INGENIO WORKING PAPER SERIES

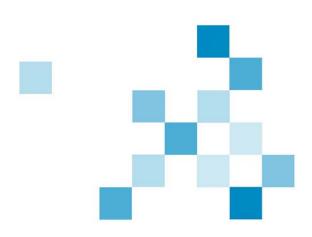


INSTITUTO DE GESTIÓN DE LA INNOVACIÓN Y DEL CONOCIMIENTO

Smoothing the lies: Do patent examiners take notice if applicants include citations?

Joaquín M. Azagra-Caro, Pauline Mattsson, François Perruchas

Working Paper Nº 2010/05



Smoothing The Lies: Do Patent Examiners Take Notice If Applicants Include Citations?

Joaquín M. Azagra-Caro^{a,b,1}, Pauline Mattsson^c, François Perruchas^d

^a European Commission, Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS), Edificio Expo, C/Inca Garcilaso 3, E-41092 Sevilla, Spain²

^b CSIC - Consejo Superior de Investigaciones Científicas, Institute of Public Goods and Policies (IPP-CCHS), C/Albasanz 26-28, E-28037 Madrid, Spain. E-mail: joaquinmaria.azagra@cchs.csic.es

^c Unit for Bioentrepreneurship (UBE) - Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institute, SE-171 77 Stockholm, Sweden. E-mail: pauline.mattsson@gmail.com

^d INGENIO (CSIC-UPV), Polytechnic University of Valencia, Edif. 8E 4°, Camino de Vera s/n, Valencia, Spain. E-mail: franperr@ingenio.upv.es

¹ Corresponding author at: CSIC - Consejo Superior de Investigaciones Científicas, Institute of Public Goods and Policies (IPP-CCHS), C/Albasanz 26-28, E-28037 Madrid, Spain. Tel.: +34 916 022 370. Email address: joaquinmaria.azagra@cchs.csic.es.

² The views expressed in this article are those of the authors; they do not necessarily reflect those of the European Commission (EC). Neither the EC nor anyone acting on behalf of the EC can be held responsible for the use that might be made of the information.

Abstract

Examiner patent citations are a popular source of indicators of technological impact and knowledge flows, despite various critiques. We analyse the distribution of examiner patent citations according to patent characteristics, to show their comparative meaningless. Our findings show that it is the science-base of the technology that determines the inclusion of applicant citations. However, this gets masked by the citations added by patent examiners, who smooth the distribution of citations across technology classes and include the 'standard' knowledge bases regardless of which references applicants cite. Some researchers have called for the use of applicant rather than examiner patent citations to build indicators of technology impact and knowledge flows. However, we show that the former are not necessarily 'better' than the latter, because applicants may 'inflate' the numbers in international patents especially when there are co-applicants. The implications are that analysts should consider alternative uses of patent citations e.g. to build indicators of trust within a research system.

1 Introduction

'Generation, diffusion and exploitation of knowledge are at the core of the research system' (EC, 2007: 16). This official recognition of a social interest underlies much academic research on the origins and the destiny of the knowledge produced. When the focus is on technological applications, key data for quantitative analyses are patent citations. Explaining the state-of-the-art requires some differentiations, starting with this:

according to direction of the citation, we can classify patent citations as *forward* or *backward* citations.

The use of forward citations is customary to express the technological impact of the patented invention (Noma & Olivastro, 1984), often as a function of the characteristics of the patent (Allison & Sager, 2007). The use of backward citations is customary to express the knowledge base of the patented invention. This introduces another important distinction:

according to the type of cited document, we can classify patent citations as patent-to-

patent or patent-to-paper citations (also known as patent vs non-patent references or literature).

Patent-to-patent citations are the most frequent and often serve as a proxy for the whole knowledge base of the invention, more properly specified as the technological knowledge base. Applications include analysis of whether they are from the same country as the applicant, which measures geographical spillovers (Jaffe & Trajtenberg, 1996). Patent-to-paper citations frequently serve as proxies for the scientific knowledge base (Narin & Noma, 1985; Hassan, 2003; Leydesdorff, 2004), usually leading to some justification of the importance of science, or at least of some cutting-edge technologies.

As Chen (2003) accurately notes, patent-to-patent backward citations are used by economists to explore knowledge spillovers, while their wider application - especially, of patent-to-paper citations – is generally in scientometrics. However, classical works use similar wording to suggest a causal effect from citation to patent: 'knowledge diffusion', 'utility of basic research to technology', etc.

The increasing use of these techniques among researchers has developed in parallel with advocates of their application to justify research funding (Kostoff, 1994). However, some qualitative and case studies recommend caution in the interpretation of results, based on another distinction among patent citations:

according to who inserted the citation, we can classify patent citations as *examiner* or *applicant* citations (the latter are also, somewhat improperly, referred to as inventor citations).

The traditional studies on patent citations rely on patent examiner citations that appear on the front pages of patent documents. Their use introduces two main problems (see e.g. Meyer, 2000; Michel & Bettels, 2001). First, examiner citations may provide biased information about knowledge flows, since numbers might vary for administrative reasons. Second, unlike the US, the patent system in Europe does not compel patent applicants to disclose complete information in patent documents, making it much more probable that patents examiners will add citations.

These critiques have not deterred traditional quantitative studies, perhaps because qualitative evidence is not sufficiently convincing. And some academics are claiming that patent citations are useful because they are more credible than paper citations (Lai & Wu, 2005) and that more efforts should be devoted to producing better-codified data

on patent citations (Stock & Stock, 2006) to facilitate its use. It has been proposed that examiner forward citations should be used to build indicators such as h-indexes of firms' technological performance (Guan & Gao, 2008).

The first contribution of this paper is to provide quantitative evidence of possible inconsistencies in examiner citations that should prevent from extended use.

Nevertheless, the critiques made about the use of patent citations have had some consequences and have inspired several quantitative research lines. The first is to adopt more careful wording to refer to the relation between patents and their citations: 'interactions', 'links' or 'linkages', 'vicinity', etc. (Tijssen, 2001; Callaert et al., 2006). A second is to use patent data in alternative ways to visualise relations within the knowledge base, e.g. through co-classification in technology classes to show perhaps that countries are not an appropriate unit of observation (Leydesdorff, 2008). A third is to promote a quantitative approach to the difference between examiner and applicant citations. This work is confirming that the use of examiner citations biases the interpretation of findings, for instance, because the knowledge base appears to be more localised if measured through applicant citations (Criscuolo & Verspagen, 2008). The degree of localisation and of differences between examiner and applicant citations depend most likely on the absorptive capacity of the region (Azagra et al., 2009).

This last stream of research pleads for the use of applicant rather than examiner citations as a better expression of knowledge flows, and links directly to the second contribution of the present paper, which is to establish whether applicant citations are a better indicator of knowledge flows.

The next section builds a conceptual framework for the distribution of patent citations according to the characteristics of patents to help understand the weaknesses of examiner and applicant citations. Using testable hypotheses, having described the research context, we present the data and methodology, our results and some conclusions.

2 Some insights into the meaning of examiner and applicant patent citations

2.1 The distribution of examiner backward citations by patent characteristics

We have reported that: (i) quantitative studies on examiner forward patent citations explore how they vary in terms of patent characteristics; (ii) qualitative studies suggest that patent characteristics influence the inclusion of examiner backward patent citations. These two aspects justify quantitative study of the distribution of examiner backward citations according to the characteristics of the patents. However, earlier investigations have been incidental, and provides descriptive statistics rather than substantial analysis. Which characteristics are included? So far, the focus has been limited mainly to time effects and technology classes.

In terms of time effects, they seem to be influential in the distribution of examiner citations. US examiner time constraints have become tighter, with the increase in the number of patent applications outpacing increases in the number of examiners (Merrill et al., 2004). Time and resource constraints necessarily limit the comprehensiveness of examiners' prior art search, producing 'citation inflation' in US Patent and Trademark Office (USPTO) patents (Hall et al., 2000). Callaert et al. (2006) confirm an increase in USPTO citations between 1991 and 2001, while for European Patent Office (EPO) patents the average number of examiner citations per patent in the same period has been constant (or even slightly decreasing).

In terms of technology classes, Callaert et al. (2006) find a larger number of EPO examiner citations in patent-to-patent citations in Mechanical Engineering and Machinery and patent-to-paper citations in Chemistry and Pharmaceuticals. In the USPTO, Process Engineering and Social Equipment ranks first for patent-to-patent citations and Chemistry and Pharmaceuticals for patent-to-paper citations. Sampat (2004) finds that examiners in the US face particular challenges in identifying prior art in emerging technological fields, focusing on nanotechnology.

Although not touched on in descriptive analysis, there are some other characteristics of patents that may condition the distribution of citations.

First, given the influence of technology classes, it could be argued that related dimensions, such as geographic region or economic sector, may be influential, i.e. the more specialised regions and sectors are in technologies where patent examiners insert more citations, the higher will be the number of examiner citations in these regions and sectors.

Second, there may be administrative reasons for the inflated number of citations. For instance, protection means and ownership regimes.

In relation to means of protection, many works consider only one protection alternative, national (USPTO) or international (EPO), probably because of the major focus on the US and Europe generally. However, when studying a single European country or region, several alternatives may be relevant, national and international, because of the home advantage effects (Criscuolo, 2006) and, also, alternative routes may be indicators of geographical patterns or technological protection (Azagra et al., 2006). Moreover, 'any non-national language documents are only cited when already having been quoted by the applicant' (Michel & Bettels, 2001: 198). Callaert et al. (2006) show that USPTO examiners include a higher number of citations per patent than do EPO examiners. The reason for this for could be that USPTO applicants have the legal obligation to provide full lists of prior art which tend to accepted by examiners, whereas there is no such obligation in the case of the EPO (Michel & Bettels, 2001).

Regarding the ownership regime, although there is little evidence about the impact of co-ownership on the number of citations in patents, it could be argued to be influential, since institutions may differ in their tendency to include citations. For instance, universities and public research organisations may show a higher propensity for citation, so co-applications involving firms and these institutions may be more likely to incorporate prior art than patent applications from firms alone.³

It is useful to synthesise this information in the following hypothesis:

Hypothesis 1. Patent characteristics play an influential role in the distribution of examiner citations.

7

³ As for forward citations, Jaffe & Trajtenberg (1996) have shown that universities on average receive more citations, followed by corporations and government institutions.

2.2 Front-page vs full-text citations: Examiners trust applicants

Another common source of criticism of information derived from examiner patent citations is that it is the patent applicants that really know the base of the invention (Jaffe et al., 2000). However, the codification of applicant citations is lagging behind that of examiner citations, making quantitative research on the former less easy. It is only recently that quantitative studies have begun to investigate examiner and applicant added citations separately, and this is due to improvements in computation facilities for identifying citations on front pages of patent documents (Sampat, 2004; Thompson, 2006; Alcácer & Gittelman, 2006; Criscuolo & Verspagen, 2008).

However, following the economic tradition of measuring knowledge spillovers referred to in the introduction, most of these studies focus on citing-cited pairs. They analyse localisation effects, regardless of whether the citations are from examiners or applicants. They are rarely interested in differences in patent characteristics although such differences are relevant, since they may be a basic reason for the different sets of patents included by applicants and examiners.

Another characteristic of these studies is that they look only at applicant citations *that* the examiner considers relevant, i.e. those on the front pages. However, the applicant citations which are included in the patent text, may be much closer to the knowledge base, since they have escaped the examiners' editing. Because of the time consuming process of retrieving the information, only a few studies consider this aspect (see exceptions such as Acosta & Coronado, 2003)⁴

So, there do exist some results related to the influence of patent characteristics on applicant citations. For instance, regarding time effects, Criscuolo & Verspagen (2008) find that the share of applicant citations in EPO patents decreased between 1985 and 1999. In terms of route of protection, the same authors conceptually justify the fact that both examiners and applicants tend to add more citations in patents applied for through the Patent Cooperation Treaty (PCT)-EPO procedure, compared to the direct-EPO procedure. The evidence from their econometric tests is inconclusive about whether this affects the probability of examiner-added citations.

8

⁴ Acosta & Coronado (2003) analyse full-text applicant citations. However, they add (Acosta & Coronado, 2003: 1794) that they also include examiner citations, which is rather confusing. What is clear is that the weight of full-text citations in their study is higher than in many other studies.

Technoeconomic characteristics (region, technology, economic sector) are also worthy of attention. Focusing on regions, Acosta & Coronado (2003) observe the concentration of patent-to-paper citations in the more developed regions in Spain, such as Madrid, Catalonia and the Basque Country, and their scarcity in regions with GDP below 75% of the EU average. However, although Acosta & Coronado do not stress this point, the distribution of patent-to-patent citations (which are more abundant than patent-to-paper citations) is relatively even across regions.

For technology classes, Sampat (2004) provides some empirical evidence and suggests that applicants are more likely to include citations in fields where patenting is important such as chemicals and pharmaceuticals. Drawing on the previous literature, Sampat argues that in fast-moving, more technology intensive classes, the ability of examiners to access current information is limited, and applicants are better informed about the closely related prior art and therefore include more citations than examiners. In technologies that are less science intensive, citation rates between examiners and applicants will generally be more similar. In the case of EPO patents, Criscuolo & Verspagen (2008) find that the share of applicant citations is higher in chemistry and materials and lower in semiconductors, telecommunication, audiovisual and information technology.

For economic sectors, Acosta & Coronado (2003) show that, in Spain, 85% of patent applicants' patent-to-paper citations are concentrated in only three sectors (chemistry, pharmaceuticals and biotechnology). However, patent-to-patent citations predominate in high and medium-high technology sectors, such as electrical engineering and instruments, which is why Leydesdorff (2004) suggests that a sector such as biotechnology is not a valid model for how university-industry interactions occur in general.

Given the scarcity of direct comparisons of citations along patent characteristics, it is difficult to establish any a priori expectations. Assuming that the administrative and technoeconomic reasons for increasing the number of citations affect examiners and applicants equally, we can start by formulating a cautious hypothesis:

Hypothesis 2. Applicants and examiners include comparable distributions of citations

9

⁵ Note that the present empirical analysis studies the Valencian Community, which is included in this group.

independent of patent characteristics.

2.3 Within front-page citations: Examiners behaviour is different depending on whether applicants disclose some knowledge or not

The best conceptual background to understand the differences between applicant and examiner citing-cited pairs is in Alcácer & Gittelman (2006). They investigate different scenarios in terms of examiners' decisions about whether to complement or supplement inventor citations. Implicitly, this approach assumes that examiners have something to complement or supplement because inventors disclose some knowledge.

This may not be an issue in the US, where applicants are subject to 'duty of candour', which forces them to be exhaustive in their inclusion of references to prior art (Meyer, 2000; Michel & Bettels, 2001). However, it is important in Europe, since there is no imposition of that kind on applicants, meaning that they are free to decide whether or not to include references. For instance, Acosta & Coronado (2002) show that only 31% of Spanish patents include applicant full-text patent-to-patent citations and 10% of Spanish patents include applicant full-text patent-to-paper citations, i.e. a small number compared to front-page citations, which appear in all patents. Similarly, Azagra et al. (2009) find that only 30% of patents from the Valencian Community (a Spanish NUTS 2 region) includes at least one applicant citation.

Thus, in the European case, it is worth studying whether examiners treat patents with and without citations equally. If treatment is equal in that in both cases the examiner transposes a standard body of references, then it is more difficult to uphold their objectivity. If treatment is not equal, we would expect the distribution of examiner citations according to the characteristics of the patent, to differ between patents with and without citations, because each will be subject to a case-by-case search report. Let us assume this situation as a starting point:

Hypothesis 3. The distribution of examiner citations is different according to whether the analysis includes patents with applicant citations or all patents.

2.4 Within full-text citations: Applicants who disclose some

knowledge are representative of all applicants

Criscuolo & Verspagen (2008) conclude their study with a plea for greater use of applicant rather than examiner citations, as indicators of knowledge flows. They rely on the assumption that their observation of applicant citations is representative of the behaviour of all applicants. The problem, which the authors are aware of since they study the European case, once again, is lack of applicant citations. What about applicants who do not include citations?

In this situation, it is questionable whether we should consider even full-text applicant citations as being representative of the knowledge base. A good indication might be if the number of patent citations was the same for different types of patents for both applicants who disclose some information and those who do not reveal any. To test for this, it is useful to formulate:

Hypothesis 4. The distribution of applicant citations is similar regardless of whether the analysis includes patents with applicant citations or all patents.

3 The research context: the Valencian Community

The Valencian Community is described as having low absorptive capacity (Azagra, 2007). The main features of the region are:

- low-tech profile of its economic structure (predominance of microfirms in services and traditional manufactures);
- weak innovation activities (innovation is mostly incremental and in the form of machinery and equipment acquisition, with little expenditure on R&D);
- scarcity of qualified personnel even in firms in the knowledge-intensive sectors;
- policy emphasis on enhancing technology transfer (similar to high-tech regions or countries).

Table 1 provides a more detailed description of the industrial structure of the Valencian Community.

{Table 1 around here}

The column presenting gross value added (GVA) shows the share of the average value

of this variable for 1999-2003 (for comparison with patents), in constant prices, by NACE activity. The data are from the Spanish National Statistics Institute's Regional Accounts. We group economic branches following (and extending) the typologies proposed by Pavitt (1984) and Breschi and Malerba (1997). The classification is arguable, but it is not the objective of this paper to justify this, and alternative sector groupings would have led to similar conclusions regarding posterior analysis.

Supplier-dominated sectors predominate, especially 'construction' and services such as 'wholesale and retail trade', 'non-market services' and 'transport, storage and communication'. Within production-intensive sectors, it 'manufacturing of other non-metallic mineral products' stands out and is based on regional strength in ceramic tiles. The contribution of science-based sectors is relatively small. The case of 'real estate, renting and business activities' is rather special, since the high GVA weighting is due mainly to the activities in the supplier-dominated sectors. However, because they include 'computer and related activities' and 'research and development' we chose to classify them under science-based sectors, which is relevant for the information in subsequent columns on patents and patent citations. First, we explain the methodology and data.

4 Methodology and data

The source of patent application information for the Valencian Community is the database of the Spanish Patent and Trademark Office (OEPM), which allows us to select according to 'province', and includes codes for the three Valencian NUTS 3 regions: Alicante, Castellon and Valencia. We recovered patents from 1999 to 2003.6

We chose patents identified as firm patents rather than including patents from all performance sectors, to allow for some institutional homogeneity and a focus on knowledge-industry interaction, and also because they represent the bulk of patenting activity. The OEPM database includes information on name(s) of patent applicant. Identifying their correspondence with firms is not straightforward and involved checking each patent individually. We classified the 1,382 patents registered for the Valencian Community between 1999 and 2003, distinguishing between firm and other types of applicants, based on an the acronyms SL (limited society) and SA (anonymous

⁶ OEPM regularly updates online data. The date of the extractions used in this paper is 23 June 2006.

society) attached to the name.

We constructed a database to include citations for these firm patents. We studied the full-text of the patent application forms for every patent, especially the description field where the applicant includes the prior art, and counted the numbers of citations. We conducted the same exercise for citations in the prior art report (included by patent examiners). This task was made easier since some of the citations already appeared in the OEPM database. This resulted in 712 applicant citations and 2,849 examiner citations, which we classified further according to the following characteristics:

- year of application: from 1999 to 2003;
- route of protection: there are three legal routes to protection of an invention in Spain included in the OEPM database: national, European, and PCT. Because of the few number of patents that apply for EPO protection, we combined them with PCT applications under the label 'international applications';8
- ownership: in the case of patents with several applicants where at least one is a
 firm, we defined cooperation type: with a firm, with a research centre, with an
 innovation/technology centre, and with an individual.
- NUTS 3 region: there are three regions in the Valencian Community –Alicante, Castellon and Valencia;
- technology class: each of the eight sections at the first level of the International Patent Classification (IPC), calculated through a fractional count.

We also attributed patents to economic sectors by linking them to the Analytical System on Spanish Balance Sheets (SABE). This database, which includes economic data on Spanish firms, includes a field for the Statistical Classification of Economic Activities (NACE), revision 1.1. We were able to match 92% of the firm applicants in our sample to the firms included in SABE and assigned to each of these patents the two-digit NACE activity code of the applicant. This allowed us to construct the following variable:

_

⁷ In the full-text of the application form, applicants provide descriptions of their inventions to demonstrate their novelty, to describe them and to explain how they were made. Although most include a section on prior art, it is not obligatory to include citations. Consequently, a patent with no citations does not mean there is no involvement of prior art, but only that it is implicitly referred to.

⁸ Note that PCT does not award patents: applications are subject only to international review and then filed at national patent offices, i.e. PCT applications become national (or not, if they are abandoned).

• economic sector: two-digit NACE activity of applicant firm, calculated with a fractional count, which provides information included in the last three columns of Table 1.

For each variable, we calculate whether the difference between numbers of applicant and examiner citations is significant, and whether there is a significant difference between variable categories using ANOVA and the non-parametric Kruskal-Wallis ANOVA.

For the analysis, we divided the sample into sub-samples - with and without applicant citations. The comparison between the full sample and the sub-sample with applicant citations allows us to test some of our hypotheses. Of course, this implies that applicants that disclose some knowledge (i.e. include at least one applicant citation) hide less knowledge than applicants who do not disclose any knowledge (i.e. do not include any applicant citations). Although it is beyond the scope of this study to verify this, we provide some evidence that this is a reasonable assumption.

5 Results

There is generally not much relation between total number of patents and GVA in Table 1. The exceptions are 'wholesale and retail trade' and 'real estate, renting and business activities', which have a high share of both patents and GVA. Other large sectors, in terms of GVA ('construction', 'non-market services', 'transport, storage and communication'), do not have high shares of patents, while other active patenting sectors do not show high shares of GVA ('rubber and plastic products', 'machinery and equipment n.e.c.', 'chemicals and chemical products'). This suggests that the technological structure of the economy is determined by the type of sectoral innovation system rather than by type of industry.

The knowledge base, proxied by examiner citations, follows the number of patents distribution, but the relation to GVA is slightly larger –mainly due to the central categories of both distributions.

The share of applicant citations follows the share of patents, but is less related to the composition of GVA, because of its high concentration in science-based sectors. Thus, there is a higher number of examiner than applicant citations in sectors with high GVA such as 'wholesale and retail trade', while the opposite is true for low GVA sectors such as 'chemicals and chemical products'. This suggests that applicants in the science-based

sectors tend to provide knowledge. This point is further addressed in the following section.

5.1 Are examiner citations based on administrative requirements? No, they reflect the technoeconomic specialisation

Table 2 shows that the average number of examiner patent citations decreases over time. So does the frequency of mode. This may be due to harmonisation with European standards, which tend towards fewer but more relevant citations, and has been encouraged since 1991 when the EPO gave the OEPM the responsibility for providing search reports for international patent applications.

With the exception of time effects, other inflationary characteristics (route of protection and ownership regime) do not influence the number of examiner citations.

There is also some regional variation: in the NUTS 3 region of Alicante, which has the lowest per capita income, patents include more examiner citations compared to the other two regions. This is because the technologies and sectors that include more examiner citations are present in Alicante.

{Table 2 around here}

Technological variation is also present and deserves some comment. On the one hand, the number of examiner citations is low in science-based technologies such as C. Chemistry; Metallurgy, G. Physics and H. Electricity. On the other hand, some non-science-based categories also have high numbers of average citations: A. Human Necessities, D. Textiles; Paper, B. Performing Operations and E. Fixed Constructions.

Table 3, by economic sector, shows that there is significant variation and, that, although differences across categories are not as clear-cut as in the case of technology classes, science-based sectors do not present many more examiner citations per patent than other sectors.

{Table 3 around here}

Overall, Hypothesis 1 is supported: examiner citations vary according to the distribution of patent characteristics. The pattern of this variation is interesting: (i) except for time

effects, administrative reasons do not seem to be influential; (ii) less science-intensive technologies have fewer examiner citations, and for economic sectors, the opposite is not true. Hence, examiners are paying attention to the technoeconomic structure of the economy.

5.2 Do examiners trust applicants? No, they correct by region, technology class and economic sector

If we take single patents as the unit of observation, the correlation coefficient between number of applicants and examiner citations is close to zero. That is, there is no relation between the numbers of citations included by the two parties. This means that Hypothesis 2 is not supported. To dig deeper into the source of differences among patents grouped by characteristics, we focus on applicant citations (2nd last column in Table 2).

In terms of patent characteristics we find that the pattern of applicant citations is different from that of examiner citations: on the one hand, the average number of citations over time does not decrease significantly; on the other hand, applicants tend to include more citations if the patent application is international and has co-applicants.

In terms of technoeconomic characteristics, applicant citations and examiner citations are similar in that technological and economic heterogeneity and more especially regional variation are significant. However, the sources of variation are very dissimilar between examiners and applicants.

Castellon, the region with the smallest numbers of examiner citations, has the highest numbers of applicant citations. The region with the largest numbers of examiner citations, Alicante, ranks second for applicant citations. Thus, from the regional distribution of citations, we can say that to an extent, examiners complement applicant citations.

In the case of technologies, there are significant differences between examiner and applicant citations in all classes. The largest differences are in non science-based technologies, where numbers of applicant citations are much smaller compared to examiners' (D. Textiles; Paper, B. Performing Operations; Transporting and A. Human Necessities). The smallest differences are in some science-based technologies (clear in the case of C. Chemistry; Metallurgy; to a lesser extent, in the case of H. Electricity).

Thus, to an extent, examiners counterweight the concentration of applicant citations in a few technologies.

These findings about complementarity are even clearer in the case of economic sectors (Table 3). Here, we can see significant differences between examiner and applicant added citations in almost all categories. In the supplier-dominated and production-intensive sectors, examiners cite more when applicants cite less, e.g. in Other supplier-dominated sectors, Manufacture of machinery and equipment n.e.c. and Manufacture of wood, paper, publishing, media. This inverse relation appears also in Manufacture of chemicals and chemical products, where most applicant citations are concentrated, and numbers of examiner citations are small. However, there are exceptions, such as Manufacture of electrical machinery and equipment n.e.c., a science-base sector with large differences between examiner and applicant citations.

These differences would seem to be the source of the overall lack of correlation between applicant and examiner citations and, overall, do not support Hypothesis 2.

Note that intergroup variation is always significant because examiners include more citations than do applicants – which is logical since the sample includes patents with no applicant citations. If we exclude this group, the results for intergroup variation are more interesting.

5.3 Do examiners behave differently if applicants include citations to disclose knowledge? No, they behave similarly

Table 4 presents the results for the sub-sample of patents with at least one applicant citation. The column labelled "average number of examiner citations" describes whether, compared to the columns with the same name in Table 2, examiners' behaviour is the same, irrespective of whether or not applicants include citations.⁹

{Table 4 around here}

In Table 4, we can see that examiners tend to include fewer citations over time than in Table 2, but that there is no significant variation between routes of protection; patents from Alicante include more examiner citations, similar to the numbers in non-science-

_

⁹ Table 4 excludes technology classes with fewer than 10 patents.

based technologies and supplier dominated sectors (Table 5). This reflects the position for examiner citations for the overall sample.

{Table 5 around here}

The difference is for co-applications: they are not significant in the overall sample but show more examiner citations in the sub-sample. However, the last column in Table 5 indicates that variations between examiner and applicant citations are not significant, suggesting that intragroup variation is not important.

The findings indicate that, in general, examiners behave similarly regardless of whether or not applicants include citations, which does not support Hypothesis 3.

5.4 Are applicants who disclose some knowledge representative of most applicants? No

Even if we focus on applicants that include at least one citation, the correlation with numbers of examiner citations is not significantly different from zero. However, as the column labelled "average number of applicant citations" in Table 4 shows, there are more similarities with examiners than in the full sample.

The average number of applicant citations for patents in Table 4 decreases significantly over time and we do not observe differences among route of protection or ownership regime. These are the same as the results for examiner citations in the full sample. For the sub-sample, intergroup variation does not tend to be significant. The differences between examiners and applicants are small along the patent characteristics mentioned above.

The differences that do exist are based on technoeconomic reasons. Among technology classes, there is significant intra- and intergroup variation: the highest difference is C. Chemistry; Metallurgy, where examiners introduce fewer citations than applicants. In H. Electricity and in B. Performing Operations; Transporting, the reverse is true. Therefore, there are indications that examiners try to complement applicants' knowledge disclosure. The findings are similar for economic sectors (Table 5).

Comparison between applicant citations in the full sample and the sub-sample is interesting. The number of citations in the full sample (Table 2) is stable over time, but

increases for international applications and applications with more than one applicant. Table 4 shows that the number of citations (in the sub-sample of patents with at least one applicant citation) significantly decreases, and does not vary based on route of protection or a positive number of co-applicants.

This does not support Hypothesis 4. Applicants that include at least one citation bias the emergent statistics on applicant citations, because they inflate the number of citations for reasons such as route of protection and ownership regime.

5.5 An unforeseen side result -a fixed number of examiner citations

The above results suggest that examiners merely replicate bodies of knowledge from patent to patent, ignoring the citations included by applicants. To verify this, we matched pairs of common examiner citations in different patents but did not find huge correspondence (less than 30%). However, the number of examiner citations in these patents was often the same. Therefore, it seems that examiners tend to add a 'fixed' number of citations, although these citations may be different. This contradicts the claim of replication of bodies of knowledge, but imply the existence of a standard practice in terms of number of citations. This result deserves further research.

6 Conclusions

The results in this paper suggest that it would be premature to defend the use of examiner citations in patents to justify research funding or evaluate performance. Our results also do not support the hypothesis that applicant citations are better sources of information. However, the results support the idea that patent citations have different meanings. We develop these ideas in the next paragraphs.

Examiner citations are related to the technoeconomic structure of the territory. If the local industry has very few leading and patenting sectors, there will be fewer examiner citations. For other sectors, even though the knowledge base exists, examiner patent citations will not capture it. Therefore, studies on examiner citations should consider the technoeconomic structure. Examiner citations may be representative of the knowledge base where there is strong industrial specialisation in highly patenting or leading sectors but require complementary analysis in every other case.

This paper looked at full-text applicant citations, which highlighted some practices of patent examiners related to adding citations, namely to include more (less) references than applicants, in patents for technology classes and economic sectors that rely less (more) on the science base. The result is a more homogeneous distribution of citations that masks the scarce importance of codified knowledge for traditional economic activities.

If we take into account the fact that examiners also assign IPC codes to patents (Kang et al., 2007), the 'noise' in citations grows exponentially. However, we also show that using applicant citations creates problems, since applicants may add large numbers of (hardly justified) citations in some patents.

Overall, this makes a case for increasing resource allocation to patent examiners' to allow for case-by-case search reports, which would increase applicants' incentives to disclose information to the EPO and the World Intellectual Property Organization - WIPO (and avoid artificial distortions among types of patents), and provide better legal standards for storing information, especially for analysing full-text applicant citations.

Many of the explanations for parts of the evidence from this study are related to the conduct of examiners and applicants. Interviews would be useful to check the consistency of explanations. Ongoing work suggests that patent examiners' personal characteristics also determine the inclusion of citations (Lemley & Sampat, 2008).

However, we would propose investigation of the differences between examiner and applicant citations as indicators of trust in the research system. For example, in Table 4 the case of technology classes: for E. Fixed Constructions, the difference is 0, possibly implying that examiners trust applicants; for C. Chemistry; Metallurgy, the difference in absolute value is equal to 2, possibly implying that examiners do not trust applicants. We would expect that the more developed the research system, the more often this indicator will take a zero value.

To what extent these results are idiosyncratic of the sample we analysed, which is based on a region with low absorptive capacity, is questionable. Elsewhere (Azagra et al., 2009), we emphasise those properties of the sample that the regional context influences more clearly. Here, we want to stress that, although some of our findings may be idiosyncratic, the importance of the hypotheses and their methodological implications should guide the debate on patent citations.

References

- Acosta, M. & Coronado, D. (2002). Las relaciones ciencia-tecnología en España. Evidencias a partir de las citas científicas en patentes. Economía Industrial, 346, 27-46.
- Acosta, M. & Coronado, D. (2003). Science-technology flows in Spanish regions. An analysis of scientific citations in patents. Research Policy, 32, 1783-1803.
- Alcácer, J. & Gittelman, M. (2006). Patent citations as a measurement of knowledge flows: The influence of examiner citations. Review of Economics and Statistics, 88(4), 774-779.
- Allison, J.R. & Sager, T.W. (2007). Valuable Patents Redux: On the Enduring Merit of Using Patent Characteristics to Identify Valuable Patents. Texas Law Review, 85, 1769-1797.
- Azagra-Caro, J.M., 2007. What type of faculty member interacts with what type of firm? Some reasons for the delocalisation of university-industry interaction. Technovation, 27, 704-715
- Azagra-Caro, J.M., Fernández de Lucio, I., Perruchas, F. & Mattsson, P. (2009). What do patent examiner inserted citations indicate for a region with low absorptive capacity? Scientometrics, 80(2), 441–455.
- Azagra-Caro, J.M., Yegros-Yegros, A. & Archontakis, F. (2006). What do university patent routes indicate at regional level? Scientometrics, 66(1), 219-230.
- Breschi, S. & Malerba, F. (1997). Sectoral Innovation Systems: Technological Regimes, Schumpeterian Dynamics, and Spatial Boundaries, C. Edquist (ed.): Systems of Innovation: Technologies, Institutions and Organizations, ch. 6. London and Washington: Pinter.
- Callaert, J., van Looy, B., Verbeek, A., Debackere, K. & Thus, B. (2006). Traces of Prior Art: An analysis of non-patent references found in patent documents. Scientometrics, 69(1), 3-20.
- Chen, C. (2003). Book Review: Patents, Citations & Innovations: A Window on the Knowledge Economy. Journal of the American Society for Information Science and Technology, 54(8), 802–803.

- Criscuolo, P. (2006). The 'home advantage' effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO. Scientometrics, 66(1), 23–41.
- Criscuolo, P. & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. Research Policy, 37, 1892-1908.
- EC (2007). Commission Green Paper 'The European Research Area: New Perspectives'. COM(2007), 161.
- Guan, J.C. & Gao, X. (2009). Exploring the h-Index at Patent Level. Journal of the American Society for Information Science and Technology, 60 (1), 35-40.
- Hall, B.H., Jaffe, A. & Trajtenberg, M. (2000). Market value and patent citations: A first look. NBER Working Paper, 7441.
- Hassan, E. (2003). Simultaneous Mapping of Interactions between Scientific and Technological Knowledge Bases: The Case of Space Communications. Journal of the American Society for Information Science and Technology, 54(5), 462-468.
- Jaffe, A. & Trajtenberg, M. (1996). Flows of knowledge from universities and federal laboratories: Modeling the flow of patent citations over time and across institutional and geographic boundaries. PNAS, 93, 12671-12677.
- Jaffe, A., Trajtenberg, M. & Fogarty, M.S. (2000). The meaning of patent citation: Report on the NBER/Case-Western reserve survey of patentees. NBER Working Paper, 7631.
- Kang, I.S, Na, S.H., Kim, J. & Lee, J.H. (2007). Cluster-based patent retrieval. Information Processing & Management, 43(5), 1173-1182.
- Kostoff, R.N. (1994). Federal Research Impact Assessment: State-of-the-Art. Journal of the American Society for Information Science and Technology, 45(6), 428-440.
- Lai, K.K. & Wu, S.J. (2005). Using the patent co-citation approach to establish a new patent classification system. Information Processing and Management, 41, 313–330.
- Lemley, M. & Sampat, B.N. (2008). <u>Examiner Characteristics and the Patent Grant Rate</u>, mimeo: http://www.nber.org/~confer/2008/si2008/IPPI/lemley.pdf, last access: 13/01/2009.
- Leydesdorff, L. (2004). The University–Industry Knowledge Relationship: Analyzing Patents and the Science Base of Technologies. Journal of the American Society for

- Information Science and Technology, 55(11), 991-1001.
- Leydesdorff, L. (2008). Patent Classifications as Indicators of Intellectual Organization. Journal of the American Society for Information Science and Technology, 59(10), 1582-1597.
- Merrill, S.A., R.C. Levin & M.B. Myers (eds) (2004). A Patent System for the 21st Century. Washington, DC: The National Academies Press.
- Meyer, M. (2000). What is special about patent citations? Differences between scientific and patent citations. Scientometrics, 49(1), 93-123.
- Michel, J. & Bettels, B. (2001). Patent citations analysis: A closer look at the basic input data from patent search reports. Scientometrics, 51(1), 185-201.
- Narin, F. & Noma, E. (1985). Is technology becoming science? Scientometrics, 7(3-6), 369-381.
- Noma, E. & Olivastro, D. (1984). Are there enduring patents? Journal of the American Society for Information Science and Technology, 36(5), 297 301.
- Pavitt, K. (1984): Sectoral patterns of technical change: Towards a taxonomy and a theory. Research Policy 13, 343-373.
- Sampat, B. (2004). Examining Patent Examination: An Analysis of Examiner and Applicant Generated Prior Art, Working Paper, School of Public Policy, Georgia Institute of Technology.
- Stock, M. & Stock, W.G. (2006). Intellectual Property Information: A Comparative Analysis of Main Information Providers. Journal of the American Society for Information Science and Technology, 57(13), 1794-1803.
- Tijssen, R.J.W. (2001). Global and domestic utilization of industrial relevant science: patent citation analysis of science–technology interactions and knowledge flows. Research Policy, 30, 35–54.
- Thompson, P. (2006). Patent citations and the geography of knowledge spillovers: Evidence from inventor- and examiner-added citations. Review of Economics and Statistics, 88(2), 383-388.

Tables:

Industrial structure, technological structure and knowledge base of the TABLE 1

Valencian Community.					
Description	NACE	Share of total gross valued added	Share of total patents	Share of total number of examiner citations	Share of total number of applicant citations
Supplier-dominated sectors					
Agriculture, hunting and forestry	01,02,05	3%	1%	1%	0%
Textiles, textile products, leather and leather products	17,18,19	3%	2%	2%	0%
Wood and wood products	20	1%	4%	4%	0%
Pulp, paper and paper products; publishing and printing	21,22	1%	4%	3%	1%
Rubber and plastic products	25	1%	10%	9%	5%
Manufacturing n.e.c.	36,37	2%	3%	4%	1%
Construction	45	9%	2%	2%	1%
Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods	50,51,52	13%	13%	12%	6%
Hotels and restaurants	55	7%	0%	0%	0%
Transport, storage and communication	60,61,62,63,64	9%	0%	0%	0%
Financial intermediation	65,66,67	5%	0%	0%	0%
Non-market services	75,80,90,95,99	12%	0%	0%	0%
Health and social work	85	3%	0%	0%	0%
Other community, social and personal service activities Production-intensive sectors	91,92,93	3%	0%	6%	4%
Food products, beverages and tobacco	15,16	2%	1%	1%	5%
Other non-metallic mineral products	26	4%	5%	4%	6%
Basic metals and fabricated metal products	27,28	2%	6%	5%	2%
Machinery and equipment n.e.c.	29	1%	12%	13%	5%
Transport equipment Science-based sectors	34,35	2%	0%	0%	0%
Mining and quarrying; coke, refined petroleum products and nuclear fuel	10-14, 23	0%	1%	1%	2%
Chemicals and chemical products	24	1%	11%	10%	37%
Electrical and optical equipment	31,32,33	1%	7%	6%	5%
Electricity, gas and water supply	40,41	2%	0%	1%	1%
Real estate, renting and business activities	70,71,72,73,74	13%	19%	16%	20%
Total		100%	100%	100%	100%

TABLE 2 Frequency of citations in patents from the Valencian Community with firm applicants – full sample. TABLE 2

Variable	Category	N° of patents	Average number of examiner citations (mode; frequency of mode)	Average number of applicant citations (mode; frequency of mode)	Significand e of intergroup variation
Year of application	1999	86	5.87 (5; 23%)	1.26 (0; 81%)	*
	2000	114	4.82 (4; 25%)	1.82 (0; 72%)	*
	2001	129	4.64 (4; 24%)	0.95 (0; 74%)	*
	2002	133	4.71 (3 and 4; 21%)	1.05 (0: 65%)	*
	2003	109	4.40 (4; 24%)	0.83 (0; 69%)	*
	Significance of the growth rate		*	n.s.	
Route of protection	National	459	4.81 (4; 24%)	1.02 (0; 74%)	*
protection	International	112	4.91 (5; 22%)	1.81 (0; 63%)	*
	Significance of intragroup variation		n.s.	*	
Ownership	No co-applicants	545	4.80 (4; 22%)	1.10 (0; 73%)	*
	With co-applicants	26	5.38 (4; 27%)	2.62 (0; 46%)	*
	Significance of intragroup variation		n.s.	*	
NUTS 3	Alicante (1)	148	5.32 (4; 21%)	1.53 (0; 70%)	*
region	Castellon (2)	99	4.65 (5; 25%)	1.81 (0; 66%)	*
	Valencia (3)	324	4.66 (4; 23%)	0.81 (0; 74%)	*
	Significance of the ratio (3)/(1) Significance of the		* n.s.	*	
Technolog y class (IPC section)	ratio (3)/(2) A. Human Necessities	152	4.48 (3; 17%)	1.18 (0; 72%)	*
	B. Performing Operations;	160	4.06 (3; 18%)	0.55 (0; 83%)	*
	Transporting C. Chemistry; Metallurgy	52	3.16 (3; 24%)	2.41 (0; 54%)	*
	D. Textiles; Paper	18	4.24 (4; 21%)	0.43 (0; 75%)	*
	E. Fixed Constructions	74	4.02 (4; 19%)	1.26 (0; 67%)	*
	F. Mechanical Engineering; Lighting; Heating; Weapons; Blasting	32	3.27 (2 and 3; 18%)	0.19 (0; 87%)	*
	G. Physics	25	3.20 (3; 24%)	0.17 (0; 82%)	*
	H. Electricity Significance of	59	3.48 (4; 22%)	0.82 (0; 61%)	*
	intragroup variation				

n.s. = not significant

TABLE 3 Frequency of citations in patents from the Valencian Community with firm applicants – sample of firms assigned to economic sectors.

Variable	Category	N° of patents	Average number of examiner citations (mode; frequency of mode)	Average number of applicant citations (mode; frequency of mode)	Significance of intergroup variation		
Economi	Supplier-dominated sectors						
c sector	Manufacture of wood, paper, publishing, media (NACE 20, 21, 22)	44	4.80 (4; 29%)	0.16 (0; 86%)	*		
	Manufacture of rubber and plastic products (NACE 25)	52	5.06 (5; 21%)	0.69 (0; 71%)	*		
	Trade, maintenance and repair (NACE 50, 51, 52)	70	4.33 (4; 27%)	0.38 (0; 78%)	*		
	Other business activities (NACE 70, 71, 74)	83	4.23 (4; 23%)	1.24 (0; 59%)	*		
	Other supplier- dominated sectors (NACE 01, 17, 18, 19, 36, 45, 63, 85) Production-intensive sec	41	5.40 (4; 28%)	0.28 (0; 79%)	*		
	Manufacture of other non-metallic mineral products (NACE 26)	27	4.22 (3 and 4; 21%)	1.29 (0; 75%)	*		
	Manufacture of fabricated metal products, except machinery and equipment (NACE 28)	30	4.60 (4; 23%)	0.53 (0; 80%)	*		
	Manufacture of machinery and equipment n.e.c. (NACE 29)	66	5.05 (5; 22%)	0.35 (0; 81%)	*		
	Other production- intensive sectors (NACE 15, 33, 34) Science-based sectors	26	4.62 (4; 19%)	1.88 (0; 62%)	*		
	Manufacture of chemicals and chemical products (NACE 24)	58	4.69 (5 and 3; 22%)	4.47 (0; 38%)	n.s.		
	Manufacture of electrical machinery and equipment n.e.c. (NACE 31)	10	4.50 (4; 30%)	0.20 (0; 90%)	*		
	R&D and computer activities (NACE 72, 73)	19	5.47 (5; 26%)	2.00 (0; 74%)	*		
	Other science-based sectors (NACE 14, 23, 32, 40)	15	4.57 (4; 29%)	1.74 (0; 65%)	*		
	Significance of intragroup variation		*	*			

Frequency of citations in patents from the Valencian Community with TABLE 4

Variable	icants – sample of pat Category	N° of patents	Average number of examiner citations (mode; frequency of mode)	Average number of applicant citations (mode; frequency of mode)	Significance of intergroup variation
Year of applicatio	1999	16	6.00 (5; 31%)	6.75 (3; 31%)	n.s.
	2000	32	5.03 (4; 34%)	6.50 (1; 22%)	n.s.
	2001	34	4.56 (5; 23%)	3.62 (1; 35%)	n.s.
	2002	46	4.72 (3; 24%)	3.04 (1; 33%)	n.s.
	2003	34	4.18 (4; 38%)	2.65 (1; 44%)	*
	Significance of the growth rate		*	*	
Route of	National	120	4.68 (3; 22%)	3.88 (1; 34%)	n.s.
protectio n	International	42	4.98 (4; 33%)	4.83 (1; 26%)	n.s.
11	Significance of intragroup variation		n.s.	n.s.	
Ownershi	No co-applicants	148	4.64 (4; 24%)	4.06 (1; 32%)	n.s.
p	With co-applicants	14	6.00 (4; 29%)	4.86 (1; 29%)	n.s.
NUTS 3 region	Significance of intragroup variation Alicante (1)	45	* 5 19 (4 and 6)	n.s. 5.04 (2; 31%)	n c
	Castellon (2)	34	5.18 (4 and 6; 22%) 5.06 (4 and 5;	5.26 (1; 32%)	n.s.
	Valencia (3)	83	26%) 4.41 (3; 25%)	3.17 (1; 36%)	*
	Significance of the ratio (3)/(1)		*	*	
	Significance of the ratio (3)/(2)		n.s.	*	
Technolo gy class (IPC section)	A. Human Necessities	42	4.52 (4; 17%)	4.34 (1; 27%)	n.s.
	B. Performing Operations; Transporting	37	3.58 (3;31%)	2.18 (1; 33%)	*
	C. Chemistry; Metallurgy	25	3.49 (3; 20%)	5.29 (1; 23%)	*
	E. Fixed Constructions	23	4.20 (4; 23%)	3.82 (2; 23%)	n.s.
	H. Electricity	23	3.25 (4; 31%)	2.16 (1; 35%)	*
	Significance of intragroup variation		*	*	

n.s. = not significant

TABLE 5 Frequency of citations in patents from the Valencian Community with firm applicants – sample of patents with at least one applicant citation and firms

assigned to economic sectors.

Variable	Category	N° of patents	Average number of examiner citations (mode; frequency of mode)	Average number of applicant citations (mode; frequency of mode)	Significance of intergroup variation
Economi	Supplier-dominated sect	tors			
c sector	Manufacture of rubber and plastic products (NACE 25)	15	5.13 (5; 27%)	2.40 (1 or 2; 33%)	*
	Trade, maintenance and repair (NACE 50, 51, 52)	14	4.66 (4; 37%)	1.74 (1; 44%)	*
	Other business activities (NACE 70, 71, 74) Production-intensive sec	34 ctors	3.79 (4; 26%)	3.03 (2; 35%)	n.s.
	Manufacture of machinery and equipment n.e.c. (NACE 29)	11	4.04 (4 and 5; 23%)	1.83 (1; 38%)	*
	Other production- intensive sectors (NACE 15, 33, 34) Science-based sectors	10	4.50 (4; 30%)	4.90 (3; 40%)	n.s.
	Manufacture of chemicals and chemical products (NACE 24)	36	4.69 (4; 28%)	7.19 (1; 19%)	*
	Significance of intragroup variation		*	*	

n.s. = not significant