2010
10

Working Paper

INSTITUTO DE POLÍTICAS Y BIENES PÚBLICOS (IPP)

KNOCKING DOWN SOME ASSUMPTIONS ABOUT THE EFFECTS OF EVALUATION SYSTEMS ON PUBLICATIONS

CARMEN OSUNA
LAURA CRUZ CASTRO
LUIS SANZ MENÉNDEZ

CONSEJO SUPERIOR INVESTIGACIONES CIENTÍFICAS (CSIC)

INSTITUTE OF PUBLIC GOODS AND POLICIES (IPP-CCHS)

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

CCHS
Centro de Ciencias Humanas y Sociales

IPP
INSTITUTO DE POLÍTICAS Y BIENES PÚBLICOS

INSTITUTO DE POLÍTICAS Y BIENES PÚBLICOS CCHS-CSIC

Instituto de Políticas y Bienes Públicos
Centro de Ciencias Humanas y Sociales
Consejo Superior de Investigaciones Científicas
C/ Albasanz, 26-28.
28037 Madrid (España)

Tel: +34 91 602 2300
Fax: +34 91 304 5710

http://www.ipp.csic.es/

How to quote or cite this document:

Osuna, C., Cruz Castro, L., & Sanz Menéndez, L. (2010). Knocking down some Assumptions about the Effects of Evaluation Systems on Publications. Instituto de Políticas y Bienes Públicos (IPP), CCHS-CSIC, Working Paper, Number 10.
Available: http://hdl.handle.net/10261/27730

2010
10
Working Paper
INSTITUTO DE POLÍTICAS Y BIENES PÚBLICOS (IPP)

KNOCKING DOWN SOME ASSUMPTIONS ABOUT THE EFFECTS OF EVALUATION SYSTEMS ON PUBLICATIONS

CARMEN OSUNA, LAURA CRUZ CASTRO & LUIS SANZ MENÉNDEZ
CONSEJO SUPERIOR INVESTIGACIONES CIENTÍFICAS (CSIC)
INSTITUTE OF PUBLIC GOODS AND POLICIES (IPP-CCHS)
CARMEN.OSUNA@CCHS.CSIC.ES, LAURA.CRUZ@CCHS.CSIC.ES& LUIS.SANZ@CCHS.CSIC.ES

ABSTRACT

In 1989 the Spanish Government established an individual retrospective research evaluation system (RES) for public researchers. This system was voluntary but involved the incentive of a permanent salary bonus. Policy makers have associated the establishment of this evaluation system with the significant increase in the volume of scientific publications attributed to Spain over the last decades. In a similar vein to the analyses of other country cases, some scholars have also claimed that the growth of Spain's international scientific publications is a result of the establishment of the new evaluation system. In this paper, we provide a methodological revision of the validity threats in previous research, including some interrupted time series analyses and control groups to investigate the effects of this policy instrument on the number of papers produced by Spanish authors. In the years following the establishment of the evaluation system, the results indicate a considerable increase in the number of papers attributed to Spanish authors among those eligible for evaluation (the "treated" group), but also in the control groups. After testing various alternative explanations, we conclude that the growth in Spanish publications cannot be attributed indisputably to the effect of the establishment of the RES, but rather to the increase of expenditure and number of researchers in the Spanish R&D system along with some maturation effects. We take this case as an example of the need to improve and refine methodologies and to be more cautious when attributing effects to research evaluation mechanisms at the national level.

Keywords: Research evaluation systems; Scientific production; Effects of evaluations; Quasi-experimental design; Spain; CNEAI

Corresponding author: Luis Sanz-Menéndez (Luis.Sanz@cchs.csic.es). Tel: +34 916022549. Fax: +34 916022971.

1. CONTENTS

1. INTRODUCTION

The high growth rate of Spanish scientific production since the early eighties has attracted the attention of scholars (Gómez et al. 1995; Méndez and Gómez 1986; Van Raan 1997; Zitt et al. 1998). In less than a decade the total contribution of Spain to the world scientific production has more than doubled and its share has gone from 0.6% in 1979 to 1.5% 1989 and 2.2% in 1995 (Van Raan 1997). In 1982, a year in which a new government took office, Spain had 5,112 publications included in the Thomson-Reuters -ISI databases-, whereas in 1989 the number increased to 10,482. The growth has continued in the nineties to a total number of 19,142 in 1995 and 26,617 in 2000. While the average world annual rate of growth of papers in the period from 1981 to 1994 was 3.7% (May 1997), in Spain the increase was more than 11 % per year, on average.

In 1982, the newly elected Spanish Government launched a significant effort to increase funding and resources for R&D, to change academic regulations and to develop a new science policy (Ballart and Subirats 1997; Sánchez-Ferrer 1997; Sanz-Menéndez et al. 1993; Sanz-Menéndez 1995a; Sanz-Menéndez 1997). Like in other countries, there was also an emerging concern regarding research evaluation, performance and results; in addition to other reforms, a *retrospective research evaluation system* (RES) of individual research results was established in 1989 (Sanz-Menéndez 1995b). As in other countries, the establishment by the Government of a national RES was associated to national objectives or seen as a means to solve some perceived policy problems, aiming at changes in the behaviour of scientists and institutions and improvements in research performance.

In some countries, analysis of RES and their impact on scientific publications started some years ago; in general, scholars have established associations between the setting up of research evaluation systems and the growth in scientific papers or the improvement of their quality (e.g. Butler 2003a, b; Moed 2008; Önder et al. 2008.). But also, too often, analysts have assumed that the institutionalisation of an evaluation system has a positive effect on research activities and performance, despite some controversies regarding the long term cost-benefits relations (Geuna and Martin 2003).

What is missing in most of the studies that address the effects of RES on scientific performance is an adequate research design that provides methodological controls of the alternative rival explanations. More precisely, what is missing is a systematic exploration and questioning of the causality attribution between the “treatment” (the set up of a research evaluation mechanism) and the supposed effects (increase in the research publications or their quality).

In Spain, while a lot of attention has been paid to describe and account for the high level

of growth in scientific production (Méndez and Gómez 1986; Gómez et al. 1995; etc.), little attention has been paid to the study of the impact of the RES on publication performance, with some exceptions, such as the work of Jiménez-Contreras et al. (2003), that helped to consolidate the idea that the existence of a research evaluation system has had significant effects on the evolution of the volume of Spanish scientific publications. Policy makers have often presumed that the growth of Spanish scientific production was associated with the implementation of the evaluation system.[1]

This paper aims to contribute to a better understanding of the impact of RES on scientific publications by proposing a methodological approach based on the control of rival explanations and the need for careful consideration of the causality. We apply a methodological approach appropriate for coping with the problem of causal attribution of effects to specific events in cases in which time is a key variable. Thus, a quasi-experimental design with interrupted time series and non equivalent control groups is applied to evaluate the short and long term effects of the evaluation system on scientific production. We use the Spanish case as an example to address the relationship and possible impacts of the evaluation systems on the aggregate volume of scientific publications. The purpose of this paper is to evaluate the effects of the establishment of a RES on the Spanish scientific production and to determine whether or not it can be claimed that the increase in the number of publications has occurred as a result of the establishment of the National Commission for Evaluation of Research Activities (*Comisión Nacional Evaluadora de la Actividad Investigadora* -CNEAI) and the implementation of its evaluation procedures.

The paper is organised as follows. In section two we depict the Spanish RES. In section three, we examine some previous work on the factors affecting scientific publications and the impact of RES. In sections four and five we present the data construction, the methodological design, and the premises of the quasi-experimental analysis. In section six we analyse the evolution of Spanish scientific publications and, using interrupted times-series and non equivalent control groups, we search for competing explanations other than the effect of the RES to account for the growth in the volume of publications. We conclude with a summary and discussion of the findings, and by drawing some analytical and policy conclusions.

2. THE SPANISH INDIVIDUAL RESEARCHERS EVALUATION SYSTEM

The basic features of the Spanish RES were related to other characteristics of the science system and shaped by specificities of the institutional construction, among others: the constitutional autonomy of universities, the civil servant status of academics and public researchers, the lack

1 For those interested in further analysis, a magazine edited by the Ministry of Education and Science between 1985-1997 *(Política Científica)* includes contributions from policy makers with this view: e.g. P. Pascual (n. 20, 1989, pp.28-31), P. Ripoll (n. 21, 1990, pp.5-7).

of individual wage negotiations and the small salary differentiation among different professional categories, a strong degree of external control of public competitive project funding by the scientific elites managed under peer review, and a rather weak institutional and organisational endowment of resources (Sanz Menéndez 1997). A debate began in the late eighties on how to incentivise research, improve the results, promote internationalisation and enhance the quality of research of permanent academics with long life employment.[2] The result of this debate was the design of an institutional system of evaluation of individual research results, which was voluntary in nature and involved a salary bonus. In this institutional context, the Spanish RES has emerged characterised by a strong focus on the retrospective evaluation of individual researchers' outputs (Cruz-Castro and Sanz-Menéndez 2007; Sanz-Menéndez 1995b).

The RES was institutionalised through the creation of the CNEAI,[3] aimed at the retrospective evaluation of the research performance of tenured university professors and National Research Council -CSIC- researchers. The resulting system was more an extension of the evaluation and quality control practices of the scientific communities than a management and steering tool (Rip and van der Meulen 1996) developed by the government. But the new evaluation system also had the objectives of increasing researchers' productivity, improving the quality of scientific publications and their visibility in international journals.

The general principles for evaluation suggest "weighing up the quality, creativity and originality" of the knowledge contributions. The research assessment is based on the examination of individual research output over a six-year period, with the identification by the applicant of five research contributions.[4] It is a peer review system organised in eleven scientific commissions.[5] The evaluation criteria, which are public, have remained quite stable over the years, but have been gradually defined more precisely and the dominant natural sciences criteria have been extended to other areas. Quality indicators were explicitly introduced in the mid-nineties giving preference to publications in journals of recognized international prestige; for most of the scientific fields, the CNEAI recommended publishing in journals included in the Journal Citation Reports (ISI), and especially in those well positioned in the rankings by impact factor. Most of the fields have defined thresholds of a minimum number of relevant contributions required to obtain a positive evaluation.[6]

2 For example, Juan Rojo, the Vice-Minister of Universities and Research, stated in an interview in a Spanish newspaper that the system was an incentive appropriate for rewarding the most productive researchers (*La Vanguardia* 24 March 1990).

3 The creation of the CNEAI was approved in 1989 to operationalise the new payroll item (*complemento de productividad extraordinaria* - Extraordinary bonus productivity) aiming to incentivise individual research activity. More information at http://www.educacion.es/horizontales/ministerio/organismos/cneai.html. Accessed 10 May 2010.

4 Main contributions to be considered as such were journal papers, books and patents.

5 Mathematics and Physics; Chemistry; Molecular and Cellular Biology; Biomedical Sciences; Natural Sciences; Engineering and Architecture; Social, Political, Education and Behavioural Sciences; Economics and Business Administration; Law and Jurisprudence; History and Art; Philosophy, Philology and Linguistics.

6 All regulations are published in the Spanish Official State Gazette (*Boletín Oficial del Estado*). The for-

The immediate effect of succeeding in the research evaluation for a researcher is an average increase, depending on their academic category, of €140 in their monthly salary for each positive evaluation (*sexenio*); approximately each *sexenio* represents 3% of a researcher's annual income (and the maximum, after 36 years of scientific activity are six periods, accounting approximately for 15% of their annual salary). Despite their limited salary effects, *sexenios* have a reputational value and influence the access to other resources and rewards, such as project grants or professional career advancement. Moreover, a minimum number of granted *sexenios* has become a formal requirement to take part in the selection committees for access to permanent university professor positions.

The CNEAI publishes the results of the evaluations with aggregate data by category, discipline and institution,[7] but not at the individual level, although the employing institution is aware of the individual results, because the bonus is to be included in the researchers' monthly cheque. This reward of performance has become a sign of status and recognition among the Spanish scientific community and is interpreted as an incentive in aligning the publication practices of researchers with the evaluation criteria. The impact of this overall system on the aggregate publications level, however, is less clear, considering the possibility that the diversity of the effects among scientific fields could be hidden by aggregate analysis, as it has been reported for other countries (Ingwersen and Jacobs 2004).

3. SCIENTIFIC PUBLICATIONS AND RESEARCH EVALUATION

There is a significant amount of literature on the factors associated to the growth of scientific publications. However, the literature about research evaluation systems and how different RES produce diverse results is limited and often country specific. We will first review some of the factors identified as important in the explanation of publication performance at the national and individual levels. Secondly, we will review some of the studies that have addressed the effects of national RES on research production and visibility.

When trying to account for the evolution of scientific productivity, studies often analyse the relations between the inputs (expenditure on R&D and the stock of researchers) and outputs

mal criteria required for obtaining the positive evaluation were clearly formulated and legally regulated to reduce judicial controversies and Court demands; for example, 1 or 2 papers in the top journals (first quartile) for natural sciences, whereas in most social sciences and humanities areas it is sufficient to have some items (contributions) in the form of papers included in the ISI databases (see for example the 2009 annual call for details at http://www.boe.es/boe/dias/2009/12/01/pdfs/BOE-A-2009-19218.pdf); accessed 3 May 2010.

7 From 1989 to 2005, 38,872 tenured university professors and 2,434 CSIC researchers were voluntarily evaluated. For each group 47% and 75% of them had all *sexenios* positively evaluated; 25% and 21% got some *sexenios* positively evaluated and 28% and 4% did not have any *sexenios*, either because the did not apply or were not approved (CNEAI 2006 a, b). More information is available at http://www.educacion.es/horizontales/ministerio/organismos/cneai/memorias-informes.html; accessed 3 May 2010. There is also some evidence that the amount of positive evaluations, at institutional level, correlate with the volume of ISI publications (Grupo Scimago 2006).

(publications). At country level, relative wealth has been associated with scientific and technological results. A high correlation between relative R&D investments and research performance measured by bibliometric indicators have also been identified (e.g. Braun and Schubert 1988; King, 2004; Moya-Anegón and Herrero-Solana 1999). It is common to find significant correlations between yearly GDP values and number of publications when using time series of different countries. Moreover, when consecutive time series are used, previous R&D investments have been reported to increase research results (Vinkler 2008). Recent analysis of the relationship between investment in science and research outputs over time, at cross-country level, have found time lags between the R&D investment and the onset of the research results and, when controlling the effect of international spillovers, evidence of diminishing returns to R&D investment in terms of publications has been found (Crespi and Geuna 2008).

At country level, the significant differences in the annual growth rates of US academic R&D investment compared with the growth in publications have been attributed to the increase in research costs in several scientific fields (Adams and Griliches 1996). R&D expenditure, especially non defence government R&D investment (Shelton 2008), was found to be more important than the number of researchers (as Crespi and Geuna 2008 suggested) in explaining publication share. For Spain, increases in human and financial resources and other sets of policy and legal changes have been reported to be associated with the growth in publications (Gómez et al. 2006).

The literature about the differences in individual productivity (Fox 1983) has concentrated on two different sets of factors: those related to environmental location and those referred to cumulative advantage processes. Some practices, such as research collaboration (Lee and Bozeman 2005), produce strong and significant effects when counting the number of papers. However the net impacts of collaboration are less clear. (Adams et al. 2005; Glanzel 2001; Katz and Hicks 1997; Leydesdorff and Wagner 2008; Persson et al. 2004). Increases in international collaboration and the fact that more papers are becoming widely distributed globally could also have a positive effect on the research impact..

Research evaluation systems are set up primarily to affect organisational or individual behaviour through incentives and controls, but there is very little evidence of how, and through which processes, research incentives and social controls affect individual productivity. Depending on the specific RES, evaluation criteria and standards are associated in various degrees to budgetary allocations and/or individual rewards, but the question is: Do reward systems or financial incentives determine the work effort and, especially, the publication results? Or, rather, is the publication behaviour of academics driven by factors mostly related to the research process itself and not so much by external (organisational or political) stimuli?

Systematic comparative analysis of the diversity of RES is a promising and relevant research area; some countries have “strong” RES while others have “weak” ones (Whitley & Gläser eds.

2007); the unit of evaluation varies and includes universities, departments or individuals. Cross-country comparative analyses of RES (Campbell 2003; Geuna and Martin 2003; Hicks 2009) and their consequences are very scarce. It is generally said that all RES increase the pressure on scientists to publish and reinforces the "publish or perish culture" among the scientific community. However, it has been argued that evaluation-based funding of research leads to "publications inflation" without necessarily improving their quality and decreasing returns in the long run (Geuna and Martin 2003**)**, and that these systems increase the number of papers by reducing their quality and lead to "salami-slicing" effects in publications **(**Liefner 2003**)**.

The expected effects of retrospective evaluation systems (RES) on scientific production will depend on their relative strength (Whitley 2007), but the responses of institutions and researchers to the implementation of RES are mediated by funding context, scientific fields and academic career stage of scientists (Laudel 2006a, b; Whitley 2003). In fact, further knowledge of the adaptation behaviours of researchers requires identifying the social mechanisms through which RES could affect knowledge production (Gläser 2007). Although little research has been done on the micro-mechanisms, nevertheless some work has highlighted that individual perceptions of RES are important factors, because intrinsic motivations for performance could decrease when incentives are perceived as "controlling" instead of "supportive" (Andersen and Pallesen 2008**).**

Most of the empirical literature refers to countries characterised by "strong" evaluation systems, in which evaluation results have significant effects on the level of funding of institutions and universities.

In the United Kingdom, since 1986, the Research Assessment Exercise (RAE) focuses on institutional performance and it has relevant consequences on university research funding. Early studies of the 1992 and 1996 RAE claimed an increased pressure to publish for academics and a general opinion that the exercise had improved university research quality (McNay 1997; Talib 2001). Evidence was also provided to show that researchers had altered their publication behaviour[8] in two ways: "targeting high impact factor journals and increasing submissions of articles prior to a RAE deadline" (Georghiou et al. 2000: 46). More recently Moed (2008) has interpreted the changing publication patterns in UK as the response of researchers to RAE changing criteria. In the years prior to the 1992 RAE, UK scientists increased their publications, while a shift from "quantity" to "quality" in the 1996 RAE evaluation produced an increase in the average impact factor of publication journals; prior to the 2001 RAE, intra-UK co-authorship increased, although publication productivity did not, a process interpreted as a "back to quantity" response to the evaluation criteria in the 2001 RAE.[9]

8 Similar evidence emerges from the analysis of specific disciplines: e.g. economics (Harley and Lee 1997) or sociology and psychology (Henkel 1999).

9 As a result of the repetition of the evaluation exercises, some learning effects and strategic behaviour by the evaluated units have emerged, such as not including the weaker researchers in the evaluation exercise, even if

After the implementation of The Association of Universities in the Netherlands (VSNU) evaluation, it has also been reported that Dutch scientists in all disciplines felt a higher pressure to publish and to do it "strategically" in international journals with high impact factors (Westerheijden 1997); Moed et al. (1999) showed that there was also an increase in the scientific publications in the natural and life sciences in the context of the VSNU system.

In Australia, the Institutional Research Grant Scheme with institutional funding formulas based on quantitative publication indicators (Gläser and Laudel 2007), has been associated to the increase in the number of Australian publications, [Butler 2003a, b, 2004; Taylor 2001]. Butler suggested that there might be a causal relationship between the Australian formula-based funding and the increase in the number of publications, although in journals with the lowest impact factor. In a system based on publications' scores with no differentiation between publication quality or impact "there is little incentive to strive for placement in a prestigious journal" (Butler 2002: 877). These trends were common to all scientific fields, but only in the university sector. [10]

This type of analysis with loosely coupled associations of growth in publications and institutional changes or evaluation systems has been extended to other countries. For example, Önder et al. (2008) state that increases in Turkish publications could be explained by recent changes in regulations regarding promotion to associate professorship and increased availability of funds for research.

Regarding Spain, Jiménez-Contreras et al. (2003) associated the establishment of the RES with the growth in Spanish publications. Based on a bibliometric analysis of the aggregate evolution of Spanish scientific production in international journals (Science Citation Index -SCI- from Thomson Reuters-ISI), they concluded that the rapid growth of Spanish research output in science in the 90s was due to the establishment of the CNEAI research evaluation system. Others have accepted and refined the argument (e.g. Gómez et al. 2006), but have stated that the evaluation system has barely produced an improvement in the quality/impact of publications (e.g. Rodríguez-Navarro 2009).

Our objective is to overcome the deficits of some of the literature that, overall, has taken an approach mainly based on simple before and after measures, without establishing the appropriate caveats. Our research question is whether we can accept the claim made by analysts, policymakers and bibliometricians that the increase in the Spanish aggregate number of publications in international journals in the 90s was the result of the establishment of the RES (CNEAI).[11] In this paper, we challenge this claim by confronting it with other alternative

that leaves the winners with less money as a result (Barker 2007).

10 She also found some differentiation effects, associated to other intervening variables, and modulated by the response of the universities to the funding formula.

11 In this paper, we adopt an aggregated level of analysis, although we acknowledge that there might be differences among scientific fields that could hide the effects in some areas. For example, there is evidence that the

plausible explanations and designing an approach based on interrupted time-series and control groups.

4. DATA CONSTRUCTION AND SOURCES

Our annual Spanish publication data are taken from the Thomson-Reuters Web of Science and were downloaded from the Science Citation Expanded Index (SCI-Expanded), Social Science Citation Index (SSCI) and Arts and Humanities Citation Index (AHCI) for the period 1980-2005.[12] We have a time-series of Spanish scientific publications, with data for every year, for the period 1980-2005.[13] The selection of this source is consistent with the evaluation criteria and the quality indicators established by the CNEAI, which emphasise international publications.

When trying to evaluate the plausibility of the causal attribution of the increase in Spanish publications to the establishment of the CNEAI, we should take into account other variables that could also be plausible explanations of the increase in the total output and which the literature has identified as important. Zitt et al. (1998) have identified transition processes in publication practices in countries that could serve as a way to assess the development of a research system (Van Raan 1997). In this context the growth over time in the total number of publications could be the result of different substantive processes:[14]

a) The rising commitment of researchers (already active in the system but with no research activity or no results published in international journals) to publish articles in international journals (as requested by the new RES), involving a change in their previous pattern of publications.

b) The increase in the "productivity" of those already publishing in international journals, who, after the establishment of the new incentive system, might publish more per capita (intensification).

c) A rapid increase of the number of researchers in the system.

The first two processes could be possibly associated to the establishment of a research evaluation

average number of ISI publications associated to granting a *sexenio* in the different areas is very diverse (Grupo Scimago 2007).

12 Despite the limitations of this database for non-English speaking countries (Moed et al. 1995), for the period selected, it seems the most suitable.

13 The figure refers to the number of documents, including all items, not just the citable documents. A Spanish paper was defined as a paper in which the list of the authors' corporate addresses contains at least one Spanish institution; thus, we use whole counting (not fractional).

14 Additionally the growth could be the result of instrumentation effects (e.g. simple growth of papers in the databases, the increase in the number of journals included in the databases or, more specifically, Spanish journals) which we also discuss in section 5. See footnote 18 on "instrumentation threats" for more details. Gómez et al. (1995) clearly demonstrated that the Spanish growth was not attributable to changes in the coverage of Spanish journals in ISI databases. The coverage of journals from editors from Spain in ISI even diminished down to 29 in SCI, 2 in SSCI and 15 in AHI in early 2000 (Gómez et al. 2006). Van Raan (1997) presented data on the growth levels of Spain (between 1979-1995), which, in index numbers (base 100) was 523, while world growth was 148, or, for example, the figure for France was 175 and for Italy, 285.

system: changing publication strategies to target international (or English) journals or increasing the number of international papers per capita as a result of more effort or more international collaboration. However, if c) occurs, even if it is simultaneously to the establishment of the RES, a more careful methodological design should be established to assess causality. The evolution of the number of researchers can be controlled through different indicators, but we adopted the number of researchers (full time equivalent –FTE-), more precisely, the number of researchers in the public sector (government and higher education), excluding those working in firms and non-profit sectors, due to the limited level of contribution to the publication output of the latter. These data are taken from the R&D national statistics. Human resources are not the only relevant input. Considering the strong association between wealth and scientific publications, we have also constructed complementary time-series data on the economic inputs for research: Total public sector R&D expenditure to represent the expenditure in higher education and government sectors. Data are taken from the R&D national statistics too.

Economic and human resources are exogenous to publications, but we should also control some endogenous changes in the publication patterns which could seriously affect the interpretation of results. The main factor refers to the diverse dimensions of scientific collaboration, especially the evolution of papers with international collaboration[15] and its impact on the measurement and growth of Spanish scientific publications (Chinchilla and Moya 2007).

Because our aim is to control for the aggregate effect in total production and asses the global impact on productivity, we do not construct data on the evolution of citations as a measure of visibility. As regards this issue, it has been reported that visibility (in terms of normalised impact factor) has improved in the last years (Gómez et al. 2006), although Spain is still lagging behind most Western European countries.

5. A METHODOLOGICAL DESIGN TO COPE WITH THREATS TO INTERNAL VALIDITY

Most of the previous analyses have addressed the issue of the relationship between evaluation and scientific publications either qualitatively, through the opinion of interviewees in surveys, or by bibliometric analyses that tend to use only before and after analytical designs. But, as any non-random design, these studies can be vulnerable to confounding variables or biases which threaten the validity of the causal inferences.

We argue that previous analyses to assess the impact of the CNEAI on the growth of publications have used research designs unsuited to confirm the causal attribution. The empirical argument

15 International collaboration involves authors from at least two different countries, with the output then being attributed in whole counting to all countries.

of Jiménez-Contreras et al. (2003) was constructed solely on the comparison between the publication trend existing before and after the treatment (the creation of the CNEAI).

Our research design aims to address some of the methodological pitfalls of prior analyses based on simple "before-after measures" and its threats to internal validity. The number of Spanish publications was 9,840, in 1998, 10,482 in 1989 and 11,228 in 1990, the year after the "treatment". Under the simplest assumptions of before and after measures, the results could appear to support the claim that there was a sharp increase in the number of Spanish publications in 1990, after the establishment of the CNEAI.

One of the main problems of using publications data with aggregate counting to test the effect of RES is the fact that in time-series, due to the non-independence of the successive measures, the regression is toward the trend-line, which is growing (Campbell 1996; Cook and Campbell 1979). Thus it is essential to test other explanations in addition to the researcher's preferred one, to question the effects of the intervention. With that purpose in mind, we review some of the threats to the internal validity (see for example Shadish, Cook and Campbell 2002: 55, table 2.4.) in connection with the measurement of publication performance at country level and its causal attribution:

1. **History** denotes specific events, other than the experimental treatment, occurring between the pre-test and post-test, which might account for the change. The growth in researchers or R&D resources, the passing of particular laws or acts, or budgetary increases in previous years are history variables affecting the growth in publications.

2. **Maturation** refers to regular changes correlated with the passage of time. The general long-term trend towards an improvement in the functioning of Spanish science, presumably due to more and better resources, increased efficacy of funding programs, etc., belong to this category of factors, as well as the results of maturation expressed by the changing patterns of publication of Spanish researchers.[16]

3. **Testing** refers to the idea that a change may occur as a result of a pre-test, even without the experimental treatment. It is conceivable that the first assessments of the S&T system (Muñoz & Ornia 1986) and publicising of the scientific publications results for 1988 or previous years[17] could have partially produced or reinforced the change of researchers' publication behaviour, even in absence of the CNEAI.

16 Previous research has identified changing patterns of scientific communication, from national to international journals, with studies at institutional level [see Sanz Menéndez and Pfretzschner (1992) for the CSIC, Bellavista et al. (1993) for the University of Barcelona and Jiménez-Contreras (1997) and Jiménez-Contreras and Ferreiro-Aláez (1996) for the University of Granada].

17 Reports made by different international consulting companies were made public [e.g. SRI International (1988) Research Activity in Spain, Portugal and Greece. A 1988 Bibliometric Model Assessment (Report prepared by C.P. Ailes; H.R. Coward and R.R. Fresne), Science and Technology Policy Programme, SRI International, Arlington (Vi), mimeo]; the increasing diffusion of country bibliometric analysis was important too (Braun, Gränzel and Schubert 1985; Schubert, Gränzel and Braun 1989).

4. **Instrumentation** refers to a shifting of the measuring instrument. Databases on publications change continuously and changes in instrumentation could relate to: a) the introduction of new journals and general growth in the databases over time, and b) a higher integration of Spanish journals in the databases.[18] However, even acknowledging that the increase of Spanish papers could not be caused only by instrumentation effects, there is an additional instrumentation threat that should be carefully assessed, which relates to the endogeneity issues associated with international collaboration changes and with the measurement criteria (whole counting).

5. **Regression** effects refer to a situation where a group has been selected for treatment just because of its extreme performance on the pre-test, and if the pre-test and post-test are imperfectly correlated, as they almost always are, it follows that on average the post-test will be less extreme than the pre-test (Campbell 1996; Mohr 2000).[19] The low levels of international publications by Spanish researchers in the 80s can plausibly be argued to have led to the creation of the CNEAI, together with other policy tools; the fact that after 1990 publications showed an improvement could be a regression effect.

Our analysis considers these threats to the internal validity in the explanation of Spanish scientific publications, for which a series of observations has been recorded for periods of time both prior and subsequent to the introduction of the RES (CNEAI). The main argument for using this type of design "is that lack of control and lack of randomization are damaging to inferences of cause and effect only to the extent that a systematic consideration of alternative explanations reveals some that are plausible" (Campbell and Ross 1968; Campbell and Stanley 1963/1995).

6. FINDING "PLAUSIBLE COMPETING EXPLANATIONS"

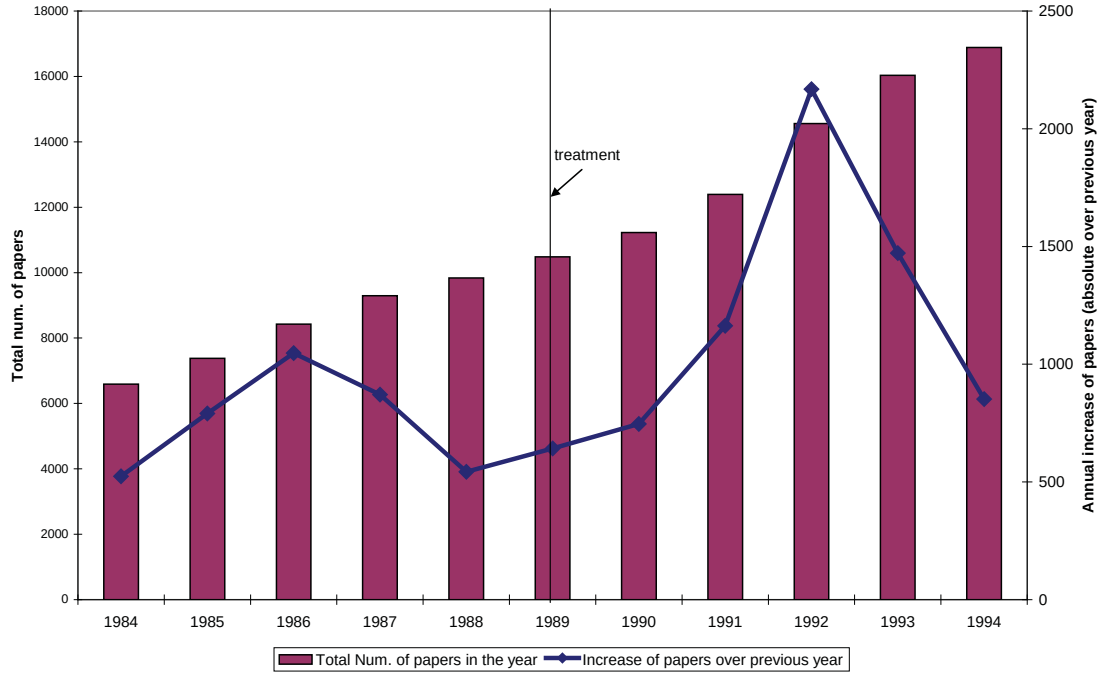
6.1. The "treatment" and scientific publications

When we analyse the data just one year before and after the treatment we find a significant growth; however it could be the case that the growth in publications was a trend already existing before the creation of the CNEAI. If we take a broad perspective and analyse, for example, the period of five years before and after the implementation of the RES system, we observe that the growing trend already existed, but the absolute values of the annual increase in publications over the previous year fluctuate significantly, both before and after.

18 We have analysed whether the number of papers published in Spanish language included in the databases could be the "cause of growth" and the argument should be rejected. Take the following figures: in 1980, 41.6% of papers (1,670) from Spain were published in Spanish, while in 1990 the share was 17.6% (1,976) and 7.25% (1,916) in 2005. While in the 80s we witnessed a small but absolute increase, later on, there was stagnation in the absolute numbers. Zitt et al. (1998) clearly present Spain as a country that has made the "transition" to publication in English in SCI-ISI; the argument is also in Gómez et al. (1995) and Gómez et al. (2006) (see also footnote 14).

19 Selection for extremity (and resultant re-test regression) can be seen as plausibly operating in our case in two ways: Firstly, among countries developing research policies in 1988, this incentive and evaluation framework was most likely to be applied in those with a very low scientific production or in those showing slowdown; secondly, the most likely time in which a new RES would be set up would be after a year in which production data were exceptionally low.

Figure 1. Spanish scientific publications: Total per year and annual growth, 1984-1994



Source: Own elaboration based on Thomson-Reuters publication databases (SCI, SSCI and A&HI)

As we can observe in figure 1 there was already a long trend of growth in the absolute number of Spanish publications which we consider a possible result of maturation and history processes.[20] We also identify a significant slowdown in the years that preceded the creation of the CNEAI (something associated to testing effects), a growing annual increase just before and after, and a big reduction in the annual growth only three years after the creation of the CNEAI, which could represent some saturation or diminishing returns effect.

In order to claim a positive impact of the CNEAI in the evolution of publications, Jiménez-Contreras et al. (2003) constructed a regression (with the publication data from 1974-1989), estimated the coefficient and then compared the prediction with the real data for 1990 to 1997. They found a difference of 7,000 more papers for 1997 than the "expected" figure if the trend to 1989 had continued over the nineties.[21]

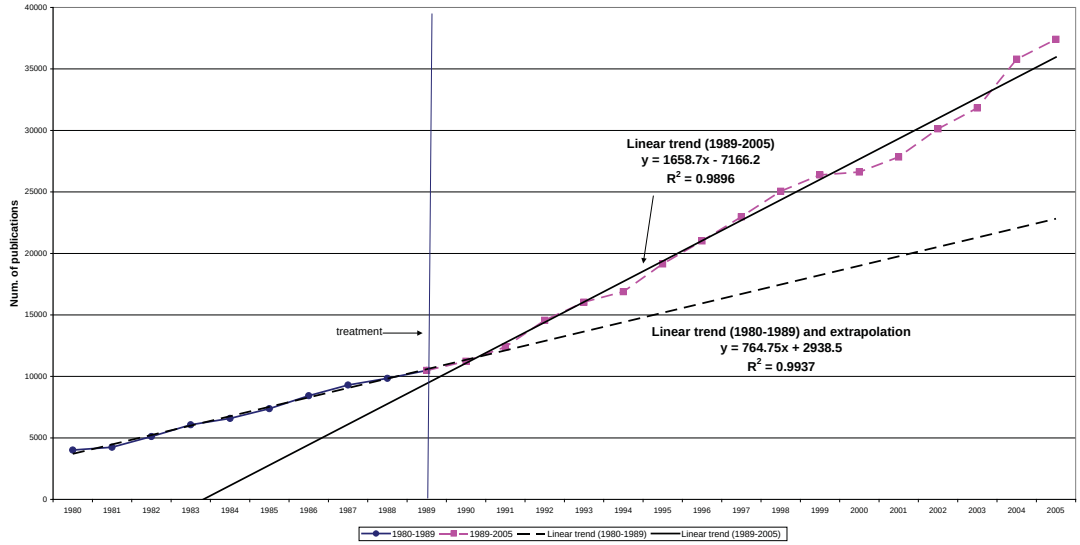
We argue that this simple method is insufficient to demonstrate that the setting up of the CNEAI caused the increase in scientific publications. In our opinion their analysis was affected by a typical case of regression artefacts. In figure 2 we show the evolution of Spain's annual scientific publications from 1980 to 2005. The number of papers increased from 4,015 in 1980 to 37,412 in 2005, a ninefold increase for the whole period. From a simple analysis we could conclude that the publications accelerated after 1989, however, if we compare the average

20 Just to make clear that it was not an instrumentation effect we have to mention that Spain had 0.62 % of the World share in 1979, 0.96% in 1984, 1.47% 1989 and 2.09% 1994 (Van Raan 1997).

21 See Figure 3 in Jiménez-Contreras et al. (2003).

values for 1985-1989 and 1990-1994, the differences are not so large. We have also elaborated two regression lines, adjusted for the period before and after 1989. The slope of the second period regression line is bigger than the one for the first period, meaning a significant change in the trend.

Figure 2. Spanish scientific publications: Total per year and trends, 1980-1989 and 1990-2005



Source: Own elaboration based on Thomson-Reuters publication databases (SCI, SSCI and A&HI)

Even if the slopes differ (to the extent that we could estimate 15,000 publications more for 2005), this evidence is not sufficient to claim causality, since the trend of increasing publications was already in place before the creation of the CNEAI. If one could claim anything, it would be just a stronger growth, an intensification/acceleration illustrated by the change in the slope, but we cannot disregard that those effects were a result of history, maturation, etc. The question is whether the acceleration of growth is explained by the impact of the RES or by other variables, as growth in resources.

6.2. Controlling other explanatory variables

The amount of resources and the number of researchers are very likely to affect the number of papers produced[22] and represent further alternative explanations to the growth of Spanish publications other than the impact of the RES.

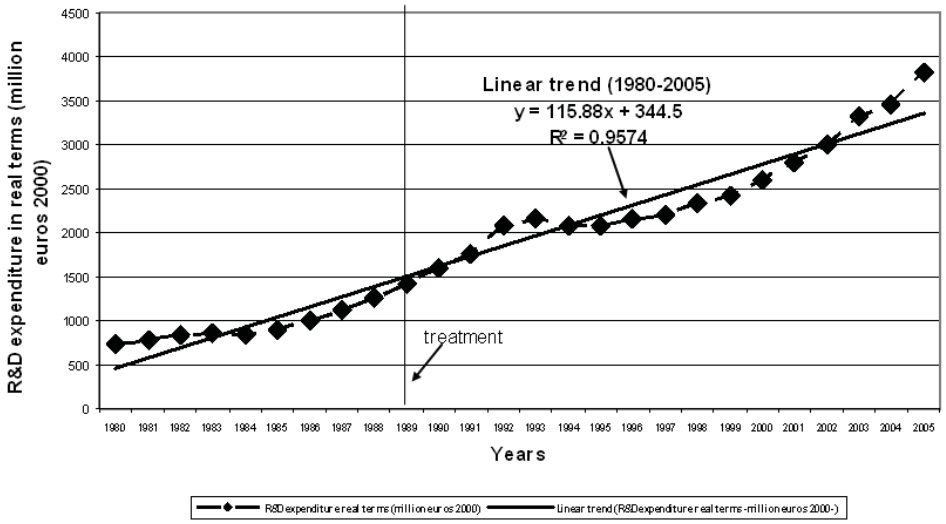
6.2.1. Investments in research (R&D)

In order to explore the evolution of resources available for research, in Figure 3 we build a long

22 This is even more so, in a context in which researchers have been increasingly socialised into publication norms for accessing tenure, accreditations and project funding (maturation effects).

time-series, from 1980 to 2005, and transform current R&D expenditures into constant prices (2000), by using implicit GDP price indices, to check for real growth and not just nominal growth. The figure shows the evolution of R&D expenditures in the public sector of research (government and higher education sectors). The period of highest growth rate in resources was 1985 to 1992, with an annual increase of more than 10% in real terms. Figure 3 represents the evolution and the linear regression; the equation presents an average annual increase of 115 million euros (in constant prices of 2000).

Figure 3. Spanish R&D expenditure in the public sector (million euros 2000 constant prices), 1980-2005



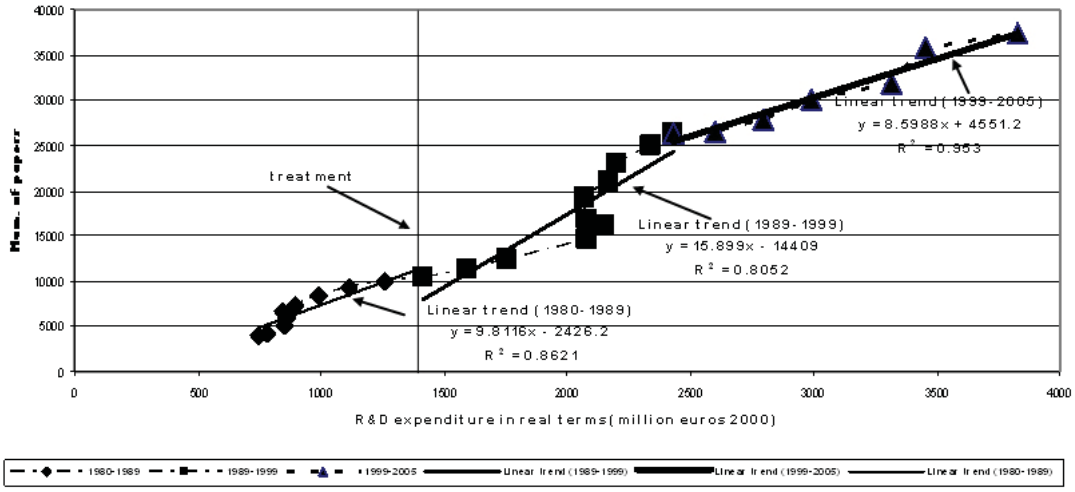
Source: Own elaboration based on National Statistical Office R&D data

Jiménez-Contreras et al. (2003) concluded that the reduction of the R&D expenditure as a share of the GDP and the lower rate of growth in the absolute R&D expenditure starting in 1993 could not account for the change and growth in the publication outputs from 1990. A more careful analysis of the data gives us the perception of the existence of some "lag effects" of the investment in R&D on the publication production (a history effect); some possible "delayed" effects of the extraordinary growth in economic resources in the 80s could have produced the growth of publications, even in a time of budgetary slowdown, to a greater degree than the RES.

Another step to control the explanation would be to test the direct relationship that exists between the two variables. If we regress, assuming linearity, the annual R&D expenditure in the public sector in real terms and the number of publications, we also find some interesting elements that undermine the assumption that the CNEAI was the cause of the growth of the latter. With two periods, before and after (1980-1989, 1989-2005), and comparing the slopes of both regressions, we observe a marginal increase in production for an additional unit of

resources. However, with 3 periods, in figure 4 we observe that in the 80s we could expect an increase of 9.8 publications for any additional million euros (year 2000) of R&D investment; whereas for the 90s, the figure is 15.8 additional publications. In the next period 1999-2005 we observe a very significant reduction (or decreasing returns).[23]

Figure 4. Spanish scientific publications and R&D expenditure in the public sector (million euros 2000 constant prices): Annual data and linear trends, 1980-1989, 1989-1999 and 1999-2005



Source: Own elaboration based on National Statistical Office R&D data and Thomson-Reuters publication data (SCI, SSCI and A&HI)

Despite the bigger slope in the 90s (which could be an "artefact" from the slowdown in R&D expenditures), we do not see big differences for the complete period 1980-2005.[24] Therefore, it is very plausible that the evolution of publications was structurally affected by growth in resources to a greater extent than it was by the creation of the CNEAI. Thus it seems necessary to analyse the association between papers and the evolution of the number of researchers, even if R&D expenditures and researchers have high correlation levels,[25] to control for the differences in the slopes before and after the treatment.

23 It is possible that the relationships between investments in R&D and publications are not linear over time; however for the sake of simplicity we continue using this type of analysis. There is evidence that after 2000, there was a decrease in returns in terms of papers for any additional investment or an increase in cost of every additional paper. This could be the result of the increasing costs of research or alternatively of decreasing returns and reduced efficiency.

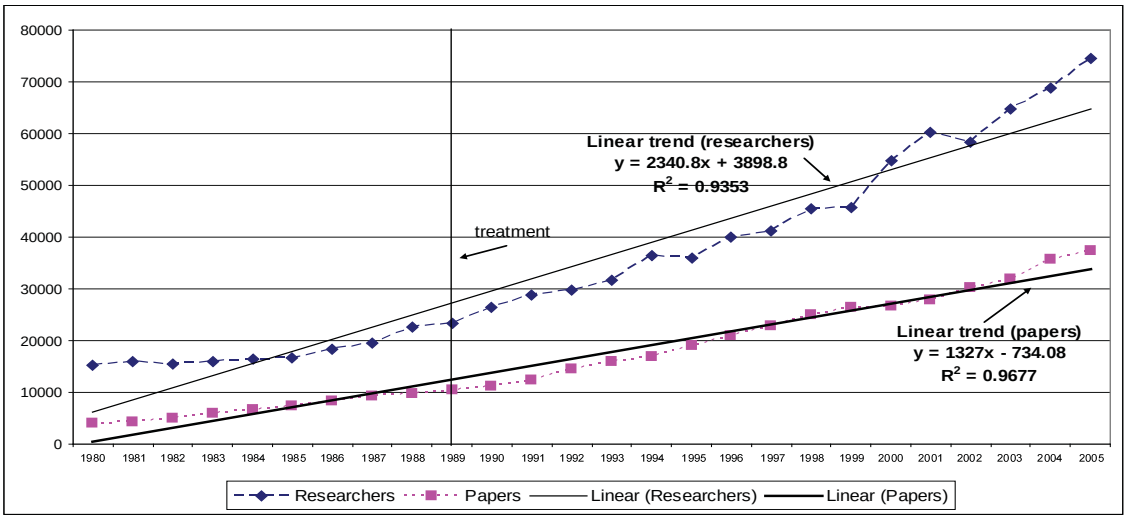
24 For the whole period 1980-2005 we find that an increase of one million euros of expenditure in real terms would produce an additional increase in publications of 11.2.

25 We should bear in mind that at the beginning of the 80s labour costs of researchers represented around 70% of the R&D expenditure, whereas later on, the figure came down to 55% of the total R&D expenditure, a change that favours the argument of increasing cost of research over time.

6.2.2. Researchers in the public sector

Although Jiménez-Contreras et al. (2003) acknowledged that between 1980 and 1998 the number of researchers in the public sector had tripled, surprisingly, they did not attribute the growth of publications to the growth in researchers and they preferred the explanation of the evaluation impact. In 1980 there were 15,300 public sector researchers and in 1989 there were 23,400, while in 2000 the figure was 54,700. That is, between 1989 (treatment year) and 2000 the stock of researchers in the public sector had more than doubled; and tripled when comparing with it 2005 (74,500). The growth in researchers largely exceeded the growth in the number of papers.

Figure 5. Spanish scientific publications and researchers in the public sector: Total per year and trends, 1980-2005



Source: Own elaboration based on National Statistical Office R&D data and Thomson-Reuters publications data (SCI, SSCI and A&HI)

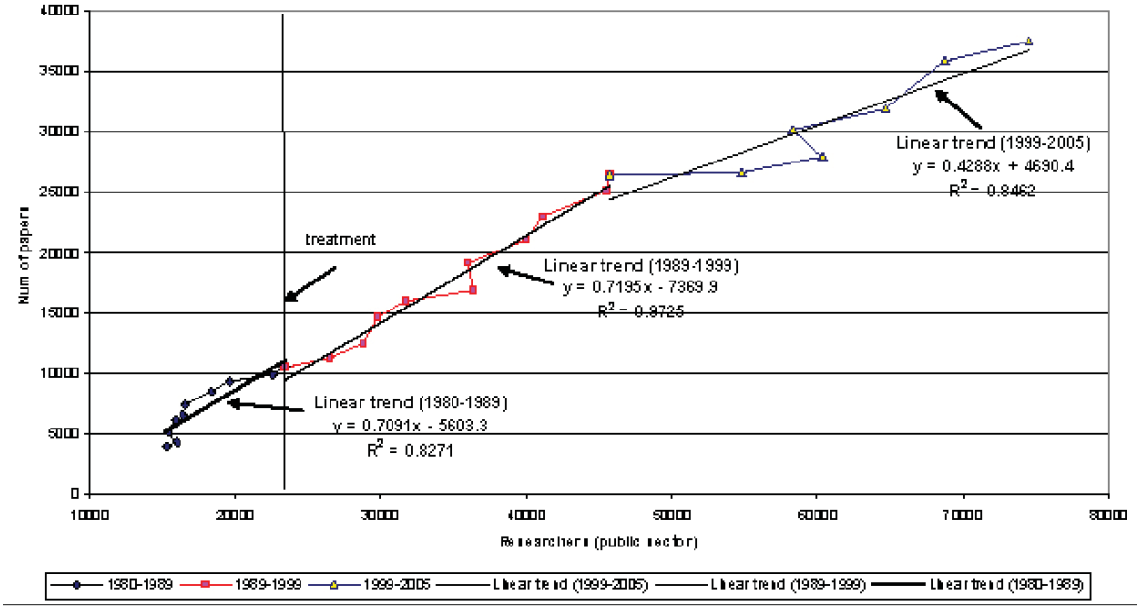
In figure 5 we compare the slopes of the two regression lines (researchers in public sector and publications), for the complete period 1980-2005, and we find that the former one is higher; that is, for any additional year estimated on a linear trend, there was an increase of 2,340 researchers but only of 1,327 additional papers.

We will now analyse the direct relationship between the evolution of papers and researchers. The correlation of the evolution of both variables over the whole period of time (assuming linearity) is very high (R^2 = 0.9774) and the equation gives us a stable estimate from which (without increases in productivity) we could expect an increase of 0.55 publications for any additional researcher.

In figure 6 we show the correlation between the number of public sector researchers and publications, before and after the creation of the CNEAI. The effect of any increase in one researcher in the 80s was only slightly lower than in the nineties, but the adjustment of the

regression was lower in the first period. Again we find evidence of decreasing returns in the next period (present decade).

Figure 6. Spanish scientific publications and researchers in the public sector: Annual data and linear trends, 1980-1989, 1989-1999 and 1999-2005



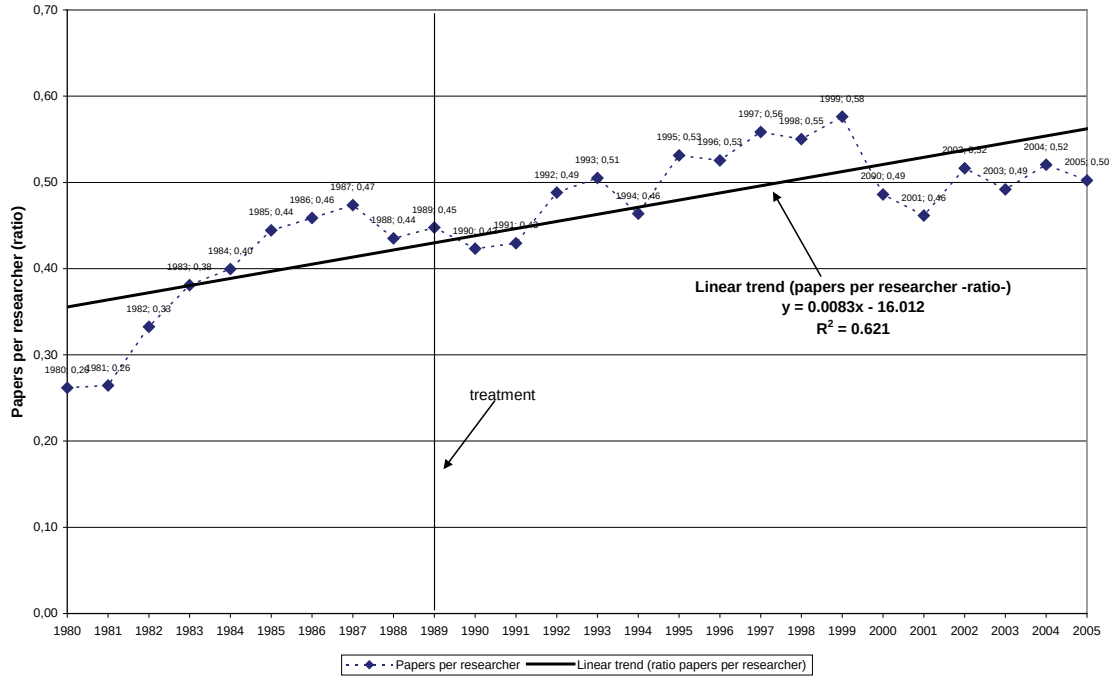
Source: Own elaboration based on National Statistical Office R&D data and Thomson-Reuters publications data (SCI, SSCI and A&HI)

Even without increases in the aggregate or individual levels of productivity (a possible behavioural effect of the CNEAI) we observe that, for any increase in the number of researchers in the 80s, there was an increase of 0.71 publications; while in the 90s, the expected increase was 0.72 publications. If one assumes that the CNEAI was the cause or had an effect, then a very small change in the coefficient represents a marginal effect.

In sum, the comparison of the evolution of papers with that of the number of researchers does not provide evidence to conclude that the growth in publications was the result of the CNEAI creation; instead, most of the growth in publications seem to have resulted from the growth in human resources for research.

If we relate the number of Spanish papers to the number of researchers, we get, as we can see in figure 7, a virtual indicator of the evolution of "productivity" of researchers (number of papers divided by number of public sector researchers) (Moya et al. 2007). It seems that there is not a clear linear annual trend, but significant fluctuations, that could be due simply to the changing patterns of publication and collaboration (more authors per paper and more international collaboration).

Figure 7. Spanish researchers' productivity (average ratio of papers per researcher), 1980-2005



Source: Own elaboration based on National Statistical Office (INE) and Thomson-Reuters data (SCI, SSCI and A&HI)

In our opinion, these results raise serious doubts about the effects of the setting up of the Spanish RES, because the highest productivity increase occurred just before 1989. All we can say from the figures is that researchers now in the system, whether tenured or not, whether affected by the RES or not, have become more productive (in the way we measure productivity) than the previous generation. Average productivity per researcher increased clearly between 1980 and 1987, and decreased slowly from 1987 to 1992. It then started to recover until 1999, when it returned to a downward trend. Observing the trend of this indicator, it could be argued that the effect of the CNEAI took a few years to impact on the average productivity values (lag effects), but even the attribution of this delayed growth to the CNEAI needs to be confirmed with control groups.

6.2.3. Endogenous issues of the measurement: patterns of co-authorship and collaboration (instrumentation effects)

Changes in the publication patterns over time could influence the results of measuring growth in Spanish scientific production. Most of those elements are related to two standard variables: a) the number of authors per paper, and b) international collaborations. Both factors are very relevant when applying whole counting at country level, but especially the latter.

Research collaboration and especially international collaboration has become a central

mechanism of socialisation of the national scientific communities into the international arena, with significant effects in the increase of national shares. Although we do not have data for the period before the establishment of the CNEAI, between 1990 and 2005, international collaboration experienced an enormous growth rate (89.9%), while national collaboration grew by 41.2% (Moya et al. 2007). Internationally co-authored papers of Spanish authors have gone from 18% in 1990 to more than 35% in 2005, meaning that those papers are also counted in other countries' records.

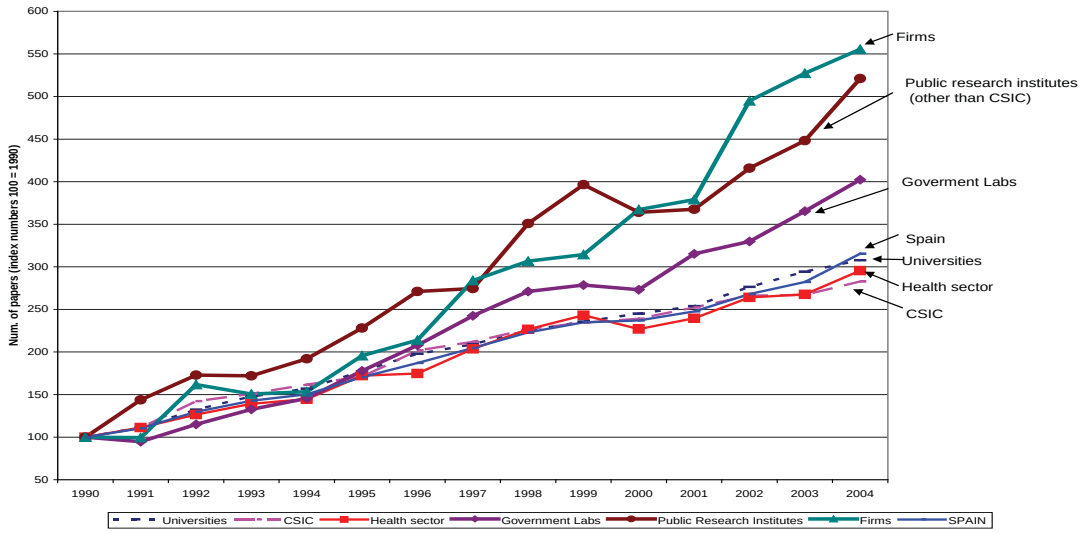
Thus, a hypothesis can be made that the endogenous change (international collaboration) and the way of counting (whole) could also plausibly account for the growth in Spanish publications. At best, the attribution effect to the CNEAI could be indirect, assuming that the collaboration was promoted indirectly by the CNEAI, as a result of the method of individual evaluation and the individual strategies to increase researchers' international publications. However, to claim this relation, we should demonstrate that the increase in the number of publications with international collaborations was specific to Spain and not just a universal phenomenon, as it happens to be (Georghiou 1998).

6.3. Non-equivalent control groups

A further element in a design aiming to confirm the claim that the RES caused the growth in publications would be to establish a "control group" among the population of public sector researchers. The control group should (by definition) be unaffected by the treatment (individual retrospective evaluation by the CNEAI).

We have differentiated the evolution of the publications, just after the treatment, among different sectors directly affected by the RES (universities and CSIC) and some other institutional sectors, for which there was no possibility of applying for the incentives (and benefits) of the CNEAI, but who were presumably subjected to the same external conditions of the Spanish research system, such as hospitals, other public research centres and government laboratories (excluding CSIC) and firms, and have used them as control groups. We acknowledge that the absolute contribution of the different sectors is very diverse, and that our "control groups" represent a small proportion of Spain's total output of papers. Nevertheless, what is important to note is the common trend of publication growth among sectors, indicating that researchers from other institutions, even those not affected by the incentives and the evaluation system, have also increased their production very significantly.

Figure 8. Spanish scientific publications by sector (index number 1990=100), 1990-2004



Source: Own elaboration based on Moya et al. (2007)

Figure 8 shows that in the evolution of the publications attributed to the different institutional sectors, there is not a positive comparative effect that could be attributed to the evaluation system in favour of universities and CSIC versus the "control groups" not directly affected by the RES incentive or rewards. Even the opposite seems to hold: the control groups appear to grow more than the "experimental" ones. In sum, the growth in those sectors was not significantly different to the growth in the institutional sectors affected by the CNEAI.

7. DISCUSSION AND CONCLUSIONS

In this paper we have analysed some plausible causal factors alternative to the effect of the research evaluation system on the evolution of Spanish scientific publications. Studies about the impact of evaluation systems face problems of causal attribution due to the difficulties to randomise the populations subjected or not to the evaluation. We have addressed these problems with a strategy of analysing other possible factors influencing the number of publications [research inputs (researchers and R&D investment) or endogenous changes in the dependent variable (changes in collaboration patterns)] in a set of interrupted time-series, and also compared the publication performance of different institutional sectors not subjected to the RES as non-equivalent control groups. We have extended the number of years of previous analyses, before and after the RES implementation, in order to consider the possibility of already existing trends in scientific publications.

The main results of our analysis confirm the difficulties of claiming the existence of a significant impact of the RES on the quantity of Spanish scientific publications at the aggregate level in the period analysed. There is no clear evidence that the researchers increased their average

productivity significantly in the years following the creation of the CNEAI, although there is evidence about Spanish transition in the nineties to English language and ISI coverage. The comparative evolution of publications by universities and CSIC with other institutional sectors as control groups does not present significant differences either. If any increase in the number of publications could be attributed to the RES it would be indirectly, through the increase in international collaboration patterns, but this endogenous change appears to be the result of a long existing trend of socialisation of our researchers and seems to be a global common trend in other countries, with and without RES.

Instead, our findings support the idea that the growth of scientific publications can be considered first the result of the growth in size of the system and, second, as the effect of the maturation and history processes of the Spanish research system, and the compounded effect of some endogenous changes, such as the expansion of international collaboration in publications.

From the evidence presented it appears that the claim that the CNEAI was the cause of the increase in the number of Spanish publications is too strong, and there are some plausible rival explanations. The CNEAI has probably played a role in the "standardisation" of the publication patterns according to the norms of international science and contributed to reinforcing a maturation process that led to a shift to English for scientific communication, but this is different to the claim that the CNEAI caused the growth.

However, some caveats should be acknowledged. Our analysis has followed a whole counting approach, with the measuring issues already commented associated to co-authorship and international collaboration patterns. Our analysis has also used aggregate data on publications: the lack of differences in the patterns of scientific communication by scientific areas cannot be taken for granted and should be empirically tested in further research, because evidence coming from other countries show that the different fields differ in their responses to common policy changes.[26]

It is well known that publication patterns differ among research fields; global data could hide the effects of the CNEAI. Although the Spanish RES was institutionalised in 1989, it did not formally apply the same evaluation criteria (international publications) and procedures for all areas until several years after. Only natural science researchers' commissions have applied international publication as simple quality proxy for the evaluation almost from the beginning. This is a problem which needs to be addressed in further research, due to the fact that researchers in social and economic sciences and arts and humanities represent around one third of the total in the public sector, although, in 2000, they only contributed 9% of the total number of Spanish publications in ISI (Cotec 2010).

The circumstances of implementation of RES in each national context must be taken into

26 E.g. Ingwersen and Jacobs (2004) for South Africa.

account. A comprehensive analysis of the effects of RES needs to consider research quality too. The impact of defining evaluation criteria without reference to quality or academic impact could produce unexpected results, as have been reported in Turkey, Australia or for UK's first RAE. The Spanish case seems somewhat different, because there are some references to the impact of journals in which researchers publish.[27] However, as with the number of papers, any increase in impact factor or citations in Spanish publications[28] would face the same causal attribution problem and would raise the same question as to whether quality improved as a direct result of the CNEAI, a question that also deserves further research.

It is common to distinguish "weak" from "strong" RES, considering their effects and their structural and financial consequences. It might be the case that the difficulties in attributing effects to the CNEAI are related to its "weak" characterisation or to its emphasis on individuals (Cruz-Castro and Sanz-Menéndez 2007). Researchers have to select only five contributions for each evaluated 6-year period, and with those small numbers we cannot expect large increases in the total amount of publications. Additionally, the fact that CNEAI is targeting individual researchers who already have a civil servant type permanent position might weaken the incentive. Moreover, there are few organisational pressures and incentives from employing organisations (CSIC and universities) that probably also interact with the former factors.

We believe our research has some policy implications. There is not a consensus in Spain as to whether or not RES that affect the immediate organisational context of researchers are more effective in increasing their productivity than those focused on the individuals themselves. There are also controversies regarding the appropriate methodologies (peer review versus bibliometric indicators) and their effect. It might well be the case that with individual researchers (with permanent status) as targets and with weak consequential effects –besides individual reputation-, we could not expect significant impacts in terms of scientific production, productivity and visibility.

It seems clear enough from our analysis that simplistic approaches such as before and after measures, often used by politicians to legitimise their "narratives", are not sufficient from a research point of view. However, further studies are necessary to control for other possible explanatory factors at the organisational or individual levels (such as diversity by areas or different research career stages). Further work based on micro data is needed to assess the real impact of the CNEAI on the Spanish researchers that were active in the early eighties. Longitudinal studies, such as a cohort analysis of that generation would be required to test the changes in terms of publications and to control attribution effects. Additional cohort

27 Recently Rodríguez-Navarro (2009) has argued that the Spanish evaluation system does not promote high quality or outstanding results, despite the fact that aggregate average impact factor of Spain has improved over the last years up to the average world (Gómez et al. 2006).

28 Citations of Spanish papers follow similar trends to those of other countries (Italy, Australia, UK, France, etc.). See Web of Science, Thomson-Reuters ISI Essential Science Indicators at http://in-cites.com/countries/index.html; access 10 May 2010.

research would be useful to identify the socialisation effects comparing the behaviour of two generations.

Lastly, a question arises regarding the effectiveness to produce effects of a policy or mechanism that has remained stable for 20 years. If the research evaluation system ever had any behavioural effects these effects have now ceased. As several studies for different countries have shown, "pervasion effects" and "learning effects" occur as response to "controls and incentives" set up by RES, independently of the methodologies and organisational models. Rethinking the CNEAI functioning and its objectives seems necessary. Incentives are more than the "bureaucratic" counting of the accomplishments of quantitative objectives. RES should contribute to the general long term improvement of the system, to the promotion of talent and to improve creativity (*highly productive environments*), and not just play the role of performance controls of a bureaucratic profession. Perhaps it is also time for a radical change, moving from a weak to a strong RES with significant impacts on funding and more focused on organisations. In this context there might be more appropriate ways to improve performance, as the experience of new research centres recently created, not based on a civil servant model, have demonstrated (Cruz-Castro et al. 2010).

ACKNOWLEDGEMENTS

This research has been funded by the Ministry of Science and Innovation (CSO-2008-03100/SOCI) and the AECID (D-026992/09). We acknowledge the contribution of Zaida Chinchilla and Félix de Moya for providing us with the annual publication data. A previous version of this paper was presented at the International Sociological Association (ISA) Conference (Research Committee 23) in Barcelona (September 2008); we acknowledge the comments, criticisms and suggestions from participants, especially those of Jochen Gläser and Maria Nedeva. Additional suggestions and criticisms from Félix de Moya and Jochen Gläser helped to improve the final version are acknowledged.

REFERENCES

1. Adams, J., Griliches, Z. (1996). Measuring Science: An exploration. *Proceedings National Academy of Sciences*, 93, 12664-12670.
2. Adams, J.D., Black, G.C., Clemmons, J.R., Stephan, P.E. (2005). Scientific teams and institutional collaboration: evidence from U.S universities. *Research Policy*, 34(3), 259-285.
3. Andersen, L.B., Pallesen, T. (2008). "Not Just for the Money?" How Financial Incentives Affect the Number of Publications at Danish Research Institutions. *International Public Management Journal*, 11(1), 28–47.
4. Ballart, X., Subirats, J. (1997). Science and technology policy for a medium-sized industrial country: the case of Spain. *Science and Public Policy* 24(3), 197-205.
5. Barker, K. (2007). The UK Research Assessment Exercise: the evolution of a national research evaluation system. *Research Evaluation*, 16(1), 3-12.
6. Bellavista, J., Escribano, L., Grabulós, M., Viladiu, C., Guardiola, E. & Iglesias, C. (1993). *Política científica y tecnológica. Evaluación del I+D de la Universitat de Barcelona*. Barcelona, Universitat de Barcelona.
7. Braun, T., Schubert, A. (1988). World flash on basic research - Scientometric versus socio-economic indicators. Scatter plots for 51 countries. 1978–1980. *Scientometrics*, 13(1-2), 1-9.
8. Braun, T., Glänzel, W., & Schubert, A. (1985). *Scientometrics Indicators. A 32-Country Comparative Evaluation of Publishing performance and Citation Impact*. Singapore-Philadelphia, World Scientific.
9. Butler, L. (2002). A list of published papers is no measure of value. *Nature*, 419, (31 October 2002), 877.
10. Butler, L. (2003a). Explaining Australia`s increased share of ISI publications. The effects of a funding formula based on publication counts. *Research Policy,* 32(1), 143-155.
11. Butler, L. (2003b). Modifying Publication Practices in response to Funding Formulas. *Research Evaluation*, 12(1), 39-46.
12. Butler, L. (2004). What Happens When Funding is Linked to Publication Counts?. In H.F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research (*pp. 389-405). Dordrecht (NL), Kluwer Academic Publisher.
13. Campbell, D.F.J. (2003). The evaluation of university research in the United Kingdom and the Netherlands, Germany and Austria. In Ph. Shapira & S. Kuhlman (Eds), *Learning from Science and Technology Policy Evaluation* (pp.98-131). Cheltenham (UK), Edward Elgar.
14. Campbell, D.T. (1969). Reforms as experiments. *American Psychologist* 24 (4) , 409-429.
15. Campbell, D.T. (1996). Regression artifacts in time-series and longitudinal data. *Evaluation and Program Planning* 19(4), 377-389.
16. Campbell, D.T., Ross, H.L. (1968). The Connecticut crackdown on speeding. Time-series data in quasi-experimental analysis. *Law and Society Review*, 3(1), 32-54.
17. Campbell, D.T., & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, Rand McNally.
18. Chinchillla, Z ., & Moya-Anegón, F. (2007). *La investigación científica española (1995-2002): Una aproximación métrica*. Granada, Editorial Universidad de Granada.
19. CNEAI (2006a). *Memoria de los resultados de las evaluaciones realizadas 1989-2005 (científicos del CSIC).* Madrid, CNEAI, 35 p., mimeo. http://www.educacion.es/dctm/ministerio/horizontales/ministerio/organismos/cneai/2005-memoria-1989-2005-csic.pdf?documentId=0901e72b8008d9e2. Accessed 10 May 2010.
20. CNEAI (2006b). *Memoria de los resultados de las evaluaciones realziadas 1989-2005 (profesores de universidad).* Madrid, CNEAI, 35 p., mimeo. http://www.educacion.es/dctm/ministerio/horizontales/ministerio/organismos/cneai/2005-memoria-1989-2005-universidad.pdf?documentId=0901e72b8008d9f6. Accessed 10 May 2010.
21. Cook, T.D., & Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin Co.

22. Cotec (2010). *Informe Cotec 2010. Tecnología eIinnovación en España*. Madrid. Fundación Cotec.

23. Crespi, G.A., Geuna, A. (2008). An empirical study of scientific production: A cross country analysis (1981-2002). *Research Policy*, 37(4), 565-579.

24. Cruz-Castro, L., Sanz-Menéndez, L. (2007), Research Evaluation in transition: individual versus organisational assessment in Spain. In R. Whitley, & J. Glaser (Eds.), *The changing governance of the sciences. The advent of the Research evaluation systems* (pp. 205-224). Dordrecht (NL), Springer.

25. Cruz-Castro, L., Sanz-Menendez, L. & Martínez, C. (2010), Research centers in transition: patterns of convergence and Diversity. *The Journal of Technology Transfer,* doi: 10.1007/s10961-010-9168-5 (forthcoming).

26. Fox, M.F. (1983). Publication productivity among scientists. A critical review. *Social Studies of Science*, 13(2), 285-305.

27. Georghiou, L. (1998). Global cooperation in research. *Research Policy*, 27(6), 611–626.

28. Georghiou, L., Howells, J., Rigby, J. Glynn, S., Butler, J., Cameron, H., Cunningham, P., Ausadamongkol, K., Thomas, D., Salazar, A., Barker, K., & Reeve, N. (2000). *Impact of the Research Assessment Exercise and the Future of Quality Assurance in the Light of Changes in the Research Landscape.* A report produced by PREST, University of Manchester for HEFCE Higher Education Funding Council for England. April 2000. http://www.mbs.ac.uk/research/innovation/publications-archive/reports.aspx. Accessed 10 May 2010.

29. Geuna, A. Martin, B. (2003). University Research Evaluation and Funding: An International Comparison. *Minerva*, 41(4), 277-304.

30. Glänzel, W. (2001). National characteristics in international scientific co-authorship relations, *Scientometrics*, 51(1), 69-115.

31. Gläser, J. (2007). The Social Orders of Research Evaluation Systems. In R. Whitley, & J. Glaser (Eds.), *The changing governance of the sciences. The advent of the Research evaluation systems* (pp. 245-266). Dordrecht (NL), Springer

32. Gläser, J., Laudel, G. (2007). Evalaution Whitout Evalautors: The Impact of Funding Formulae on Australian University research. In R. Whitley, & J. Glaser (Eds.), *The changing governance of the sciences. The advent of the Research evaluation systems* (pp. 127-151). Dordrecht (NL), Springer

33. Gómez, I., Fernández, M.T., Zulueta, M.A., Camí, J. (1995). Analysis of biomedical research in Spain. *Research Policy*, 24 (3), 459-471.

34. Gómez, I., Sancho, R., Bordons, M., Fernández, M.T. (2006). La I+D en España a través de publicaciones y patentes. In J. Sebastián & E. Muñoz (Eds.), *Radiografía de la investigación pública en España* (pp. 275-302). Madrid, Biblioteca Nueva.

35. Grupo Scimago (2006). Producción ISI y tramos de investigación: cómo combinarlos en un nuevo indicador. *El profesional de la información*, 15(3), 227-228.

36. Grupo Scimago (2007). Producción ISI y tramos de investigación: cómo combinarlos en un nuevo indicador (II). *El profesional de la información*, 16(4), 510-511.

37. Harley, S., Lee, F.S. (1997). Research Selectivity, Managerialism, and the Academic Labor Process: The Future of Nonmainstream Economics in U.K. Universities. *Human Relations*, 50(11), 1427-1460.

38. Henkel, M. (1999). The Modernisation of Research Evaluation: The case of the UK. *Higher Education*, 38(1), 105-122.

39. Hicks, D. (2009). Evolving regimes of multi-university research evaluation. *Higher Education*, 57(4), 393-404.

40. Ingwersen, P., Jacobs, D. (2004). South African research in selected scientific areas: Status 1981-2000. *Scientometrics*, 59(3), 405-423.

41. Jiménez Contreras, E. (1997). *Universidad de Granada: 1975-1987. La transición científica*. Granada, Universidad de Granada.

42. Jiménez-Contreras, E., Ferreiro-Aláez, L. (1996). Publishing abroad: fair trade or short sell for non-english-speaking authors: A Spanish Study. *Scientometrics*, 36(1), 81-95.

43. Jiménez-Contreras, E., Moya-Anegón, F.; López-Cózar, E. (2003). The Evolution of Research Activity in Spain. The impact of National Commission for the Evaluation of Research Activity (CNEAI). *Research Policy*, 32(1), 123-142.

44. Katz, J., Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric Model. *Scientometric*, 40(3), 541-554.

45. King, D.A. (2004). The scientific impact of nations. What different countries get for their research spending. *Nature*, 430 (15 July, 2004), 311-316.

46. Laudel, G. (2006a). The 'Quality Myth': Promoting and Hindering Conditions for Acquiring Research Funds. *Higher Education*, 52(3), 375-403.

47. Laudel, G. (2006b), The Art of Getting Funding: How Scientists Adapt to their Funding Conditions. *Science and Public Policy*, 33(7), 489-504.

48. Lee, S., Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5), 673-702.

49. Leydesdorff, L., Wagner, C.S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, 2(4), 317-325.

50. Liefner, I. (2003). Funding, Resources Allocation, and Performance in Higher Education Systems. *Higher Education*, 46(4), 469-489.

51. May, R.M. (1997). The Scientific Wealth of Nations. *Science*, 275(5301) (Feb. 7, 1997), 793-796.

52. McNay, I. (1997). *The Impact of the 1992 Research Assessment exercise on Individual and Institutional Behaviour in English Higher Education: Summary Report and Commentary.* Chelmsford: Anglia Polytechnic University.

53. Méndez, A., Gómez, I. (1986). The Spanish scientific productivity through eight international databases. *Scientometrics*, 10(3-4), 207-219.

54. Moed, H. (2008). UK Research Assessment Exercises: Informed Judgements on Research Quality or Quantity?. *Scientometric*, 74(1), 153-161.

55. Moed, H.F., Debruin, R.E., Van Leeuwen, T.N. (1995). New bibliometric tools for the assessment of national research performance –data base description, overview of indicators and first applications. *Scientometrics*, 33(3), 381-422.

56. Moed, H.F.; Leeuwen, T.N., Visser, M.S. (1999). Trends in Publication Output and Impact of Universities in the Netherlands. *Research Evaluation*, 8(1), 60-67.

57. Mohr, L.B. (2000). Regression artifacts and other customs of dubious desert. *Evaluation and Program Planning* 23(4), 397-409.

58. Moya-Anegón, F., Herrero-Solana, V. (1999). Science in America Latina: A comparison of bibliometric and scientific-technical indicators. *Scientometrics*, 46(2), 299-320.

59. Moya-Anegón, F. et al. (2007). *Indicadores Bibliométricos de la Actividad Científica Española (1990-2004).* Madrid: FECYT.

60. Muñoz, E. & Ornia, F. (Eds) (1986). *Ciencia y Tecnología: una oportunidad para España*. Madrid, Aguilar.

61. Önder, C., Sevkli, M., Altinok, T., Tavukçuoglu, C. (2008). Institutional change and scientific research: A preliminary bibliometric analysis of institutional influences on Turkey's recent social science publications. *Scientometrics*, 76(3), 543-560.

62. Persson, O., Glänzel, W., Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421-432.

63. Rip, A., Van der Meulen, B.J.R. (1996). The post-modern research system. *Science and Public Policy*, 23 (6), 343-352.

64. Rodriguez-Navarro, A. (2009). Sound Research, Unimportant Discoveries: Research, Universities, and Formal Evaluation of Research in Spain. *Journal of the American Society for Information Science and Technology,* 60(9) 1845-1858.

65. Sánchez-Ferrer, L. (1997). From Bureaucratic Centralisation to Self-Regulation: The Reform of Higher Education in Spain. *West European Politics*, 20(3), 164-184.

66. Sanz-Menéndez, L. (1995a). Policy choices, institutional constraints and policy learning: The Spanish science and technology policy in the eighties". *International Journal of Technology Management*, 10 (4/5/6), 255-274.

67. Sanz-Menéndez, L. (1995b). Research actors and the state: research evaluation and evaluation of science and technology policies in Spain. *Research Evaluation*, 5(1), 79-88.

68. Sanz Menéndez, L. (1997). *Estado, ciencia y tecnología en España (1939-1997)*. Madrid: Alianza Editorial.

69. Sanz Menéndez, L., Muñoz, E., Garcia, C. E. (1993). The vicissitudes of Spanish science and technology policy. *Science and Public Policy*, 20(6), 370-380.

70. Sanz Menéndez L., Pfretzchner, J. (1992). Política científica y gestión de la investigación: El CSIC (1986-1990) en el sistema español de ciencia y tecnología. *Arbor*, 557 (Mayo), 9-51.

71. Schubert, A., Glänzel, W. Braun, T. (1989). Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981–1985. *Scientometrics*, 16(1-6), 3-478.

72. Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston-New York, Houghton Mifflin Co.

73. Shelton, R.D. (2008). Relations between national research investment and publication output: Application to an American paradox. *Scientometrics*, 74(2), 191-205.

74. Talib, A.A. (2001). The Continuing Behavioural Modification of Academics since the 1992 Research Assessment Exercise. *Higher Education Review*, 33, 30-46.

75. Taylor, J. (2001). The impact of performance indicators on the work of university academics: evidence from Australian universities. *Higher Education Quarterly*, 55(1), 42-61

76. Van Raan, A.F.J. (1997). Science as an international enterprise. *Science and Public Policy*, 24(5), 290-300.

77. Vinkler, P. (2008). Correlation between the structure of scientific research, scientometric indicators and GDP in EU and non-EU countries. *Scientometrics*, 74(2), 237-254.

78. Westerheijden, D. (1997). A Solid Base for Decisions: Use of the VSNU research Evaluation in Dutch Universities, *Higher Education*, 33(4), 397-413.

79. Whitley, R. (2003). Competition and pluralism in public sciences: the impact of institutional frameworks on the organization of academic science. *Research Policy*,32(6), 1015-1039.

80. Whitley, R. (2007). The changing governance of the public sciences: the consequences of research evaluation systems for Knowledge production in different countries and scientific fields". In R. Whitley, & J. Glaser (Eds.), *The changing governance of the sciences. The advent of the Research evaluation systems* (pp. 3-27). Dordrecht (NL), Springer..

81. Whitley, R., & Glaser, J. (Eds.) (2007). *The changing governance of the sciences. The advent of the Research evaluation systems*. Dordrecht (NL), Springer.

82. Zitt, M., Perrot, F., Barré, R. (1998). The Transition from "National" to "transnational" Model and Related Measures of Countries' Performance. *Journal of the American Society of Information Science*, 49(1), 30-42.

Working Papers published in:
Instituto de Políticas y Bienes Públicos (IPP), 2010

1. **Cruz Castro, L. & Sanz Menéndez, L.** Endogamia, Productividad y Carreras Académicas.

2. **Corrochano, D.** Guía Bibliográfica sobre Inmigración en España (1990-2009). Datos y Reflexiones sobre la Institucionalización de una Comunidad Académica.

3. **Golob, S.R**. Evolution or Revolution? Transitional Justice Culture Across Borders.

4. **Arias Aparicio, F.** Organización y Producción del Conocimiento Científico en los Organismos Públicos de Investigación Agraria: El Instituto Nacional de Investigación Agraria y Alimentaria (INIA).

5. **Moreno, L.** Welfare Mix, CSR and Social Citizenship.

6. **Martínez, C. & Rama, R.** The control and generation of technology in European food and beverage multinationals.

7. **Del Pino, E. & Colino, C.** National and Subnational Democracy in Spain: History, Models and Challenges.

8. **Closa, C.** Negotiating the Past: Claims for Recognition and Policies of Memory in the EU.

9. **Martínez, C., Cruz Castro, L. & Sanz Menéndez, L**. Convergencia y diversidad en los centros de I+D.

10. **Osuna, C., Cruz Castro, L. & Sanz Menéndez, L**. Knocking down some assumptions about the effects of evaluation systems on publications.