. This implies that merely looking at the inputs and output of the CNR project alone would lead to an over-estimate of the marginal product of research effort. Similarly, the selected units may indicate as an ouptut of the programme publications supported by other funds, leading to "double counting" and an over-estimate. Conversely, if there are positive spillovers across projects, one will end up with an under-estimate.

Since we do not observe the funding for any of the unit's projects other than the one studied here, we will formally assume that the CNR project is independent of its other activities. As a practical matter, this is a reasonable approximation. The average funding for selected project in our sample is 105 million Lire over a period of three years, which is equal to about $65,000 (see Table 2). We do not have a firm benchmark to compare but this clearly is a small enough number to imply that this particular CNR programme was not intended as the major funding source for most of the units. Similarly, based on data from the *Science Citation Index* (provided to us by the ISI), we estimate that the average PI among the selected units in our sample produced about 4.7 publications (unweighted) per year in the five year period prior to the programme [12]. By comparison, for this programme, the annual publication output for an entire research unit was only 1.3 during 1989-1991. Note also that during this time interval, total publications in the area of bio-medical were growing at over 2% per annum, implying that the programme-related publication output of the unit as a whole is less than a fourth of the publication output credited to the typical PI.

Moreover, in the foregoing, the precise meaning of 'independent' should be understood as applying to the "direct cost" aspects of the CNR project. What we assume here is that research outputs that have been directly financed from other sources are not attributed by the unit to its project in the CNR's B&B programme; and, likewise, that funds obtained from the latter programme are not diverted to support research and publications that

11. Indeed, if either or both sources of positive feedback are sufficiently strong, the microlevel dynamics of the stochastic process governing publication output and reputational status will become non-ergodic and path dependent. See David [1994: pp. 80-84] for further discussion.
12. The PIs of the selected units produced 8197 publications during the period 1983-1987. The total publication from the programme between 1989-1991 was 1367.

fall outside that programme's (and hence would not be reported among the CNR project outputs). Our assumptions here recognise the possibility that there are elements of "joint-production" in the research unit's operations. The latter are, indeed, quite likely in the case of the larger units which hold a number of grants and/or contracts; the costs of indivisible elements of the "infrastructure" (both staff and facilities) may be met through a policy of levying what are in effect "internal overhead" charges on all concurrently running projects in the unit.

4.2. **The Model**

Our objective here is to derive three equations that can be estimated. The first two are the equations for budget requested and budget granted. The third equation is the production function of publication output.

• **Research Units.** We assume that the research units maximise their expected research output. They have two choice variables – the size of the project, which we measure by the budget asked, and the amount of research effort (unobserved by us), also reckoned in monetary units. We assume that increasing the size of the project is costly, and these costs are borne by the research unit. Furthermore, actual research effort will differ from the *ex ante* research effort because the units cannot fully predict their actual research budget. In a fully specified model, the units would have to satisfy an inter-temporal budget constraint. Since we only have a cross-section, and since it is possible that units may have other sources of research support, we specify in (3) and (3a) below an *ad hoc* rule linking expected budgets, and the *ex ante* and actual research effort. By substituting in this rule, the unit's decision problem effectively reduces to the choosing the optimal size of the research project, B_A.

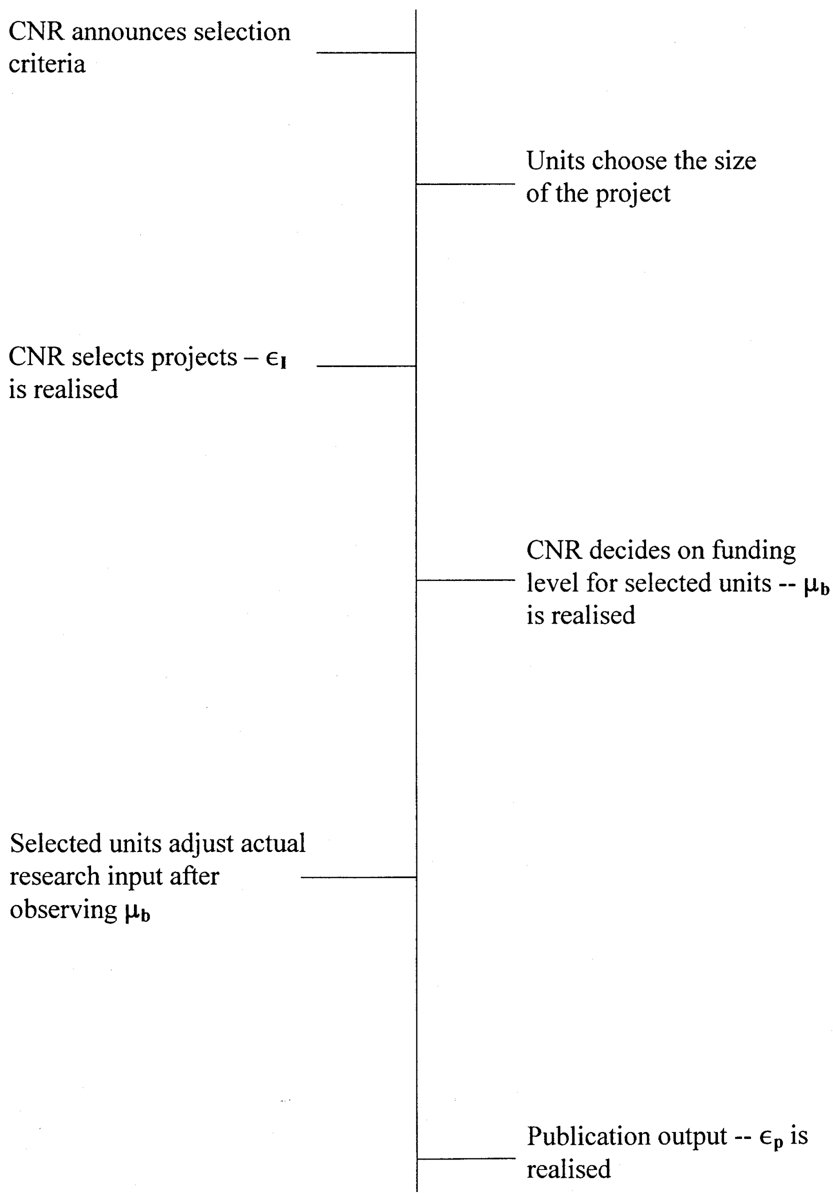
• **CNR Decision Making Process.** The CNR is assumed to follow a two-step decision procedure. The first is a dichotomous decision, whereby projects are either selected for funding, or rejected. In the next step, the actual funding levels for the projects selected in the first step is decided. We assume that the budget requested for the project is not included as a criterion in the first step. Although we have chosen this representation for modeling convenience, we believe that this is a fairly accurate representation of the actual decision making procedures followed by public research agencies. Although excluding the budget requested from the first step may appear to be a very strong restriction because projects with very large (or very small) budget requests may be deselected at the first stage, in practice there is informal communication between agencies like the CNR and the research units. In this way, otherwise worthy projects may be modified to fit into any budget criterion that the agency may have implicitly imposed. Note that if this informal (and unobserved) communication takes place, the situation is difficult to distinguish empirically from one where no budget criterion exists.

In a fully specified model, one would also derive the CNR's decision rule as an optimal response to the decision rules followed by the units. That, however, would entail the introduction of a great deal more structure, which presently we lack the necessary data to identify econometrically.

This approach must therefore be deferred for future research based on a still more extensive data set.

Figure 3 shows the sequence of actions from the launch of the programme to production of publications. When the programme is started, the agency sets a rule for selecting projects and for allocating budgets. This amounts to defining the two-step procedure characterised by (1) and (2) below. Given these rules, each unit chooses the optimal size of the research project so as to maximise its utility. The CNR then selects from amongst the projects,

FIGURE 3



and allocates budget. These decisions depend in part upon characteristics of the unit and project observed only by the CNR (ε_I and μ_b), in part upon characteristics observed by the CNR and the research unit (ε_k), and in part upon publicly observed characteristics of the unit and the project, denoted by $\mathbf{X}$. While we assume that μ_b is distributed independently of ε_I, we allow for ε_k to be correlated with ε_I. Units observe the actual budget, and adjust their planned research effort according to (3). Finally, ε_p is realized, and publication output is determined according to (4). We allow ε_p to be correlated with ε_I, as well as with ε_k.

4.3. Notation

Let:

R	research effort of unit
R^*	"planned" research effort
B_G	budget granted by the agency
B_A	budget requested by units
I^*	latent selection variable
I	index with $\mathbf{I = 1 \leftrightarrow I^* > 0}$
X, Y, Z	other characteristics of unit and project
$C\,(A, Z)$	application and set-up costs of a project of size A
P	publication output
K	past publication output
H	other characteristics of the units that influence the production of publications

We also define the following expressions.

- **Selection equation**

$$I^* = \pi \ln Y + \varepsilon_I \tag{1}$$

where $\varepsilon_I \sim N\,(0,\,1)$. Define ϕ, Φ to be the standard normal and the standard cumulative normal such that $\Phi = \mathrm{Prob}\,(I = 1)$.

- **Budget granted equation**

$$\begin{cases} B_G = B_A^\lambda \, X^\theta \, e^{\varepsilon_k + \mu_b} & I = 1 \\ B_G = 0 & I = 0 \end{cases} \tag{2}$$

where $\varepsilon_k \sim N\,(0,\,\sigma_k)$ is a measure of quality, observed by the units and the CNR and not by the econometrician. The other term, $\mu_b \sim N\,(0,\,\sigma_b)$, is independent of ε_k and ε_I, and represents the uncertainty, from the viewpoint of the research unit, in the actual budget granted. It captures any unobserved

CNR preferences that are not directly related to the publication performance of the research units [13].

- **Research effort**

$$\begin{cases} R = R^* \, e^{\delta \mu_b} & I = 1 \\ R = 0 & I = 0 \end{cases} \tag{3}$$

where R^* is "planned" research, defined as

$$R^* = E_{\mu_b} \left(B_G - C | I = 1, \, \varepsilon_k \right) \tag{3a}$$

Note that this implies that research effort is not simply proportional to the allocated budget. Instead, the relationship between research effort and the budget depends, indirectly, also upon the costs, and other characteristics of the research unit.

- **Production Function of Publications**

$$\begin{cases} \ln \left(PUB \right) = \alpha \cdot \ln R + \gamma \cdot \ln H + \varepsilon_p & I = 1 \\ \ln \left(PUB \right) = 0 & I = 0 \end{cases} \tag{4}$$

where $\varepsilon_p \sim N \left(0, \, \sigma_p \right)$ accounts for stochastic factors in the production of publications. Note that we allow α, the elasticity of output with respect to research effort, to vary with quality of the unit.

- **Cost equation**

$$C \left(A, \, Z \right) \equiv c \left(Z \right) \cdot A \equiv Z^{-\eta} \cdot A \tag{5}$$

4.4. Optimal Project Size

We assume that units choose B_A to

$$\operatorname{Max} E_I \, E_{\mu_b, \, \varepsilon_p} \left(\ln PUB | \varepsilon_k \right) \equiv \Phi \, E_{\mu_b, \, \varepsilon_p} \left(\ln PUB | I = 1, \, \varepsilon_k \right)$$

Notice that $E_{\mu_b, \, \varepsilon_p} \left(\ln PUB | I = 1, \, \varepsilon_k \right) \equiv \alpha \ln R^* + \gamma \ln H + E_{\mu_B, \, \varepsilon_p} \left(\alpha \delta \mu_b + \varepsilon_p | I = 1, \, \varepsilon_k \right)$. Since B_A does not affect selection, Φ and $E_{\mu_B, \, \varepsilon_p} \left(\alpha \delta \, \mu_b + \varepsilon_p | I = 1, \, \varepsilon_k \right)$ are also independent of B_A. Thus the problem of the units boils down to

$$\operatorname{Max} R^* \equiv E_{\mu_b} \left(B_G - C | I = 1, \, \varepsilon_k \right) = B_A^{\lambda} \, X^{\theta} \, e^{\varepsilon_k} \, E_{\mu_b} \left(e^{\mu_b} \right) - Z^{-\eta} \, B_A$$

The first order condition of this problem yields

$$C \left(\cdot \right) = \lambda \, E_{\mu_b} \left(B_G | I = 1, \, \varepsilon_k \right) \tag{6}$$

---

13. We use the conventon that ε's denote errors whose expected value depends upon I, and μ's errors whose expected value does not depend on I.

4.5. **The Estimated Equations**

We now derive our three equations to be estimated for selected units [14]. To derive the budget requested and budget granted equations, substitute for B_A from (6) in the expression (2) for B_G, conditional on $I = 1$. This gives

$$\text{(7)} \quad \ln B_A = (1-\lambda)^{-1} \left(\ln \lambda + \sigma_b^2/2 + \frac{\theta}{1-\lambda}\, x + \eta \cdot z + \sigma_{IK}\, \frac{\phi}{\Phi} + \mu_k \right)$$

$$\text{(8)} \quad \ln B_G = (1-\lambda)^{-1} \left(\lambda \cdot \ln \lambda + \lambda \cdot \sigma_b^2/2 + \frac{\theta}{1-\lambda}\, x + \lambda \cdot \eta \cdot z + \sigma_{IK}\, \frac{\phi}{\Phi} + \mu_k \right) + \mu_b$$

where σ_{IK} is the covariance between ε_1 and ε_k, and $\mu_k \equiv \varepsilon_k - \sigma_{IK}\, \frac{\phi}{\Phi}$ [15]. Combining this with (3), (3a), and (6), we can write

$$\text{(9)} \quad \ln PUB = \alpha \left[\ln (1-\lambda) + b_e^* \right] + \gamma \cdot \ln H + \sigma_{IP}\, \frac{\phi}{\Phi} + \alpha\,(K) \left(\frac{\mu_k}{(1-\lambda)} + \delta \mu_b \right) + \mu_p$$

where σ_{IP} is the covariance between ε_p and ε_I, $\mu_p \equiv \varepsilon_p - \sigma_{IP}\, \frac{\phi}{\Phi}$, and $b_e^* \equiv \ln E_{\mu_B, \varepsilon_k} (B_G | I = 1) = \hat{b} + \frac{\sigma_b^2}{2} + \frac{\sigma_{IK}}{1-\lambda}\, \frac{\phi}{\Phi}$, where $\hat{b}$ stands for the first four terms of equation (8).

Equation (9) is the production function of publications that we estimate. Note that our model enabled us to transform the production function defined by equation (4), which depended on the unobserved research effort of the units, into an expression that depends on a variable b_e^*, the expected budget conditional on selection and a parameter λ which can be retrieved from the estimated parameters of equations (7) and (8). In addition, equation (9) enables us to estimate the elasticity with respect to the budget (or research effort) after specifying a functional form for $\alpha\,(K)$.

5 Estimation and Results

5.1. **Estimation Strategy, Regressors, and Identification**

We first estimate the selection equation (1) as a probit, using all observations. This produces estimates of ϕ and Φ evaluated at $\hat{\pi} y$, where

14. This procedure implies that information from the non funded units is only used to estimate the selection equation.
15. Note that by accounting for the covariance between ε_I and ε_k, we are following what amounts to the standard Heckman-Mills procedure for correcting sample selection. In our case, selection correction has to be done in the budget equations as well as the publication equations.

$\hat{\pi}$ is the estimated value of π. We then estimate (7), (8), and (9) for observations $I = 1$ after substituting the estimated ϕ and Φ. This two-step, Heckman-Mills correction procedure was chosen because it is convenient and robust. Using conventional terms, selection correction has to be applied both in the budget equations, as well as in the publication equation.

We next estimate (7) and (8) jointly by GLS. We use the estimated values of λ, $\hat{b}$, σ_b^2, and σ_{IK} to compute $\ln(1-\lambda) + b_e^*$ which we substitute in (9), and estimate (9) as a Tobit using a variety of specifications for $\alpha(K)$ [16]. Since the data are distributed as log-normal, we use the log-log specification throughout [17]. We also estimated specifications in levels but the fit to the data was poorer, although the point estimates of elasticities were remarkably similar to those reported here.

• **Regressors.** As discussed earlier, we use all exogenous variables (vector $\mathbf{Y}$ below) except size and budget asked to predict selection. We also exclude *DPRO7* because CNR did not commit *ex ante* to select given percentages of applicants from each sub-programme, but examined the projects altogether and selected them according to characteristics such as quality, and transferability. We impose this restriction primarily to preserve degrees of freedom; it is not critical for identification and a formal test of the restriction implies that the restriction is not rejected.

Recall that the variables in $\mathbf{Z}$ affect the cost of preparing a project of "size" B_A; the variables in $\mathbf{X}$ account for factors that affect the fraction of B_A granted; the vector $\mathbf{H}$ accounts for exogenous variables that affect the productivity of publications. We used the following specification for $\mathbf{Z}$, $\mathbf{X}$, and $\mathbf{H}$ (corresponding θ, η, and γ parameters in parenthesis):

Z-regressors	**X**-regressors	**H**-regressors
• const (η_0)	• const (θ_0)	• const (γ_0)
• DPRO7 (η_7)	• DPRO7 (θ_7)	• DPRO7 (γ_7)
• DCNR (η_{CNR})	• DCNR (θ_{CNR})	• DSOUTH (γ_S)
• DUNI (η_{UNI})	• DUNI (θ_{UNI})	• TRANSF (γ_T)
• DSOUTH (η_S)	• DSOUTH (θ_S)	• ln(N) (γ_N)
• TRANSF (η_T)	• TRANSF (θ_T)	• ln(K) (γ_K)
• ln(K) (η_K)	• ln(K) (θ_K)	• ln(RCF) (γ_C)
• ln(RCF) (η_C)	• ln(RCF) (θ_C)	• ln(NUIST) (γ_U)
• ln(NUIST) (η_U)	• ln(NUIST) (θ_U)	• ln(PROV_POP) (γ_P)
• ln(PROV_POP) (η_P)	• ln(PROV_POP) (θ_P)	• ln(AGEPI) (γ_A)
• ln(AGEPI) (η_A)	• ln (AGEPI) (θ_A)	
• ln(SIZE) (η_N)		

The **Z**- and **X**-regressors include all the variables used for selection. Unlike $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{Z}$ include *DPRO7*. Bio-instrumentation projects are more costly to prepare. To present a "credible" proposal in this area, units have to show that they will be able to utilize expensive equipment or facilities

16. We estimate (9) separately by Tobit because more than 10% of the selected units produced zero publications.
17. For variable that take on a value of zero, we added one to that variable when taking logs.

to carry out development and testing activities. Thus, the organization of the proposal may require, to a greater extent than projects in the other more "scientific" sub-programs, time and resource consuming steps like arranging for the use of such equipment or even for their rental or purchase. Finally, we included the size of the team in $\mathbf{Z}$. Larger teams can write larger grants because of greater specialisation amongst its members, or because *SIZE* proxies for other resources available to the group. We assume that conditional on the budget asked and other covariates, *SIZE* does not affect the fraction of the budget granted. In the production function of publications (vector $\mathbf{H}$) we include all exogenous variables except *DCNR* and *DUNI*. We also include *DSOUTH* in $\mathbf{H}$ to account for any disadvantages that units in Southern Italy may face.

• **Identification.** Note that we identify our parameters through some key exclusion restrictions. First, we exclude the budget asked and the size of the unit from the selection equation. As discussed above, the exclusion of budget asked is justified if there is informal communication between the units and CNR prior to the formal application process. The exclusion of the size of the unit is also a plausible restriction because given budgets and other characteristics of the units, size does not affect output. Therefore, it is rational for the CNR not to consider size in selection [18].

Second, we assume that the size of the team does not affect the fraction of budget granted. This is a plausible restriction, inasmuch as the fraction of the budget granted is hypothesised to depend upon measures of quality and transferability. Third, we identify the elasticity with respect to the budget in the publication equation by assuming that differences in institutional types, *DCNR* and *DUNI* do not directly influence the productivity of publications, whereas they do influence the fraction of budget granted. University based units and CNR units may face greater costs of writing large grants because they are subject to a variety of constraints on hiring of temporary personnel. These units also are more likely to obtain larger fractions of budget asked. Typically there are formal and informal relationships between a public funding agency like CNR and other public research institutions. This implies that CNR is better informed about the activities and reputation of these groups and their projects. Such groups may, of course, have been in a better position to "lobby" for their projects. Once other factors are controlled for, however, there is no obvious reason why the productivity of research groups should differ according to institutional type. A likelihood ratio test implies that the restriction cannot be rejected [19].

Finally, we identify the budget asked equation by assuming that the cost of writing a proposal is linear in B_A. Identification through functional form is clearly not very attractive. However, it is difficult to conceive of observable variables that would affect outcomes such as selection and publication output but would not affect budget requested. Indeed, although our exclusion restrictions are very plausible, our sample is small and our

18. As discussed earlier, DPRO7 is also excluded. However, this exclusion is not critical for identification but for preserving degrees of freedom.
19. The value of the chi-square statistic is 2.98 with two degrees of freedom, so that the exclusion restriction we impose cannot be rejected.

estimates may therefore be sensitive to the interaction between the exclusion restrictions and functional form. Accordingly, for each estimated equation (except the selection equation) we tested for robustness by estimating a linear specification as well. The results do not change much, although the log-log specification fits the data somewhat better. Table A1 in the appendix shows the reduced form estimates using levels. By comparing them to Table 3 and 4, one can see that the log specification is a better fit, but that the qualitative properties of the empirical model are not driven by the specification.

5.2. **Results: Selection, Budget Asked, Budget Granted**

Table 5 shows the results of the probit estimates. Note that the selection estimates strongly support the notion that selecton is driven by the objectives of the programme. These objectives in turn do not appear to be substantively different from those of comparable public research programmes in the US. Thus, variables correlated with scientific merit (*K*, *COLLAB*) are quantitatively and statistically significant. Likewise, "industrial transferability" increases the probability of selection. Interestingly enough, units from the South of Italy do not appear to be particularly advantaged, even though

TABLE 5

Selection_Equation (PROBIT).

Dependent variable: $I = 1$ if unit is funded by CNR

Parameter	Estimate
Const	−2.813
	(1.302)
DCNR	0.466
	(0.168)
DUNI	0.293
	(0.143)
DSOUTH	0.090
	(0.128)
TRANSF	0.378
	(0.098)
ln(COLLAB)	0.224
	(0.075)
ln(*K*)	0.398
	(0.052)
ln(NUIST)	−0.034
	(0.055)
ln(PROV_POP)	0.169
	(0.063)
ln(AGEPI)	−0.065
	(0.303)
Log Likelihood	−475.4
N.obs	797
(of which positive)	(347)

Heteroskedastic consistent standard errors in parenthesis.

providing such a preference is an explicit announced aim in many CNR programmes.

Table 6 reports the results of joint-estimation of (7) and (8). Note that the estimates are similar to the reduced form estimates. Larger research groups have lower marginal costs in applying for grants (η_N), but they gain no advantage from being part of larger institutions (η_U) or from being in more populated areas (η_P). Externalities amongst research groups in the same institution or in the same city are not pronounced. CNR acts consistently and puts no weight on NUIST or PROV_POP in the funding decision (θ_U and θ_P). Instrument development projects are more expensive to set up (η_7), and CNR grants them a larger fraction of budget (θ_7). Transferability and past publications increase the fraction of the budget granted (θ_Γ and θ_K). CNR units also obtain a larger fraction of expected budget (θ_{CNR}). Unobserved characteristics of the units matter as well, although the statistical significance of σ_{ik} is not high. The estimated value of λ suggests that non research (indirect) costs are of the order of 32% of the expected budget conditional upon selection. The coefficients of *DSOUTH*, η_S and θ_S, are

TABLE 6

GLS Estimation of* Budget Requested *and* Budget Granted *equations (Equations (7) and (8)).

Dependent variables: ln(B_A) and ln(B_G). No of obs. $= 347$, $I = 1$.
Log of likelihood function $= -522.6$.

Parameters	Estimates	Parameters	Estimates
Const (η_0)	– 0.051	Const (θ_0)	1.769
	(0.887)		(1.252)
DCNR (n_{CNR})	– 0.210	DCNR (θ_{CNR})	0.322
	(0.115)		(0.146)
DUNI (η_{UNI})	– 0.004	DUNI (θ_{UNI})	0.097
	(0.116)		(0.132)
DSOUTH (η_S)	0.358	DSOUTH (θ_S)	– 0.185
	(0.103)		(0.088)
DPRO7 (η_7)	– 0.272	DPRO7 (θ_7)	0.369
	(0.156)		(0.125)
TRANSF (η_T)	0.014	TRANSF (θ_T)	0.130
	(0.073)		(0.107)
ln(K) (η_K)	0.013	ln(K) (θ_K)	0.136
	(0.038)		(0.107)
ln(COLLAB) (η_C)	0.076	ln(COLLAB) (θ_C)	0.049
	(0.053)		(0.078)
ln(NUIST) (η_U)	0.023	ln(NUIST) (θ_U)	– 0.026
	(0.045)		(0.041)
ln(PROV_POP) (η_P)	– 0.033	ln(PROV_POP) (θ_P)	0.065
	(0.052)		(0.061)
ln(AGEPI) (η_A)	0.126	LN(AGEPI) (θ_A)	– 0.115
	(0.218)		(0.191)
LN(SIZE) (η_N)	0.366	σ_{1K}	0.315
	(0.120)		(0.389)
ln(A) (λ)	0.319		
	(0.151)		

Heteroskedastic consistent standard errors in parenthesis.

respectively 0.36 and –0.19, and they are well measured. These estimates clarify the reasons for the pattern revealed by the reduced form estimates in Table 3. However, the positive sign of η_S suggests that units in the South prepare larger projects for one reason. All else constant, they have a lower cost of writing larger grants.

5.3. **Results: Production Function of Publications**

We experimented with different specifications for $\alpha(K)$ in estimating (9). In the end, we settled on a logistic specification, $\alpha(K) = \frac{\alpha_0}{1+e^{-\beta(1+K)}}$, where α_0 and β are parameters to be estimated. The logistic fit the data better than either a constant elasticity $\alpha(K) = \alpha_0$, or a linear specification with interaction, $\alpha(K) = \alpha_0 + \beta \ln(1+K)$. In each case, the point estimates of the elasticity are almost the same as those reported here [20]. The logistic specification also has the appealing property that it allows the elasticity to vary but within bounds, namely that with $\beta > 0$, α_0 is an upper bound of the elasticity of budget.

For the logistic specification reported in Table 7, the estimated value of α_0 of 1.01 implies that the elasticity of research budgets has an upper bound of about 1. More interestingly, the statistical significance of β indicates that the elasticity of research budget does increase with the stock of past publications K of the PI [21]. In turn, this means that the distribution of the elasticity of research budgets ought to mimic the skewed distribution of past performance. Indeed, as Figure 4 shows, the distribution of our logistic $\alpha(K)$ evaluated at the estimated parameters, α_0 and β, is skewed towards the left. Although the estimated $\alpha(K)$ for our sample ranges between 0.51 and 1.01, its value at the median K of the population of 347 selected units is 0.58. Moreover, about 90% of these units have an output elasticity with respect to research budget that lies in the range below 0.8 [22].

This estimated elasticity is consistent with a characterisation of the scientific enterprise as a "star" system. While the productivity of the large majority of our research groups falls within a limited range around the median of the distribution, a small fraction of the research groups displays higher productivities. The skewed distribution of $\alpha(K)$ suggests that the marginal product of total budget in a given research programme may vary substantially with changes in the resource allocations among research units. Thus, as we show in the next section, changes in resource allocation schemes can change aggregate output.

20. Table A2 in the appendix reports our estimated production function of publications using the constant and log-linear elasticities.
21. Note that estimate of β is statistically significant even though $\ln(K)$ appears as a separate regressor.
22. We also estimated (9) by using the actual levels of research budgets instead of instrumenting for it. This amounted to using actual values of $\ln B$ instead of the expression for b_e^* in (9). We found that in this case α_0 is 0.67 and β is 0.007, and they are both statistically significant. Our model then predicts higher elasticities of research budget (and a distribution with greater spread) than if one did not instrument for research grants.

TABLE 7

Publication Equation – Logistic $\alpha\,(K)$ (Equation (9) – TOBIT).

Dependent variable: $\ln(PUB)$ N obs. $= 347$, for $I = 1$.

	$\alpha\,(K) = \alpha_0/(1 + e^{-\beta(1+K)})$
const	1.841
	(2.393)
α_0	1.013
	(0.255)
β	0.009
	(0.004)
DPRO7 (γ_7)	– 0.990
	(0.231)
DSOUTH (γ_S)	– 0.564
	(0.179)
TRANSF (γ_T)	– 0.198
	(0.180)
ln(SIZE) (γ_N)	– 0.002
	(0.140)
ln(*K*) (γ_K)	0.082
	(0.221)
ln(*COLLAB*) (γ_C)	0.257
	(0.122)
ln(NUIST) (γ_U)	– 0.001
	(0.057)
ln(PROV_POP) (γ_P)	– 0.057
	(0.098)
ln(AGEPI) (γ_A)	– 0.515
	(0.380)
σ_{IP}	0.146
	(0.621)
LogLik	– 515.75

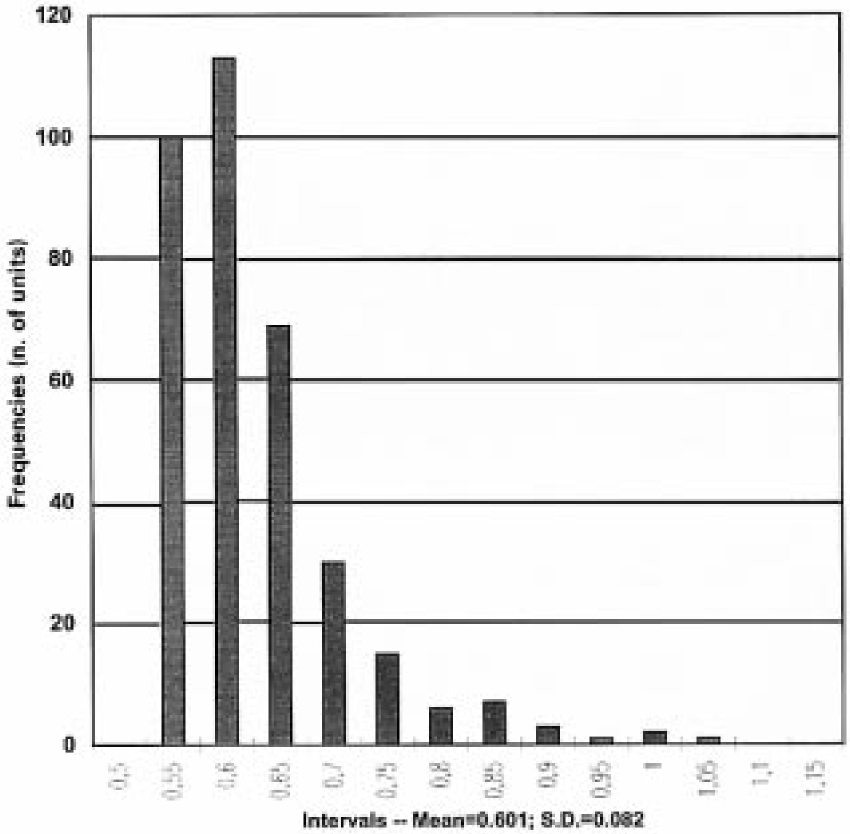
Heteroskedastic consistent standard errors in parenthesis.

As far as the other parameters of the production function are concerned, Table 7 shows that bio-instrumentation projects and southern units have lower output (*DPRO7* and *DSOUTH*), whereas collaborations with foreign institutions increases output (*COLLAB*). Consistent with our identifying assumptions, the size of the team does not affect productivity. To the extent that *SIZE* proxies for other resources available to the units, this effect manifests itself largely through the size of the project, budget asked, rather than directly affecting research output. This results is also consistent with our earlier assumption that these are "small" projects. We also found that externalities within the same institution or from being part of large metropolitan areas do not play a significant role (*NUIST* and *PROV_POP*).

It was noted, at the end of Section 4.1, that the marginal product of the budgetary resources provided by comparatively small grants from the CNR would tend to be raised when the research unit in question was able to maintain more "infrastructure" by spreading its costs over projects funded from other sources. It should be clear that *SIZE* alone would not serve as a proxy for the capacity to obtain the margins of funding above direct

FIGURE 4

Alpha(K) (Frequency distribution, 347 observations for I=1)



research costs that would be required for such purposes; the same factors that affect the probability of receiving a grant, and the magnitude of the project budget, would most likely be just as relevant in obtaining funding from other sources. Among those factors we have found the "knowledge capital" of the unit, K, to be significant in the case of the CNR project. By symmetry, this would imply that some part of the estimated direct effect of K on the unit's productivity in using CNR programme-provided resources may, in some part, be reflecting the positive role of past research performance in furnishing the unit with a better physical and human "infrastructure". Of course, without the necessary data on other sources of research support held by the units, this must remain a conjectural interpretation and the hypothesised "infrastructure effect" of the unit's accumulated knowledge capital cannot be identified separately within the direct production effect of K.

Finally, although their statistical significance is not high, the point estimates of *TRANSF* and *AGEPI* are negative and large in magnitude. The negative impact of the age of the PI may reflect not just life cycle effects, but also selection effects at the tails of the age distribution. More reputed and productive researchers are likely to become PIs at younger ages. Thus, compared with the median age, at lower ages, the expected quality of PIs is likely to be higher. No further interpretation can be given to this coefficient in a cross section data set such as ours. The point estimate coefficient of *TRANSF* suggests that, other things being equal, there is a trade-off between industrial transferability and publications: projects aimed at industrial applications produce about 20% fewer publications.

6 Optimal Resource Allocation and Returns to Past Performance

6.1. Optimal Allocation of Resources and the Aggregate Productivity of Research Budgets

As has already been noticed, the skewed distribution of $\alpha(K)$ suggests that the average productivity of research budgets at the aggregate level may vary considerably with the distribution of research grant allocations, even when the total size of the research budget does not change. Given our estimated parameters, in this section we ask which allocation of resources would equalise the selected units' respective marginal productivities – reckoned in terms of expected quality adjusted-publications [23].

It should be understood at the outset that this is a very short-run allocative criterion, and that qualification must be borne in mind when interpreting our references to the magnitude of the actual CNR allocations' departure from "optimality". Maximising conditional expected aggregate output in this way would not take into consideration the effects upon the research units' subsequent capabilities, their future access to funding (from all sources) and the expected future trajectory of their productivity. Nor does it allow for generalised training effects, and possible long-run spillovers that depend upon the presence of groups pursuing a diversity of approaches, including approaches that have yet to show payoffs in terms of past publication performance measures. It is by no means obvious that the dynamic effects would run in the same direction as the first round consequences of a reallocation that equalised marginal (expected) outputs across the population of funded units; shifting funding towards the presently most productive units might have deleterious effects upon the development of others whose future productivity potential is far higher – even were it to be suitably discounted. Nevertheless, obtaining a sense of the magnitude of the short-run reallocation effects remains an important starting point in any attempt at a more complete dynamic analysis.

23. We perform this experiment only for the units that were funded by the CNR. This amounts to taking the actual selection decision of CNR as given, and looking for the allocation of resources that maximise the expected total publications of the programme.

To "re-allocate" resources amongst our selected units we obtained an estimate of their marginal product of publications, MP. Since $\ln E(B_G) = E(\ln B_G) + \sigma_b^2/2 - u_b$, one can use the following expression as an estimate of the marginal product of budget

(10)
$$\hat{M}P(B, K) = \alpha(K) \exp(E(\ln PUB|I=1, B_G) \equiv \alpha(K)\,\Psi\, B_G^{\alpha(K)-1}$$

where $\Psi = \exp\{\alpha \cdot (K)[\ln(1-\lambda) + \frac{\sigma_b^2}{2}] + \gamma \cdot h + \sigma_{IP}\frac{\phi}{\Phi}\}$

Note that the marginal product of publications depends on all characteristics of the units, not simply B_G and $\alpha(K)$. Hence, even though $\alpha(K)$ increases in K, the marginal product need not.

We proceed by ranking our selected units according to K, and comparing the units located at the mid point of each quartile. The total budget received by those units was then reallocated so that their estimated marginal products were equalised. Table 8 shows that to maximise aggregate publications in this way, 85% of total budget should be allocated to the top quartile. In the resulting, short-run "re-optimised" allocation, a very large share of the total available budget would be given to a small percentage of highly productive teams. Also, as shown by Table 8, the efficient allocation to the top unit is roughly of the same magnitude of the amount of budget asked by that unit (496 *vs.* 477 million Lire). These "optimal" short-run allocations are thus of reasonable size, and do not entail grant awards in excess of the (self-assessed) "absorptive capacities" of the units [24].

We also compared the output produced by the "re-optimized" allocation with the benchmark case of equal allocations. Given the total budget actually allocated to our four units, Table 8 reports what their expected total publications would have been, had they obtained identical shares of that those funds. Rather surprisingly, it turns out that this "egalitarian" distribution of funding would yield the same expected publications output as that obtained with the actual CNR allocations; in effect it "remedies" the disproportionately large budget that was allocated to a relatively low marginal product unit. But, from Table 8 one may also see that the "re-optimized" allocation produces about 19% more total publications (in quality-adjusted units) than both the actual and "egalitarian" allocations.

Needless to say, the foregoing calculations are intended merely to illustrate what our econometric results imply, and are not presented in support of any policy prescriptions. As the introductory caveats in this section have indicated, maximising the existing research community's output of publications is unlikely to be the sole objective of a sensible national science

24. We experimented with many different observations around the mid-point of each quartile, and the results discussed here are robust.

policy [25]. What our analysis shows is that by making use of data generated in the management of the public funding process, it is possible to quantify the trade-off between various goals in terms of foregone production of scientific publications.

6.2. **Total Returns to Past Performance**

Our second experiment is to compute the total returns to past performance K. As discussed in the introduction, the indirect effect of past performance through budget can amplify differences in scientific productivity. The magnitude of this effect may then be suggestive of the extent to which the observed skewed distribution of publications in the scientific enterprise is influenced by the institutional mechanisms of resource allocation in this sector.

The elasticity of output with respect to past performance is

$$\rho_k \equiv \frac{\partial E(\ln PUB)}{\partial \ln K} = \Phi \left\{ \frac{\partial \alpha(K)}{\partial \ln K} [\ln(1-\lambda) + b^* + \delta \mu_b] + \gamma_k \right\} + \phi\, \pi_k\, p + \Phi \alpha(K) \frac{\partial b^*}{\partial \ln K} \tag{11}$$

where π_k is the elasticity of K in the selection equation, and $b^* = \ln E(B_G | I = 1, \varepsilon_k)$. Since we do not observe b^* (because we do not observe ε_k), and we do not observe μ_b and ε_p, we cannot compute ρ_k directly. However, we can estimate it by taking the expectation of (11) over ε_k, μ_b, and ε_p, given $I = 1$. Using b_e^* to denote the expected value of the budget conditional upon selection (but not upon ε_k), the estimated elasticity is

$$\begin{aligned} \hat{\rho}_k = {} & \Phi \left\{ \frac{\partial \alpha(K)}{\partial k} [\ln(1-\lambda) + b_e^*] + \gamma_k \right\} \\ & + \phi \pi_k \left\{ \alpha(K) [\ln(1-\lambda) + b_e^*] + \gamma\, h + \sigma_{IP} \frac{\phi}{\Phi} \right\} \\ & + \Phi \alpha(K) \frac{\theta_k}{1-\lambda} \end{aligned} \tag{11a}$$

In (11a), the first term measures the direct effect of past performance given the budget, the second term measure the indirect effect through increases in

25. One can list a number of reasons. Papers produced by different PIs, even though of the same quality, may not be perfect substitutes. One may wish to encourage certain fields rather than others. In our particular programme we saw that the agency also wanted to encourage industrial transferability of scientific research, and our estimates of the production function of publications suggest that industrial transferability has a negative impact on publication output. Second, one may wish to have a diversified portfolio to minimize risk. Third, one may wish to encourage young talent, even at the cost of a short run reduction in output. This could be either if there is learning-by-doing in research, or if by funding young scientists, the public agency can get a signal about their true productivity, and hence, can make more informed funding decisions in the future. (See for instance ARORA and GAMBARDELLA [1997].)

TABLE 8

Effects of Alternative Reallocations of Research Budget Resources (347 selected units ranked by past publications* K*, mid-points of each quartile: positions 43, 130, 207, 304).

Marginal product MP (*)	$\alpha\,(K)$	Actual allocation $B_G\,(+)$	Efficient allocation $B'_G\,(+)$	Marginal product of efficient allocation $MP\,(B'_G)$	Expected #pubs. using actual allocation $EPUB\,(B_G)$	Expected# pubs. Using efficient allocation $EPUB\,(B'_G)$	Equal share allocation $B''_G\,(+)$	Expected #pubs. using equal share allocation $EP\,(B''_G)$	Past publ. $K\,(\hat{})$	Budget asked $B_A\,(+)$
0.109	0.677	156	496	0.0875	29.3	64.2	145.75	28.0	74.0	477
0.006	0.604	100	23.5	0.0875	8.2	3.4	145.75	10.3	41.0	620
0.189	0.561	120	38.5	0.0875	11.4	6.0	145.75	12.7	22.8	126.2
0.049	0.531	207	25	0.0875	12.6	4.1	145.75	10.5	10.3	562
		Total: 583	Total: 583		Total: 61.5	Total: 77.7	Total: 583	Total: 61.5		

(*) Computed as $\alpha(K) \cdot PUB/B_G$ using actual values of PUB and B_G and estimated $\alpha\,(K)$, logistic.

(+) Millions of Italian LIre.

($\hat{}$) Quality-adjusted number.

probability of selection (given expected budget conditional upon selection) and the third term measures the indirect effect through the increase in expected budget conditional upon selection.

Figure 5a shows the distribution of $\hat{\rho}_k$ for our sample of selected units. Table 9 presents descriptive statistics of $\hat{\rho}_k$, and of its three components. The average value of the total elasticity of publication output with respect to past publication performance is 0.64. This value is determined largely by the direct effect $\hat{\rho}_{k1}$, whose sample mean is 0.27, and the indirect effect through selection $\hat{\rho}_{k2}$, whose sample average is 0.31. The indirect effect through budget given selection, $\hat{\rho}_{k3}$, is smaller; its average value in the sample is 0.07.

TABLE 9

Descriptive Statistics for Elasticity with Respect to Past Performance (347 observations for $I = 1$).

	Mean	Std Dev	Minimum	Maximum
$\hat{\rho}_{k1}$ (direct effect)	0.272	0.223	0.011	0.993
$\hat{\rho}_{k2}$ (indirect effect selecton)	0.305	0.096	0.067	0.549
$\hat{\rho}_{k3}$ (indirect effect budget)	0.066	0.030	0.011	0.161
$\hat{\rho}_k$ (total effect)	0.643	0.304	0.109	1.466

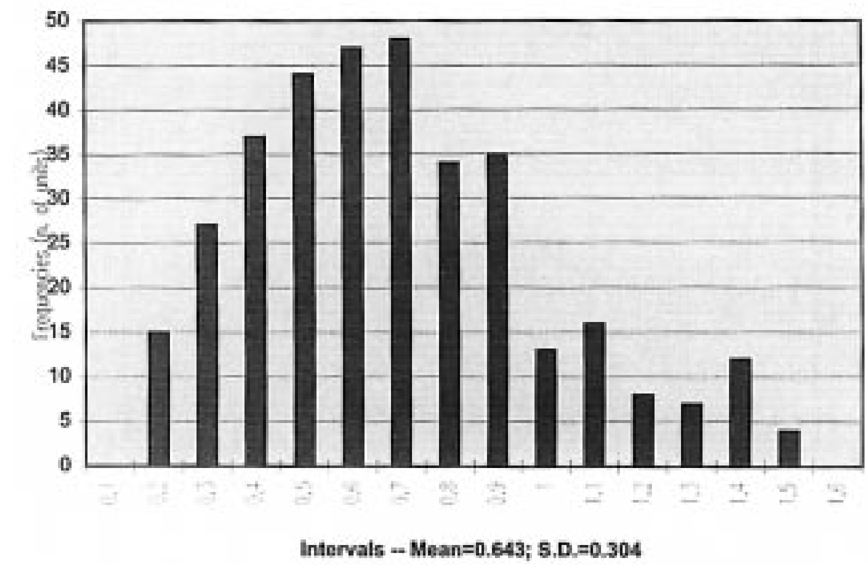
This suggests that there are important reinforcing effects of past performance in the scientific sector, operating primarily by increasing the chances of selection. The relative insensitivity of the expected budget conditional upon selection may reflect a measurement issue – it is difficult to find variables that affect the budget conditional upon selection, but not selection itself. Even so, we believe that our results point to the importance of the selection process, where reputation and quality of the research unit appear to play a very prominent role.

Although our cross-sectional data do not allow us to take dynamic considerations into account, the indirect effects seem to be serious enough to be capable of generating increasing returns to past performance. Consider the following suggestive piece of evidence: the total elasticity of publication output with respect to past publications in our sample is greater than unity for 47 of the 347 selected units, and, as shown by Figure 5b, these values are associated with higher K. Since the estimated direct effect is well below one, the institutional mechanisms for resource allocation in the scientific enterprise may be critical in creating the appearance of increasing research returns to the accumulation of knowledge capital at the level of the individual research unit. The marked skewness in the distribution of publication outputs of individual scientists, and of research groups, thus may be due in large part to the way in which academic-style scientific activities are funded. In turn, this points to the importance of controlling for funding levels and for selection in estimating the "competences" of research organizations and the research "abilities" of individual scientists. Failure to do so will produce greatly exaggerated estimates of the dispersion in the underlying distribution of innate research capabilities.

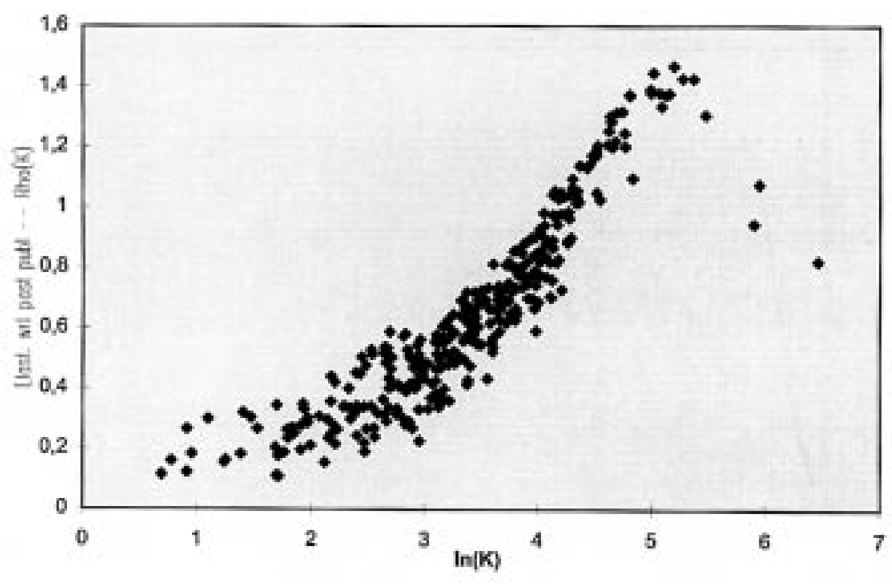
FIGURE 5

Distribution of Rho(K), and plot on past publ. (347 observations for I = 1).

a) Distribution of Rho(K).



b) Plot of Rho(K) on past publications.



7 Conclusions

Empirical studies of resource allocation in the Republic of Science are still in their infancy. While economists have ignored it, quantitative sociologists of science have largely been concerned with the determinants of the scientific productivity of individuals and their career paths. In the natural sciences and engineering, research is increasingly a group activity, one whose continuity depends upon success in mobilizing not just human resources but costly instruments, materials, and facilities as well.

This paper has examined the determinants of the publication performance of publicly funded scientific research groups. We modeled the process of resource allocation and production of scientific research output. Our analysis allows for unobserved differences across research units and corrects for “funding selection biases”, which have been found to be quantitatively quite important.

We estimate that for a large fraction of the research groups in our sample the elasticity of quality-adjusted publications with respect to the budget given is about 0.6. However, for a small fraction of researchers, with high values of quality-adjusted past publications, this elasticity is higher, and it approaches unity. This implies that the aggregate publication output may vary with the distribution of research grants.

This relatively high responsiveness of output to research budgets points to an important indirect route through which past performance influences future performance. Superior performance on the part of the group’s leader in the past increase the probability of the research proposal being selected. This indirect effect turned out to be substantial. While the combined elasticity of publication output with respect to past publications is on average 0.64, the indirect component accounts for almost 60% of this figure. Moreover, the elasticity varies across research units; for a small fraction of our applicant units, it is greater than 1.

The nature of the analysis carried out here is not normative, either in its intent or its conclusions. We most certainly would not advance it as suggesting that the institutional mechanisms for funding scientific research are inefficient, or that using past performance to estimate productivity is incorrect. Neither do we claim that the characteristic skewness of the distribution of scientific publications is socially undesirable. Absent systematic micro-level time series data on inputs (funding levels), and research performance in various scientific fields, questions of that kind simply cannot be answered. Our objective in this paper has been to demonstrate the possibilities of quantitatively describing the relationships that proximately govern productivity in scientific research groups. In doing so, we have made a start towards fully uncovering how competence (both innate and acquired), reputation, and the institutional mechanisms for the funding of academic research interact in the production of scientific knowledge.

APPENDIX

Results for Alternate Specifications

TABLE A1

Reduced Form Equations: OLS Budget Granted and Publication

	B_G	B_G $(I=1)$	*PUB*	*PUB* $(I=1)$	*PUB*
Const	9.90	93.14	– 0.14	15.42	– 0.80
	(14.66)	(21.33)	(3.56)	(7.50)	(3.32)
DPRO7	27.39	54.79	– 1.47	– 6.21	– 3.37
	(10.17)	(16.54)	(0.80)	(1.54)	(1.01)
DCNR	30.55	26.18	– 1.32	– 3.84	– 3.34
	(10.90)	(13.44)	(1.42)	(2.94)	(1.39)
DUNI	9.72	– 2.16	0.18	– 0.68	– 0.46
	(5.717)	(10.34)	(1.44)	(3.52)	(1.33)
DSOUTH	– 0.20	– 8.70	– 2.65	– 6.64	– 2.64
	(5.65)	(8.96)	(0.90)	(1.87)	(0.87)
TRANSF	17.15	6.12	0.58	– 0.46	– 0.55
	(4.65)	(6.30)	(0.97)	(1.74)	(0.93)
SIZE	1.32	1.07	0.05	0.00	– 0.03
	(0.34)	(0.40)	(0.07)	(0..09)	(0.06)
K	0.32	0.08	0.23	0.25	0.21
	(0.08)	(0.06)	(0.04)	(0.04)	(0.04)
COLLAB	4.38	1.27	0.70	0.66	0.41
	(1.32)	(1.43)	(0.29)	(0.42)	(0.28)
NUIST	0.01	0.35	– 0.06	– 0.09	– 0.06
	(0.18)	(0.28)	(0.03)	(0.07)	(0.03)
PROV_POP*	4.29	– 0.86	0.55	0.06	0.25
	(2.45)	(4.21)	(0.44)	(0.95)	(0.41)
AGEPI	– 0.46	– 0.49	0.00	– 0.12	0.03
	(0.26)	(0.40)	(0.05)	(0.12)	(0.05)
B_G					0.07
					(0.01)
No of obs	797	347	797	347	7.97
Adj. R^2	0.14	0.13	0.35	0.42	0.42

(*) Measured in hundreds of thousands.
Heteroskedastic consistent standard errors in parenthesis.

TABLE A2

Publication equation – Constant and logarithmic $\alpha(K)$ specification (TOBIT).

Dependent variable: ln(*PUB*). No of obs. $= 347$, for $I = 1$.

	$\alpha(K) = \alpha_0$	$\alpha(K) = \alpha_0 + \beta \ln(K)$
const	– 0.593 (5.879)	5.816 (11.10)
α_0	0.611 (0.988)	– 0.690 (2.101)
β	–	0.293 (0.390)
DPRO7 (γ_7)	– 1.014 (0.482)	– 0.812 (0.583)
DSOUTH (γ_S)	– 0.504 (0.234)	– 0.578 (0.257)
TRANSF (γ_T)	– 0.047 (0.182)	– 0.120 (0.204)
ln(SIZE) (γ_N)	0.20 (0.214)	0.062 (0.223)
$\ln(K)(\gamma_K)$	0.563 (0.176)	– 0.766 (1.798)
ln(COLLAB) (γ_C)	0.301 (0.131)	0.266 (0.143)
ln(NUIST) (γ_U)	– 0.002 (0.062)	– 0.014 (0.064)
ln(PROV_POP) (γ_P)	– 0.029 (0.099)	– 0.049 (0.105)
ln(AGEPI) (γ_A)	– 0.517 (0.400)	– 0.558 (0.413)
σ_{IP}	0.671 (0.921)	0.284 (1.096)
Log of Likelihood	– 520.11	– 519.81

Heteroskedastic consistent standard errors in parenthesis.

• References

ALLISON, P., PRICE, SOLLA, D. J. de, GRIFFITH, B., MORAVCSIK, M., STEWART, J. (1976). – "Lotka's Law: A Problem in its Interpretation and Application", *Social Studies of Sciences,* Vol. 6, pp. 269-276.

ALLISON, P., LONG, S., KRAUSE, T. (1982). – "Cumulative Advantage and Inequality in Science", *American Sociological Review,* Vol. 47(5), pp. 615-625.

ARORA, A., GAMBARDELLA, A. (1997). – "Public Policy Towards Science: Picking Stars or Spreading the Wealth?", *Revue d'Économie Industrielle,* N. 79, pp. 63-75.

DASGUPTA, P., DAVID, P. A. (1987). – "Information Disclosure and the Economics of Science and Technology", in FEIWEL, G. (ed.), *Arrow and the Ascent of Modern Economic Theory,* New York University Press, New York, pp. 519-542.

DASGUPTA, P., DAVID, P. A. (1994). – "Towards a New Economics of Science", *Research Policy,* Vol. 23, pp. 487-521.

DAVID, P. A. (1993). – "Knowledge, Property and the System Dynamics of Technological Change", in *Proceedings of the World Bank Annual Conference on Development Economics: 1992,* Summers, L. and Shah, S. (ed.), Washington DC, March.

DAVID, P. A. (1994). – "Positive Feedbacks and Research Productivity in Science: Reopening Another Black Box", in GRANSTRAND, O. (ed.) *Economics of Technology,* North-Holland, Amsterdam and London.

European Report on Science and Technology Indicators (The) (1994), European Commission, DG XII, EUR 15897 EN, Luxembourg.

JAFFE, A. (1989). – "Real Effects of Academic Research", *American Economic Review,* Vol. 79, (5), pp. 957-970.

LEVIN, S., STEPHAN, P. (1991). – "Research Productivity over the Life Cycle: Evidence for Academic Scientists", *American Economic Review,* Vol. 81, (1), pp. 114-132.

LOKA, A. (1926). – "The Frequency Distribution of Scientific Productivity", *Journal of the Washington Academy of Sciences,* Vol. 16, (12), pp. 317-323.

MANSFIELD, E. (1991). – "Academic Research and Industrial Innovation", *Research Policy,* Vol. 20, (1), pp. 1-12.

MERTON, R. (1968). – "The Matthew Effect in Science", *Science*, Vol. 159, (3810), pp. 56-63.

NARIN, F., OLIVASTRO, D. (1992). – "Status Report - Linkage between Technology and Science", *Research Policy,* Vol. 21, (3), pp. 237-249.

NELSON, R. (1986). – "Institutions Supporting Technical Advance in Industry", *American Economic Review Proceedings,* Vol. 76, (2), pp. 186-189.

OECD (1994). – *Main Science and Technology Indicators,* OECD, Paris.

PRICE, D. J. de Solla (1963). – *Little Science, Big Science,* Columbia University Press, New York.

PRICE, D. J. de Solla (1976). – "A General Theory of Bibliometric and Other Cumulative Advantage Processes", *Journal of the American Society for Information Sciences,* Vol. 27, (5/6), pp. 292-306.

STEPHAN, P., LEVIN, S. (1992). – *Striking the Mother Lode in Science: The Importance of Age, Place and Time,* Oxford University Press, New York.

STEPHAN, P. (1996). – "The Economics of Science", *Journal of Economic Literature,* Vol. XXXIV, pp. 1199-1235.