

REPUTATION AND COMPETENCE IN PUBLICLY FUNDED SCIENCE:{PRIVATE }
ESTIMATING THE EFFECTS ON RESEARCH GROUP PRODUCTIVITY

Ashish Arora
Heinz School, Carnegie Mellon University,
Pittsburgh PA

Paul David
All Souls College, Oxford University, and
Stanford University, Stanford CA

Alfonso Gambardella
Istituto di Studi Aziendali, University of Urbino,
Urbino, Italy

First draft: July 1994
This draft: March 1996

We thank the Director of the CNR Programme "Biotechnology and Bio-instrumentation", Prof. Antonio De Flora, for his co-operation. An earlier version of this paper was presented in seminars at Carnegie Mellon, LSE, NBER, and the University of Pittsburgh. We gratefully acknowledge the helpful comments of Zvi Griliches and Bronwyn Hall. We also acknowledge financial support from the University of Urbino (MURST 60%), the EC Human Capital Mobility Programme (Contract N.ERBCHRXCT920002), and MERIT, University of Limburg. Fabrizio Cesaroni, Marco Cioppi, and Aldo Geuna provided excellent research assistance.

Abstract

This paper studies an unusually comprehensive data-set of the universe of research groups that applied to an Italian 1989-93 research programme in biotechnology and bio-instrumentation. We develop a model to test whether past publications influence the expected grants obtained by the units in this programme, and hence their publication output. We find that this effect is important, and suggest that institutional features of resource allocation in public research may have serious implications for long term divergence in research performance. We also find that to maximise the total expected publications of the programme about 85% of its budget should be allocated to the top quartile of the selected units (about 10% of applicants) ranked according to past publications. Our estimated elasticity of publications on budget is 0.6, and we find evidence of a trade-off between publications and "industrial transferability": all else held constant, projects conducted with industrial applications in mind produce about 20% fewer publications.

1. managers. median age of in US economics
(See David scientific industry. literature on
INTR [1992] for a papers cited R&D as a
ODU recent by US patents Given this problem of
CTIO review.) has declined. recognition, it resource
N Narin and Nelson is surprising allocation by
Today the Olivastro [1986], Jaffe how little firms. The
complex and [1992] found [1989], and neglect of
multi- a significant Mansfield been paid by science is
dimensional upward trend [1991] have economists to even more
links between in the found that the remarkable
technological citations of university determinants when one
progress and scientific research, of resource notices the
basic research publications university allocation and substantial
are being in US patents patents, and productivity resources
recognised by during the important differences in allocated
economists as 1980s. academic scientific through the
well as by Moreover, discoveries research. By institutional
science they found have a sizable contrast there mechanisms
administrators that in the impact on is an that are found
and business 1980s the R&D payoffs extensive in the

Republic of Science. In the US and Western Europe, 40-45 percent of the total R&D expenditure is financed by the public sector (OECD [1994]), and of that, a major fraction is allocated through programmes in which public agencies offer research contracts to competing applicants.¹

¹ See Dasgupta and

This paper develops a structural model of the process by which research units apply for, and receive funds.

We use data on about 800 research units _____ David [1987] for a discussion of this disparity and its sources; David [1994], and Dasgupta and David [1994] examine the efficiency implications of the institutional structures and reward systems characterising academic science.

that applied to a 1989-1993 research programme, Biotechnology and Bio-instrumentation, instituted by the Italian agency for public research -- Centro Nazionale delle Ricerche (CNR), the Italian equivalent of the NSF. We study the effects of past performance (measured by quality-adjusted past publications) on the quality

adjusted publication output of the units in this programme. In addition to a direct effect on the productivity of research effort, there are two indirect effects. First, units with better past records are more likely to be successful in getting research grants from CNR. Second, knowing this, units with a better track record are

more likely to invest in applying for larger grants. Both effects imply higher expected research budgets for these units, which increases (stochastically) their publication output. The reason why these effects deserve attention is that institutional features of the system which give rise to what Robert

K. Merton [1968] labelled the "Matthew Effect"² may underlie phenomena widely observed in studies of scientific

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 25:29.)

productivity: that the distributions of publications among cohorts of scientists in a given field are extremely skewed. In other words, a small fraction of the individuals account 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 are a part of the reason, the institutional resource allocation mechanisms may contribute significantly as well -- that is, productive disparities also reflect the outcome of stochastic processes which cumulate advantage and amplify differences.³

³ Allison et al. [1982] surveys the sociological literature in cumulative advantage

In this paper we also depart from the view of science as the pursuit of solitary researchers linked in "invisible colleges"; today science is a collaborative enterprise in very visible processes. For economists' view on this subject, see the discussions in Arora and Gambardella [1992], David [1994], and Stephan and Levin [1994: 29-30].

teams organised around expensive physical facilities and instruments. Even the recent careful work of Levin and Stephan [1991], and Stephan and Levin [1992] continues the traditional focus on the life cycle productivity of the individual academic scientist. They seek to estimate the effects of ageing through

analysis of panel data, using fixed effects type procedures to control for unobserved differences. However, their analysis does not allow for the effects of other time varying covariates, most notably research funding, as well as characteristics of research groups. Furthermore, although the relationship between the demographic composition of the scientific community and the total research productivity is interesting, the policy relevance is less easy to assess. Policy options for manipulating the demographic composition, essentially those affecting entry into the profession, are slow and costly. By contrast, policy instruments targeting resource allocation and reward are likely to impinge more immediately on research productivity. Put differently, one would like to know how shifting the marginal research dollar, across groups of different characteristics, would affect total research productivity. Our data allow us to match a given research budget with the specific publication output arising from it. Moreover, Biotechnology and Bio-instrumentation was the first major public research programme in biotechnology in Italy. Practically all Italian research groups in molecular biology and genetic engineering applied to it, and we obtained information on both the characteristics of the units that were selected for funding, and of those that were rejected. Thus, not only can we relate inputs to scientific output, but we can also correct for selection effects, and we do this by using a sample that is close to the universe of the national scientific community in these fields. Because of the unusual

features of these data, we chose to start our paper by describing our data set, and by showing reduced form regressions linking budget granted and publication output to several covariates (section 2). We use these correlations to motivate our structural model in section 3. Section 4 presents our estimates of the model. Section 5 computes the

direct and indirect elasticities of past performance, and simulates the allocation of resources that would maximise total expected publications of the programme. Section 6 concludes.

2.

DAT
A
AND
PREL
IMIN
ARY
REG
RESS
IONS

2.1 Data Biotechnology and Bio-Instrumentation (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 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. The majority were university laboratories (about 62%); about 15% were CNR laboratories,

and the rest were other non-profit research institutions such as foundations, hospital research labs, and a few 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 show some stability over time and across grants.

They were not formed to carry out just this project.

Of the original 858 units, CNR selected 360 for funding.

Because of missing data, our final sample is composed of 797 units, of which 347 were selected for funding.

We collected most of our data from the application

forms. The latter contained information such as the budget requested for each of the five years of the programme, the geographical location of the unit, the type of unit (i.e. whether university lab, CNR lab, etc.), the name and age of the principal investigator (PI), the number of people employed for

the project, the total membership of each member of the team, the number of research collaborations, the number of collaborations (and names) with potential users of the results produced by the project. The units had to list (separately) all publications of the PI and of other members of the unit in the previous five

years (1983-1987).

An explicit goal of B&B was to encourage industrial "transferability".

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. Many CNR programmes have policy

objectives, like encouraging research in the less advanced regions of the country, and particularly in the Italian "Mezzogiorno". We then constructed a dummy variable, *DSOUTH*, which takes the value 1 for units located in the South. Because of the peculiar features of sub-programme 7, we created a dummy

DPRO7 for the units in that sub-programme. We also created dummies *DCNR* and *DUNI* for CNR and university labs. Finally, we defined *AGEPI* to be the age of the PI, *NUIST* to be the number of units from the same institution (e.g. University of Rome, CNR of Naples) that applied to the programme, and *POPPR*

to be the population of the province wherein the unit is located. The latter two variables account for the size of the institution and the size of the metropolitan area to which the unit belongs. CNR supplied us with the list of units that were selected, and the total budget granted to each of them 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 acknowledge d the financial support of this programme. As B&B really 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). According, we defined our budget requested and budget granted variables as the amounts pertaining to the first three years of the programme. Since annual budgets tended to constant over

time, this involved little more than simply scaling the variables by three fifths.

We weighted the publication to account for quality. We used the 1987 impact factor (*IF*) of scientific journals computed by the Science Citation Index (SCI).⁴ An

⁴ The *IF* is the ratio between the number of citations of the journal by other journals, and its number of citations to

alternative possibility was to use the number of citations of the papers. However,

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. In addition to journal articles, articles in books and working papers were also listed. These have a nominal *IF* of 0.

because the papers produced in the programme were all relatively recent, citations measures may be biased. We chose instead to weight by the quality of journals. To be precise, $P_i = \sum_j (s_{ij} \omega_j)$, where the s_{ij} is the number of publications of the i th unit in the j th journal, and ω_j is a linear function of

the impact factor of the j th journal. We experimented with a variety of specifications for ω_j . Here we report results using $\omega_j = 0.5 + \frac{IF_{ij}}{1 + IF_{ij}}$.⁵ We experimented with $1 + IF_{ij}$ 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,

As a measure of past performance of the group, we used the 1983-1987 publications of the PI listed in the application form. These were adjusted for quality in the same way as the publication output. Even though these publications are "older", and the citations more complete, we the results were substantially unchanged.

chose to be consistent with the measure used for publication output. We also obtained the 1983-1987 publications of the PI from an electronic search in the SCI. The latter appeared to be less complete. At any rate, the results obtained by using this measure were similar to the ones obtained using the publications

listed in the application forms. We use the research collaborations (*RCF*) of the unit with foreign non-profit institutions as another measure of quality. In fact, one could think of the past publications of the PI as a measure of the quality of the unit, while *RCF* measures the quality of the project. (Although clearly the

two overlap.) As with other self reported variables, we assume that research collaborations are exogenously given, at least in the short run.

2.2

Variables and correlations

Table 1 defines all the variables used in our empirical analysis.

Table 2 presents

descriptive statistics. Fig.s 1-4 show the distributions of budget asked, budget granted, past publications of the PI, and publication output. All four distributions are skewed, especially the two distributions of publications. These appear to be consistent with an assumption of log-normality in the error terms.

Tables 3 and 4 show reduced form regressions of budget asked, budget granted, and publication output (in logs). We show these regressions for our total sample (797 units), and for the sample of units that were selected for funding (347 units). Because *P* and *B* take values 0 for unselected units, we used

[TABLE]

Tobit rather than OLSQ in the corresponding full sample regressions, and we set $\ln B = 0$ for $B = 0$. Thirty-four selected units also produced zero publications. Hence, we used $\ln(I+P)$ and Tobit in the publication equations for the selected units as well. Because two PIs in our sample had zero past publications, we used

$\ln(I+K)$, and we used $\ln(I+RCF)$ because $RCF = 0$ for some units. Finally, we show our publication equations with and without $\ln B$ among the regressors. Tables 3 and 4 show fairly significant correlations of our dependent variables with a number of covariates. For instance, both budget asked and

budget granted are correlated with transferability ($TRANSF$), past publications, and size of the unit ($\ln N$). In all publication equations, the elasticity of $\ln(I+K)$ retains its statistical significance even after we add budget granted amongst the regressors. This suggests that past performance (which is correlated

with budget, as shown by Table 3) may have a direct and an indirect effect on publication output. The elasticity of budget varies considerably in the full vs. sample (1.06 vs. 0.34). The problem arises because we do not observe the publication output and the resource inputs of the unselected units. The full sample

estimate of this elasticity is based on the extreme assumption that their observed budgets and publications (zero) are their "true" budgets and publications. But the restricted sample elasticity is based on the other important assumption that there are only random differences between the productivities of the selected and

[TABLES 3-4 ABOUT HERE]

unselected groups. If selection was random, the results of the restricted equation would not change in systematic ways if one added information about the unselected units, and 0.34 would be a consistent estimate of the elasticity of budget. If instead the unselected units were less productive, this would be

an underestimate of the true elasticity.

By using information about the characteristics of the unselected units, we can test whether selection is sensitive to variables that influence the productivity of budget, and control for it in the publication equation.⁶ In

⁶ After we control for selection the estimated elasticity of budget ought to become

addition, budget might be endogenous. Our structural model below attempts to address both problems: selection and endogeneity of budget.

3. THE MODEL

closer to the one that would be obtained by using more aggregate data. For instance, if we observed the total research inputs and outputs of our units, there would be no need to control for selection.

AND THE ESTIMATE D EQUATION S
3.1 Two caveats *Knowledge vs. reputation capital*. In this paper we cannot empirically distinguish between knowledge capital or "competence", and "reputation" capital. Past performance may be related to future performance both because they reflect

the quality of the researchers, and because superior performance enhances reputation, and is rewarded by more generous funding. This opens up the possibility that past performance may reflect, not just inherent differences in quality, but also small idiosyncratic differences in early funding success unrelated to

talent. If one were to allow for learning by doing, and other types of learning in research, the state dependence would become even more marked. 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

reflecting the working of the institutional mechanisms. *Marginal vs. total inputs and outputs.* In a given year, a unit may choose to work on more than one projects. 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 reserach 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. Conversely, if there are positive spillovers across projects, one will end up with an under-estimate.

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 reserach 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. Conversely, if there are positive spillovers across projects, one will end up with an under-estimate.

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. 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 project other than the one studied here, we will formally assume that the CNR project is independent of others -- there are no positive or negative spillovers across projects. As a practical matter, this is not a bad 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). Although we do not have a firm benchmark to compare, this is clearly a small number. This would imply that the CNR programme is not intended as the major funding source for most of the units. Therefore, it seemed best

to neglect the possible spillovers across projects. 3.2 The model We model the decision process of the units that apply to a given research programme. Our objective here is to derive three estimable equations -- one for budget requested, one for budget granted, and one for

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 for, and the amount of research effort, also reckoned in monetary units, but unobserved by the

econometrician. 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, A .

■ *CNR*. 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 in the first step, the budget requested for the project is not included as a criterion. Although we have chosen this representation for modelling convenience, we believe that this is a fairly accurate representation of the actual decision making procedures followed by agencies such as the *CNR*.

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 the *CNR* and the research units. In this way, otherwise worthy

projects may be modified to fit into any budget criterion that the *CNR* 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. Fig.5 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 decides on the optimal size of the research project so as

to maximise its utility. The CNR then selects from amongst the projects, and allocates budget. These decisions depend in part upon characteristics of the unit and project observed only by the CNR (ϵ_l 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_l , we allow for ϵ_k to be correlated with ϵ_l . 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_l , as well as with ϵ_k . 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. However, that would require putting a great deal more structure in two ways. First, it would require us to

specify CNR's objective function. Second, it would imply taking explicit account of the CNR's budget constraint, and hence introduce inter-dependence across the different project proposals. In turn, it would require us to specify how research units form expectations about the behaviour of other units, given their estimates of

the unobserved (by us) quality of other units. We defer this for future research.

3.3 Notation

Let:

- R research effort of units
- R^* "plan ned" research effort
- B budget granted by the agency
- A budget requested by units
- I^* latent selection variable
- I index with $I = 1 \Leftrightarrow I^* > 0$ & $I = 0 \Leftrightarrow I^* < 0$
- X, Y, Z

We also define the following expressions.

■

Application selection

equation

where $\varepsilon_1 \sim N(0, 1)$. Define ϕ, Φ to be the standard normal and the standard cumulative normal evaluated at $\pi \cdot \ln Y$ -- i.e. Φ

$$= \text{Prob}(I = 1).$$

■

$$\{EMBED Equation.2\}\{\text{seq Equation } \backslash * \text{ Arabic}\}$$

(3)

$$I^* = \pi \ln Y + \varepsilon_1 \quad (1)$$

Budget grant equation ion

where $\varepsilon_k \sim N(0, \sigma^2)$ is a measure of quality, observed by the units and the agency and not by the econometrician. The other

$$\text{term, } \mu_b \sim N(0, \sigma^2)$$

■

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.⁷

⁷ We use the convention that ε 's denote errors whose expected value depends upon I , and μ 's errors whose

■

Research effort

where $\{EMBED Equation.2\}\{s$ eq Equation $\backslash * \text{ Arabic } \backslash h\}$ is "planned"

research, defined as expected value does not depend on I .

Note that this of the production of t Since the
implies that research unit. publications, i selection does
research ■ and we allow o not depend
effort is not the elasticity n on A, then
simply **Prod** effort to {EMBED
proportional **uctio** depend on Equation.2}{s
to the **n** past 3.4 Optimal eq Equation
{EMBED Equation.2}{seq Equation * Arabic} (4)

allocated **functi** publications K project size
{seq Equation * Arabic} {EMBED Equation.2}

budget. **on of** through a The relevant * Arabic \h}
Instead, the **publi** function $\alpha(\cdot)$. decision is the and {EMBED
relationship **catio** choice of A. Equation.2}{s
between **ns** ■ We assume eq Equation
research that units * Arabic \h}
effort and the **C** choose A to are also
budget **o** independent
depends, where $\varepsilon_p \sim$ **s** of A, and the
indirectly, $N(0, \sigma^2)$ **t** problem of
also upon the Notice that

{seq Equation * Arabic} {EMBED Equation.2}
{EMBED Equation.2}{seq Equation * Arabic} (5)

costs, and $O(\cdot)$ **e** the units boils
other accounts for **q** down to
characteristics stochastic **u**
factors in the **a**

{seq Equation * Arabic} {EMBED Equation.2}{seq Equation * Arabic} (6a)

this problem units.⁸ (From conditional on (3), (3a), and (9))

yields now on lower $I = 1$, (6),

Note that if μ_b is normal, then $exp(\mu_b)$ is log-normal, and its expected value is $exp[E(\mu_b) + \sigma^2/2]$.

Recalling that μ_b is assumed to be

{EMBED Equation.2}{seq Equation * Arabic} (10a)

{EMBED Equation.2}{seq Equation * Arabic} (10b)

To derive the

3.5 The budget where σ_{IK} is where
estim requested and the {EMBED
ated budget covariance Equation.2}{s
equati granted between ϵ_I eq Equation
ons equations, and ϵ_K , and * Arabic \h}.

We now substitute a {EMBED
derive our from (6a) in Equation.2}{s
three the expression eq Equation
equations to (2) for b , * Arabic \h}.

{EMBED Equation.2}{seq Equation * Arabic} (8)

This

distributed independently of ϵ_i , it is easy to show that the first order

condition of

be estimated, *viz* (8), (9), (12). They will be defined for
procedure implies that information from the non funded units is only used to estimate the selection equation, and
to be equal to the first four terms of (9), and notice that

{EMBED Equation.2}{seq Equation * Arabic} (6)

selected probability of selection. $I = 1$, given {EMBED
Equation.2}{s

and produces observations $I = 1$ after substituting the estimated ϕ and Φ .

$$\{ \text{Estimate} \}_{s=1}^I \quad (11)$$

This two step procedure was chosen because it is convenient

$$\{ \text{Estimate} \}_{s=1}^I \quad (12)$$

and robust. In the second step, we first estimated (8) and (9) jointly by GLS. We used the estimated values of λ , $\{ \text{Estimate} \}_{s=1}^I$ is the covariance between $\{ \text{Estimate} \}_{s=1}^I$ and $\{ \text{Estimate} \}_{s=1}^I$. We then estimated (8), (9), and (12) for

σ_{IK} to compute {EMBED Equation.2}s eq Equation * Arabic \h}, which we substituted in (12). We then estimated (12) by Tobit using different specifications for $\alpha(K)$.⁹ Because the errors in (8), (9), and (12) are

⁹ The reason why we estimate (12) separately by Tobit is that, as suggested earlier, some selected units produced zero publications.

heteroskedastic, we used a robust procedure to compute the standard deviations of the estimated parameters (Eicker-White).¹⁰

¹⁰ We also tried the following estimation procedure. First, (8) is estimated using all observations.

The fitted value of a is then replaced in the structural form of the budget equation, (2), and we estimated the budget equation by maximum likelihood using all observations,

Regressors. To predict selection (vector \mathbf{Y}) we allowed for institutional differences, *DCNR* and *DUNI*, policy objectives, *DSOUTH* and *TRANSF*, and differences in quality,

taking into account that for $I = 0$, $b = 0$. This produces estimates of ϕ and Φ , which we used along with the estimated $r\{eq \ O()\}$ to estimate (12) by Tobit. Our results did not change.

$\ln(I+K)$ and $\ln(I+RCF)$. We also accounted for the fact that research groups in larger institutions, $\ln(NUIST)$, or in metropolitan areas, $\ln(POPPI)$, or with older PIs $\ln(AGEPI)$, may be more likely to be selected because of greater visibility, lobbying capabilities, or reputation not accounted for by past

publications or other scientific reasons.

As far as equations (8), (9), and (12) are concerned, recall that: the variables in \mathbf{Z} affect the cost of preparing a project of "size" A -- equation (5); the variables in \mathbf{X} account for factors that affect the fraction of A granted -- equation (2); the vector \mathbf{H} accounts for variables that affect the

productivity of publications -- equation (4). We used the following regressors for **Z**, **X**, and **H** (corresponding θ , η , and γ parameters in parenthesis):

Z - regressors

X - regressors

H - regressors

· const (η_o)

const (θ_o)

· DSOUTH	·	ln(AGEPI)
(η_s)	ln(1+RCF)	(θ_A)
const (γ_o)	(θ_C)	· ln(N) (η_N)
· DPRO7 (η_7)	·	·
DSOUTH	ln(NUIST)	The Z - and X
(θ_s)	(γ_U)	- regressors
·	· ln(NUIST)	contain all the
DPRO7 (θ_7)	ln(N) (γ_N)	variables used
· TRANSF	(η_U)	for selection.
(η_T)	·	We used
DPRO7 (γ_7)	ln(NUIST)	these
· DCNR	(θ_U)	variables in
(η_{CNR})	·	the X vector
·	ln(POPPR)	because we
DCNR	(γ_P)	have no
(θ_{CNR})	· ln(POPPR)	special
·	(η_P)	reasons to
DSOUTH	·	exclude from
(γ_s)	ln(POPPR)	the budget
· DUNI	(θ_P)	decision
(η_{UNI})	·	variables that
·	ln(AGEPI)	affect
·	(γ_A)	selection.
·	· ln(AGEPI)	Likewise, we
DUNI (θ_{UNI})	(η_A)	have no
·	·	special
·	ln(1+RCF)	reasons to
const (θ_o)	(η_C)	exclude them
·	·	
TRANSF (γ_T)	·	

from **Z**. In fact, apart from $\ln(1+K)$, $\ln(1+RCF)$, *TRANSF*, *AGEPI*, and *DSOUTH*, units in larger institutions, $\ln(NUIST)$, or metropolitan areas, $\ln(POPPR)$, may take advantage of scale, shared facilities or other externalities which encourage them to write larger grants. One reason for including *DCNR* and *DUNI* in **Z** is

that in Italy universities and especially CNR units, which are public research institutions, have to comply with elaborate bureaucratic procedures. For instance, new people can be hired only through complicated contractual arrangements, and the purchase of new equipment or materials demand involved accounting

practices. Other things being equal, CNR units and partly universities then face greater adjustment costs than private research institutions, like foundations or firms, in expanding the size of their projects. The vectors **X** and **Z** include *DPRO7*. The reason why *DPRO7* was excluded from selection

is that 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, transferability, etc.. At the same time, 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 utilise expensive equipment or facilities to carry out development and testing activities. Thus, the organisation of the proposal may require, to a greater extent than projects in the other more "scientific" sub-programmes, time- and resource-

consuming steps like arranging for the use of such equipment or even for their rental or purchase. Apart from *DPRO7* affecting the cost of writing a grant of size *A*, we allowed for the possibility that, in realising this, CNR supplies these projects with a greater share of funds, if they are selected. Finally, we included the

size of the team, $\ln N$, in **Z**. Larger teams can write larger grants because of greater specialisation amongst its members, or because *N* proxies for other resources available to the group. The productivity of publications (vector **H**) is affected by *DPRO7* and *TRANSF* because instrumentati

on or more transferable projects demand more time on activities other than publications. We also included *DSOUTH* in **H**. Southern units enjoy fewer externalities or extra-resources. For instance, unlike the North, in the South of Italy there are practically no research-intensive pharmaceutical, chemical or

biotechnology firms. Hence, to the extent that geographical distance matters, these units are less favoured in obtaining support from industry, which constitutes an important source of funding in fields like biotechnology. Apart from $\ln(I+K)$ and $\ln(I+RCF)$, we used $\ln N$ as another proxy for additional resources available to

the unit.¹¹ While $\ln(NUIST)$ and $\ln(POPPR)$ control for externalities, $\ln(AGEPI)$ may reflect conflicting factors, like experience of the PI or declining age productivity. *Identification.* Note that we identified our parameters through two

¹¹ The assumption here is that the size of the group is exogenous, at least in the short-run.

key exclusion restriction. First, the size of the team does not affect the fraction of budget granted, i.e. X does not contain N . This is a plausible restriction, given that the fraction of the budget granted is allowed to depend upon measures of quality and transferability. Second, we identify the elasticity of budget in the

publication equation by assuming that differences in institutional types, *DCNR* and *DUNI*, do not directly influence the productivity of publications, but they influence the fraction of budget granted. We already suggested that CNR and university units may face greater costs of writing larger grants. These units are also 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, and this is especially true of CNR units -- i.e. the research laboratories of CNR itself. Apart from enhancing their "lobbying" capabilities,

this implies that CNR is better informed about the activities and reputation of these groups and their projects, which reduces uncertain about their outcomes, and encourages the agency to provide them with greater expected grants. Once we control for other factors, there is no critical reasons why the productivity

of research groups should differ according to institutional types. Finally, notice that while we identify the budget granted equation by assuming that there are variables in Z that are not in X , we identified the budget asked equation by assuming that the cost of writing a proposal is linear in A -- equation (5). As one can

see from (6), this implies that in the structural form of the equation for budget asked, the elasticity of expected budget is imposed to be equal to 1.¹²

¹² We also want to note that we experimented with many different covariates in our Y , X , Z , and H regressors, including variables in the application forms that were not described in the paper. Although we did obtain some differences in the

4.2

R
e
s
u
l
t
s

Table 5 shows the results of the probit estimates. The probability of being selected increases with past publications and research collaborations. _____ estimations, all qualitative results discussed below did not change.

More transferable projects, units in more populated areas, CNR units and university units are also more likely to be selected.

Table 6 reports the results of joint-estimation of (8) and (9). 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 that pronounced. CNR acts consistently and puts no weight on NUIST or POPPR in the funding decision (θ_U and θ_P).

Prototype development projects are more expensive to set up (η_7), but CNR grants them a larger fraction (θ_7). Transferability and past publications increase the fraction of the budget granted (θ_T and θ_K). CNR units also obtain a larger fraction of expected budget (θ_{CNR}).

[TABLE 5 ABOUT HERE]

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 respectively 0.36 and -

0.19, and they are well measured. The positive sign of η_S is puzzling, as it would suggest that southern units have lower costs of writing larger grants.¹³ All

¹³ One possibility is that the southern units believed that the budget rule (2) implied a greater elasticity on *DSOUTH* than the one settled by the CNR. According to this interpretation, while southern units believe that the elasticity of *DSOUTH*

other variables in Table 6 have small and statistically insignificant impacts. We estimated the production function of publications (12) by experimenting with different specifications for $\alpha(K)$. In the end, we settled on a logistic specification, $\{EMBED$ in (2) is 0.17 (i.e. $\theta_S + \eta_S = -0.19 + 0.35$), the latter is in fact $\theta_S = -0.19$.

Equation.2}\{s eq Equation * Arabic \h}, where α_o and β are parameters to be estimated. The logistic performed better than, for instance, a constant elasticity $\{EMBED$ Equation.2}\{s eq Equation * Arabic \h} and a logarithmic specification, $\{EMBED$ Equation.2}\{s eq Equation * Arabic \h}, which we attribute to the fact that it allows $\alpha(K)$

to capture the allegedly skewed distribution of productivities in scientific research. The logistic specification also has the interesting interpretation that, with $\beta > 0$, α_o is an upper bound to the elasticity of budget. Table 7 reports our results using the logistic specification, and Fig.6 shows the distribution of $\alpha(K)$. At the

median value of K for the selected units the logistic produced an elasticity of budget of 0.58. Table 7 also shows that bio-instrumentation projects and southern units have lower productivities, whereas collaborations with foreign institutions improves productivity. The size of the team does not affect productivity. To the extent that N proxies

for other resources available to the units, this project appears to be insulated from other activities of the group. In turn, this is consistent with our assumption suggested earlier that these are "small" projects. Externalities within the same institution or from being part of large metropolitan areas do not play a

significant role either. The magnitude of the corresponding coefficients is small. Although their statistical significance is not impressive, the point estimates of $TRANSF$ and $\ln(AGEPI)$ are negative and of non-negligible size. 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 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. In a cross section data set such as ours, no

further interpretation can be given to this coefficient. The coefficient of $TRANSF$ suggests that, other things being equal, there is a trade-off between industrial transferability and publications: projects with industrial applications in mind produce about 20% fewer publications.

5.

RET

TABLE 7 AND FIGURE 6 ABOUT HERE

URN Our through were rejected. that equalises
S TO methodology budget can It may then the marginal
PAST enables us to seriously amplify be interesting products of
PERF perform some differences in to assess the the expected
ORM experiments. scientific extent to publications
ANC In section 5.2 production. which this of the
E below we indirect effect selected units.
AND compute the But, to our depends upon Because we
OPTI total returns knowledge, the conditions do not
MAL to past no study has of entry into observe the
RES performance measured this the system publication
OUR *K*. These are effect, and we (selection) output of the
CE the returns on do not know vis-a-vis the unselected
ALL the ex ante much about size of grants units, we
OCA expected its magnitude given cannot
TION publication and selection. estimate their
5.1 output -- i.e. importance. marginal
before the Moreover, in Section 5.3 products, and
Exper selection and this performs another we performed
imenti allocation programme, experiment this
ng decisions take like in most experiment
with place. As publicly funded parameters only for the
our discussed in research programmes, units that
estim the introduction, more than we ask what were funded
ated the indirect more than would it be by the CNR.
param the indirect more than the allocation This amounts
eters effect of past half of the of resources to taking the
performance applications of resources actual

selection of the CNR as given, and looking for the allocation of resources that maximise the expected total publications enterprise, and its mechanisms for resource allocation, these experiments are suggestive of how our methodology could be used To compute the returns to past performance, the unconditional expectation of publications is $\{EMBED Equation.2\}$ $\{EMBED Equation.2\}$ $\{EMBED Equation.2\}$ the selection equation, and using our logistic specification for $\alpha(K)$, $\{EMBED Equation.2\}$ $\{EMBED Equation.2\}$ $\{EMBED Equation.2\}$ through increases in probability of selection (given expected budget conditional upon selection) and the third term

$$\{EMBED Equation.2\} \{seq Equation \ \{EMBED Equation.2\}\} \quad (13)$$

of the programme. The efficient allocation can then be compared with the actual one. as a practical tool for ex post evaluations of publicly funded research programmes. $\{EMBED Equation.2\}$. Using (11), the elasticity with respect to past performance is $\{EMBED Equation.2\}$. In (13), the first terms measures the direct effect of past performance on publications given the expected budget, the second term measure the indirect effect through the increase in expected budget conditional upon selection.

Apart from better understanding the scientific production 5.2 Elasticity w.r.t. past publications where $k \equiv \ln(1+K)$, π_k is the elasticity of $\ln(1+K)$ in 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. But we can estimate it by taking the expectation of (13) over ε_k , μ_b , and ε_p , given $I = 1$. This yields

selected units. Table 8 presents descriptive statistics of $\{eq \backslash O()\}_k$, and of its three components, $\{eq \backslash O()\}_{k1}$, $\{eq \backslash O()\}_{k2}$, $\{eq \backslash O()\}_{k3}$. The Table shows that the critical components of $\{eq \backslash O()\}_k$

budget given selection is smaller. Moreover, the total elasticity of past publications is greater than 1 for 47 of our 347 units, and as shown by Fig.8 these values are associated with higher

smaller again.) Highly reputed PIs exhibit increasing returns to past performance. Although our cross-sectional data do not allow

consistent with the observed skewness in the distribution of publication performance. As shown by Table 8, the estimation of the direct effect tends to be well below one. The reinforcing

$$\{EMBED Equation.2\}\{seq Equation \backslash * Arabic\} \tag{14}$$

Fig.7 shows the distribution of $\{eq \backslash O()\}_k$ for our sample of

are the direct effect, and the indirect effect through selection. The indirect effect through

$\ln(I+K)$. (Although for very high values of past publications $\{eq \backslash O()\}_k$ becomes

us to linger on dynamic consideration, increasing returns for highly reputed PIs would be

effect due to the institutional mechanisms of resource allocation is then critical in

producing that affect the 5.2 Given our We used the
 increasing budget
 returns, and conditional Optim of the following
 the most upon al production expression to
 estimate the

$$\{\text{EMBED Equation.2}\}\{\text{seq Equation * Arabic}\} \quad (15)$$

significant selection but alloca function, marginal
 impact takes not selection tion $\{\text{EMBED}$ product of
 place at the itself. Even of Equation.2}\{s publications
 selection so, we believe resour eq Equation given budget
 stage. It must that the ces * Arabic \h}.
 be noted that results point To "re-

$$\{\text{EMBED Equation.2}\}\{\text{seq Equation * Arabic}\} \quad (16)$$

the relative to the great allocate" Moreover,
 insensitivity importance of resources using (11), where
 of the the selection amongst the and the fact $\{\text{EMBED}$
 expected process, selected units that Equation.2}\{s
 budget where in this $\{\text{EMBED}$ eq Equation
 conditional reputation programme Equation.2}\{s * Arabic \h}
 upon and quality of we obtained eq Equation is the
 selection may the research an estimate of * Arabic \h}, exponential of
 simply reflect unit appear to their marginal the three
 a play a very product of terms in (16)
 measurement prominent publications, and that do not
 issue -- it is role. *MP*, for given involve *b*.
 difficult to levels of Note that in
 find variables budget.

from B and $\alpha(K)$, the marginal product of publications depends on the other characteristics of the units that are in {EMBED Equation.2}{s eq Equation * Arabic \h}. (Hence, even though $\alpha(K)$ increases in K , the marginal product need not.)

We ranked our selected units according to K , we picked the mid-

points of each quartile, and re-allocated budgets to equate (17) for given total budget.

Table 9 shows that to maximise publications about 85% of total budget should be allocated to the unit with the largest K .

Note that the top quartile of the selected units correspond to about 10% of the total applicants to the programme. Also, as

shown by Table 10, 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). The efficient allocation then seems to be of reasonable size, and we are not proposing grants that are far beyond the actual "productive"

capacity of the units.¹⁴

Needless to say, this result should be interpreted with caution. Allocations based only on past performance may prevent entry into the system, and hence

¹⁴ We experimented with many different observations around the mid-point of each quartile, and the results discussed here are fairly robust.

discourage the rise of talented new scientists. [TABLE 9 AB

Moreover, there could be discouraging effects produced by previous successes -- e.g. researchers who had just obtained important accomplishments may be less motivated to seek new ones. In fact, the correct way to interpret our result is to say that if in this particular programme

CNR had wanted to maximise publications, this is the allocation that would have had to be produced. At the same time, this opens the question whether maximising the number of publications ought to be the main objective of science policy. Publications may not be perfect substitutes with one another. For

instance, it may be desirable to encourage certain fields rather than others, or certain groups (e.g. younger people) rather than others. Put differently, we are simply providing evidence of the extent of the trade-off between total publications per se and other goals in publicly funded research. Table 9 also shows the

actual allocations, and the estimated total publications of the four units, as well as the estimated total publications if the four units obtained the same share of funds.¹⁵ Note that CNR's actual allocation

¹⁵ The estimated total publications are computed from (16) after replacing b with its corresponding values in Table 9.

produces about the same total publications, in expected terms, as the equal share allocation. The optimal allocation produces about 19% more total publications (in quality-adjusted units) than the equal-share or CNR allocation. The similarity between the equal share and the CNR allocations appears to be consistent with our point

made earlier that in this programme strict evaluation criteria were enforced primarily at the selection stage. After selection less effort was exercised in optimising the distribution of resources. We do not have the information to make meaningful comparison, but we wonder whether the same applies to the public research

programmes of institutions like the NSF, the British Council or the European Union, whereby quality standards and other criteria are implemented mainly at the selection stage vis-a-vis the allocation decision given selection. We hope to be able to address this question in future research.

6. 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 mobilising not just human resources but costly instruments, materials, and facilities as well. Studies of research performance under the norms of "open science" must develop along lines that take

these factors into account. This paper 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 selection biases which turned out to be quantitatively important. In the particular research programme that we have studied there are direct and indirect effects of past performances on publication. The indirect effect works through the institutional process by which research budgets are

allocated. Superior performances of the group leader in the past increase the probability of the research proposal being selected, and also increase the fraction of the budget award. These indirect effects turned out to be substantial. While the combined elasticity is on average 0.64, the indirect component accounts for almost 60%

of this figure. Moreover, the elasticity varies across research units, and for a small fraction of our applicants (about 6%), it is greater than unity, even though the direct effect would still imply an elasticity well below unity. This suggests that the institutional mechanisms of resource allocation are a vital contributor to the existence

of "increasing returns to knowledge capital" in scientific research. The marked skewness in publication output of scientists might then be due, to a significant extent, to the way in which scientific research is funded. In turn, this points to the importance of controlling for funding levels, and for selection in estimating the

productivity of scientists. Failure to do so will produce exaggerated estimates of the distribution of innate ability. Our analysis is not a normative one. We do not mean to suggest that the institutional mechanisms for funding scientific research are inefficient, nor that using past performance to estimate

productivity is incorrect. Neither do we claim that exaggerations in the skewness of publication output is socially undesirable. Absent systematic time series data on inputs (funding levels), these questions simply cannot be answered. Our objective in this paper was to uncover how ability, reputation, and the way in which

scientific
research is
funded
interact in the
production of
scientific
knowledge.

Allison, P., Price, D.J. de Solla, Griffith, B., Moravcsik, M., and Stewart, J. [1976] "Lotka's Law: A Problem in its Interpretation and Application", Social Studies of Science, Vol.6, pp.269-76.

Allison, P., Long, S. and Krauze, T. [1982] "Cumulative Advantage and Inequality in Science", American Sociological Review, Vol.47(5), pp.615-25.

Arora, A., and Gambardella, A. [1992] "Public Policy Towards Science:

Picking Stars or Spreading the Wealth?", in Antonelli, C. and David, P. (editors), forthcoming, Path-Dependence and Economic Analysis, Kluwer Academic Publishers, 1995.

Dasgupta, P. and David, P. [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.519-42.

Dasgupta, P. and David, P. [1994] "Towards a

References
Economics of Science", Research Policy, Vol.23, pp.487-521.

David, P. [1992] "Knowledge, Property and the System Dynamics of Technological Change", in Proceedings of the World Bank Annual Conference on Development Economics: 1992, published as a Supplement to the World Bank Economic Review, Summers, L. and Shah, S. (eds.), Wahington DC, March 1993.

David, P. [1994] "Positive Feedbacks and Research

Productivity in Science: Reopening Another Black Box", in Granstrand, O. (ed.) Economics of Technology, North-Holland, Amsterdam and London.

Jaffe, A. [1989] "Real Effects of Academic Research", American Economic Review, Vol.79 (5), pp.957-70.

Levin, S. and Stephan, P. [1991] "Research Productivity over the Life Cycle: Evidence for Academic Scientists", American Economic Review, Vol.81 (1), pp.114-32.

Lotka, A. [1926] "The Frequency Distribution of Scientific Productivity", Journal of the Washington Academy of Sciences, Vol.16 (12), pp.317-23.

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. and Olivastro, D. [1992] "Status Report - Linkage between Technology

and Science",
Research
Policy,
Vol.21 (3),
pp.237-49.

Nelson, R.
[1986]
"Institutions
Supporting
Technical
Advance in
Industry",
American
Economic
Review
Proceedings,
Vol.76 (2),
pp. 186-89.

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
Science,
Vol.27(5/6),
pp.292-306.

Stephan, P.
and Levin, S.
[1992]
Striking the
Mother Lode
in Science:
The
Importance of
Age, Place
and Time,
Oxford
University
Press, New
York.

Stephan, P.
and Levin, S.
[1994]

Table 1. Definition of variables used for the analysis (PRIVATE)

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

	Mean	Standard Deviation (797 observations)	Minimum	Maximum
	0.435	0.496	0	1
	350.244	275.743	25	4224
I	105.035	62.147	3	519
A (*)	17.468	22.442	0	199.19
B (*) (+)	(6.761)	(6.065)	(0)	(50)
P (+)	0.107	0.309	0	1
(non quality-adj. number)	0.154	0.361	0	1
DPRO7	0.645	0.479	0	1
DCNR	0.189	0.392	0	1
DUNI	0.375	0.484	0	1
DSOUTH	195.260	126.290	100	3447
TRANSF	4.018	3.323	0	25
POPPR (§)	31.672	43.556	0	636.79
RCF	(13.065)	(17.020)	(0)	(226)
K	12.055	6.221	2	99
(non quality-adj. number)	25.152	18.377	1	63
N	52.287	8.490	34	85
NUIST				
AGEPI				

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

Table 3: Reduced Form Estimates
(Budget asked, budget granted)

{PRIVATE }	OLSQ lnA	OLSQ lnA (I=1)	TOBIT lnB	OLSQ lnB (I=1)
const	3.497 (0.530)	3.691 (0.722)	-10.91 (4.667)	3.742 (0.782)
DPRO7	0.165 (0.079)	0.141 (0.106)	1.089 (0.572)	0.413 (0.121)
DCNR	-0.060 (0.067)	0.033 (0.093)	1.587 (0.567)	0.244 (0.111)
DUNI	-0.038 (0.057)	0.045 (0.092)	0.934 (0.525)	0.050 (0.115)
DSOUTH	0.274 (0.051)	0.231 (0.077)	0.164 (0.410)	-0.127 (0.078)
TRANSF	0.122 (0.040)	0.109 (0.054)	1.282 (0.340)	0.095 (0.061)
ln(N)	0.547 (0.059)	0.543 (0.093)	0.842 (0.375)	0.177 (0.074)
ln(1+K)	0.065 (0.021)	0.105 (0.027)	1.489 (0.176)	0.092 (0.033)
ln(1+RCF)	0.102 (0.030)	0.118 (0.039)	0.752 (0.258)	0.042 (0.043)
ln(NUIST)	-0.066 (0.026)	0.006 (0.037)	-0.144 (0.206)	-0.017 (0.042)
ln(POPPR)	0.041 (0.027)	-0.001 (0.039)	0.643 (0.235)	0.031 (0.046)
ln(AGEPI)	0.101 (0.129)	0.037 (0.160)	-0.483 (1.070)	-0.089 (0.199)
N.obs.	797	347	797	347
Adj.R ²	0.242	0.308	--	0.089
LogLik	-648.5	-237.0	-1249.5	-293.9

Heteroskedastic consistent standard errors in parenthesis.

Table 4: Reduced Form Estimates
(publication output)

{PRIVATE }	TOBIT ln(1+P)	TOBIT ln(1+P) (I=1)	TOBIT ln(1+P)	TOBIT ln(1+P) (I=1)
const	-6.530 (2.990)	3.280 (1.701)	-1.287 (1.521)	2.005 (1.713)
ln(B)	--	--	1.063 (*) (0.031)	0.342 (.114)
DPRO7	-0.004 (0.356)	-0.764 (0.223)	-1.186 (0.217)	-0.913 (0.224)
DCNR	0.756 (0.358)	-0.030 (0.186)	-0.243 (0.182)	-0.113 (0.185)
DUNI	0.420 (0.336)	-0.073 (0.186)	-0.087 (0.172)	-0.091 (0.181)
DSOUTH	-0.352 (0.276)	-0.618 (0.178)	-0.468 (0.182)	-0.576 (0.178)
TRANSF	0.498 (0.223)	-0.137 (0.123)	-0.209 (0.121)	-0.170 (0.123)
ln(N)	0.525 (0.238)	0.130 (0.144)	-0.038 (0.147)	0.070 (0.143)
ln(1+K)	1.118 (0.121)	0.455 (0.070)	0.357 (0.071)	0.423 (0.070)
ln(1+RCF)	0.611 (0.165)	0.233 (0.084)	0.202 (0.082)	0.220 (0.082)
ln(NUIST)	-0.090 (0.129)	-0.002 (0.067)	0.003 (0.065)	0.002 (0.066)
ln(POPPR)	0.341 (0.147)	-0.074 (0.085)	-0.072 (0.075)	-0.084 (0.082)
ln(AGEPI)	-0.435 (0.685)	-0.550 (0.392)	-0.407 (0.368)	-0.522 (0.382)
N.obs. (positive obs.)	797 (313)	347 (313)	797 (313)	347 (313)

LogLik	-1000.7	-520.4	-544.2	-515.1
--------	---------	--------	--------	--------

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

**Table 5: Selection Equation
(PROBIT)**

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

{PRIVATE }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(1+RCF)	0.224 (0.075)
ln(1+K)	0.398 (0.052)
ln(NUIST)	-0.034 (0.055)
ln(POPPR)	0.169 (0.063)
ln(AGEPI)	-0.065 (0.303)
Log Likelihood	-475.4
N.obs (of which positive)	797 (347)

**Table 6: GLS estimation of *Budget Requested* and *Budget Granted* equations
(Equations (8) and (9) -- GLS)**

Dependent variables: $\ln(A)$ and $\ln(B)$. N obs. = 347, for I = 1.

Log likelihood function = -522.6

{PRIVATE }Parameters	Estimates	Parameters	Estimates
const (η_o)	-0.051 (0.887)	const (θ_o)	1.769 (1.252)
DCNR (η_{CNR})	-0.210 (0.115)	DCNR (θ_{CNR})	0.322 (0.146)
DUNI (η_{UNI})	-0.004 (0.116)	DUNI	0.097 (0.132)
DSOUTH (η_s)	0.358 (0.103)	DSOUTH (θ_s)	-0.185 (0.088)
DPRO7 (η_7)	-0.272 (0.156)	DPRO7 (θ_7)	0.369 (0.125)
TRANSF (η_T)	0.014 (0.073)	TRANSF (θ_T)	0.130 (0.107)
$\ln(1+K)$ (η_K)	0.013 (0.038)	$\ln(1+K)$ (θ_K)	0.136 (0.107)
$\ln(1+RCF)$ (η_C)	0.076 (0.053)	$\ln(1+RCF)$ (θ_C)	0.049 (0.078)
$\ln(NUIST)$ (η_U)	0.023 (0.045)	$\ln(NUIST)$ (θ_U)	-0.026 (0.041)
$\ln(POPPR)$ (η_P)	-0.033 (0.052)	$\ln(POPPR)$ (θ_P)	0.065 (0.061)
$\ln(AGEPI)$ (η_A)	0.126 (0.218)	$\ln(AGEPI)$ (θ_A)	-0.115 (0.191)
$\ln(N)$ (η_N)	0.366 (0.120)	σ_K	0.315 (0.389)
$\ln(A)$ (λ)	0.319		

(0.151)

Heteroskedastic consistent standard errors in parenthesis.

**Table 7: Publication equation -- logistic $\alpha(K)$
(Equation (12) -- TOBIT)**

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

{PRIVATE }	$\alpha(K) = \alpha_0 / (1 + \text{EMBED Equation.2}) \{ \text{seq Equation } \backslash^* \text{h} \}$
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)
$\ln(N)$ (γ_N)	-0.002 (0.140)
$\ln(1+K)$ (γ_K)	0.082 (0.221)
$\ln(1+RCF)$ (γ_C)	0.257 (0.122)
$\ln(\text{NUIST})$ (γ_U)	-0.001 (0.057)
$\ln(\text{POPPR})$ (γ_P)	-0.057 (0.098)
$\ln(\text{AGEPI})$ (γ_A)	-0.515 (0.380)
σ_{IP}	0.146 (0.621)
LogLik	-515.75

Heteroskedastic consistent standard errors in parenthesis.

Table 8: Descriptive statistics for elasticity of past performance
(347 observations for I = 1)

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

Table 9: Optimal allocation of resources

(347 selected units ranked by past publications K , mid-points of each quartile: positions 43, 130, 207, 304)

{PRIVAT E }Marginal product $MP (*)$	$\alpha(K)$	Actual allocation $B (+)$	Efficient allocation $B' (+)$	Marginal product of efficient allocation $MP(B')$	Expected # public.s using actual allocation $EP(B)$	Expected # public.s using efficient allocation $EP(B')$	Equal share allocation $B'' (+)$	Expe public using shr. alloca $EP(B$
0.109	0.677	156	496	0.0875	29.3	64.2	145.75	
0.006	0.604	100	23.5	0.0875	8.2	3.4	145.75	
0.189	0.561	120	38.5	0.0875	11.4	6.0	145.75	
0.049	0.531	207	25	0.0875	12.6	4.1	145.75	
		Total: 583	Total: 583		Total: 61.5	Total: 77.7	Total: 583	

- (*) Computed as $\alpha(K) \cdot P/B$ using actual P and B and estimated $\alpha(K)$, logistic.
- (+) Millions of Italian Lire
- (^) Quality-adjusted number

Filename: CNR7PAP.DOC
Directory: C:\ALF\CNR\CNRPAP
Template: C:\WINWORD\TEMPLATE\NORMAL.DOT
Title: REPUTATION AND COMPETENCE IN PUBLICLY
FUNDED SCIENCE:
Subject:
Author: The Heinz School
Keywords:
Comments:
Creation Date: 05/29/96 11:09 AM
Revision Number: 2
Last Saved On: 05/29/96 11:09 AM
Last Saved By: The Heinz School
Total Editing Time: 2 Minutes
Last Printed On: 05/29/96 11:11 AM
As of Last Complete Printing
Number of Pages: 52
Number of Words: 9,212 (approx.)
Number of Characters: 52,512 (approx.)