THE DETERMINANTS OF THE PRIVATE VALUE OF PATENTED INVENTIONS*

Alfonso Gambardella
Bocconi University, Milan
alfonso.gambardella@unibocconi.it

Dietmar Harhoff
Ludwig-Maximilians-Universität (LMU) München
harhoff@lmu.de

Bart Verspagen
Maastricht University
b.verspagen@algec.unimaas.nl

Version January 2011

Abstract

By using survey data on a large sample of European patents, we estimate the determinants of a composite indicator of patent value that summarizes information from many indirect indicators commonly employed by the literature. The elasticity of man-months on the value of one patent is small, 4%, while that on the number of patents in the portfolio of technically connected patents is higher, about 20%. This is in line with Rosenberg's view of the innovation process: firms can better control the number of innovations than their economic value, which depends on many factors (demand or complementary inventions) outside their control. The paper also finds that more experienced inventors, with more past citations, exploit incremental trajectories, as implied by the fact that they produce very efficiently many innovations of lower value. In addition, we find that education (PhD) replaces many years of age-related experience, and that portfolios of higher values are more likely to come from PhDs (or older inventors) who have not yet produced major inventions.

JEL Classification: L20, O31, O33, O34

Keywords: Patents, Inventors, Technical Change, Intellectual Property Rights

* We have benefited tremendously from comments and suggestions made by Ashish Arora, Iain Cockburn, Wesley Cohen, Jesse Giummo, Dominique Guellec, Bronwyn Hall, Thomas Hellman, Jacques Mairesse, Susanne Prantl, Mike Scherer, Scott Stern and Manuel Trajtenberg. All responsibilities remain ours. We thank the European Commission, Contract N. HPV2-CT-2001-00013, for supporting the creation of the PatVal-EU dataset. A.G. also acknowledges financial support from the Italian Ministry of University Research and Bocconi University. D.H. acknowledges financial support from the Deutsche Forschungsgemeinschaft through its SFB/TR 15 program (Project C2).

1. INTRODUCTION

Despite a plethora of research in this field, the process by which investments in innovation translate into economic values is not fully understood. A common clichè is that the innovation process is uncertain, and thus the value of innovation outputs is volatile. However, since profit-maximizing firms rationally invest in R&D, another truism is that investments in these resources matter as well. At the same time, we do not know much about the weight of different factors. Is the value of innovations largely determined by the sector or type of technology, or are there differences depending on the individual inventors or the organization in which they are employed? How much does the value of innovation depend on investments in resources or is it largely unpredictable? How important are inventor characteristics compared to the organization in which the invention is produced?

In this paper we address these issues by estimating the determinants of the economic value of a large sample of patents held by firms. Patents are not the same thing as innovation ouput because not all innovations are patented. In addition, the mapping from innovations to patents is often not one-to-one. This means that we will only be able to estimate the determinants of the value of patents. However, not only patents cover a relevant share of innovation output, but they are an important asset of the firm. Also, this paper does not test a specific theory, but it is largely a measurement paper. The data that we have collected enable us not only to employ unique information on patents, but also on the applicant organizations and above all the inventors. Rather than turning our research into a theory-paper, there is value in showing new empirical patterns and relationships and the magnitudes of the impacts.

We employ a unique and comprehensive dataset of patents drawn from a large scale survey of European inventors (PatVal-EU). The survey collected data on an inventor self-assessed measure of the value of the patent and a broad set of characteristics of the inventors and the context of the invention. We combined these data with other data on firms, inventors, industry and other factors.

Our measure of value stems from a question in the survey where we ask the inventors to report the minimum price at which they think that the applicant would have sold their patents to a potential competitor on the day in which the patent was granted. This, however, is not the measure that we use in our regressions. Since it may be affected by subjective assessments of the respondent, we regress it on a large set of indirect indicators of patent value commonly employed by the literature. We find that our survey-based measure is highly correlated with these objective indicators, which gives us confidence about its ordinal ranking. We then use the expected value of this regression as the dependent variable of our regressions of the determinants of patent value. This is like employing a composite indicator of the indirect measures, as the one estimated by Lanjouw and Schakerman (2004). However, unlike them, we anchor the indicators to an index provided by the inventors. Also, because most of our confidence is on the ordinal ranking of our measure, our analysis focuses on the factors that produce variations in our dependent variable (the "determinants" of patent value), and the direction of these variations.

Our main contribution deals in particular with three factors affecting the value of the patent: 1) the resources (man-months) invested in the project leading to the patented invention; 2) the number of technically connected patents; 3) the characteristics of the inventor. As far as 1) is concerned, our survey asks the inventors to indicate the man-months specifically dedicated to the research underlying the innovation output. As a result, this is one of the few papers testing the impact of the specific investment in resources associated to the research leading to the patent. In a seminal article, Hausman et al. (1984) find a strong correlation between R&D input and patent output at the firm level. This suggests that while innovation output has a fair degree of randomness, the amount of resources invested in the process is an important determinant of such an output. Our analysis provides further evidence on this matter using detailed patent-level data.

Point 2) introduces a novelty in patent analyses. Patents are often technically related. There are related inventions, or extentions of basic inventions, as well as interconnected innovations stemming from the same research program, which are also patented. Our survey asks the inventors to indicate the number of patents connected with the focal invention, and we estimate whether the number of connected patents affects the value of the focal patent. In addition, since the given PatVal-EU patent under consideration is a random draw of the set of connected patents, we can take its value to be a noisy measure of the average value of the patents in the connected set, and we can study the determinants of the value of the entire set of connected patents. This paper is then one of the first attempts to estimate the value of patent sets rather than individual patents.

Finally, under 3), we test the impacts of several of inventor characteristics, which we obtain from our survey, viz. the inventor patenting experience or ability (measured by the citations to her past patents), her age or academic degree, her motivations, her experience within the organization in which she is employed. In a world in which human capital and the talent of individuals appear to play a growingly important role, we assess how much these individual characteristics explain patent value.

We treat man-months, the number of connected patents, and the inventor's past citations as endogenous. In addition, we employ firm- and project-level controls. To our knowledge, this is the first attempt to determine the impact, and the relative importance, of such a comprehensive set of factors on the value of patents. We test the effect of many factors that were ignored in previous studies that employed mainly variables collected from patent documents.

We find that the man-months invested in the specific patent project affect the value of the patent.

However, the elasticity is small, 4%. Interestingly, when we instrument for the number of connected patents and the inventor past citations they are not significant. This prompted us to investigate further the determinants of the number of connected patents and the role of the inventor past citations. The PatVal-EU survey also asked the inventor to indicate the man-months employed in the production of the entire set of connected patents. We therefore estimated an equation for the number of connected patents as a function of controls, the man-months invested in the whose set of connected patents, the man-months invested in the focal patent, and the inventor past citation, with these three variables treated as endogenous. The rationale for using both measures of man-months is that the former captures the whole investment in the patent set, while the latter proxies for the average size of investment in each patent of the set. For example, the total man-months invested in the set may increase because of investments in new technically related patents or because of increases in the average scale of the given set of patents. We also show the results of estimating equations for the total man-months invested in the set and the man-months invested in the specific PatVal-EU patent. The covariates are controls and the investor past citation are treated as endogenous.

With this additional exercise we can study the determinats of the total value of the portfolio of connected patents, viz. the number of connected patents times the value of the PatVal-EU patent, which proxies for the average patent value in the portfolio. We obtain some intriguing results. First, the elasticity of man-months increases to about 24%, with the difference from 4% explained by the effect of man-months on the number of connected patents. Thus, man-months raise the value of a patent portfolio largely because of the effect on the number (quantity) rather than on the average value (quality). Rosenberg (1982) noted that it is hard to predict the value of technologies because of the many factors that the inventor or the firm cannot fully control – e.g., demand, the presence of complementary assets or products, network externalities. By contrast, inventors or firms can control better the number of inventions that they produce, a point also raised by other studies (Audia and Goncalo, 2007; Mariani and Romanelli, 2007; Conti et al., 2011). Our analysis does not distinguish whether the patent set is composed of genuinely different inventions or of patents that protect a core invention for strategic reasons (e.g., Ziedonis, 2004). However, this does not affect our point that it is easier to control the number of patents produced than their value.

Second, we find that inventors with higher past citations produce portfolios of lower value – in terms of both number and average value of patents. This suggests that these inventors exploit trajectories of incremental innovations from assets, investments or discoveries that are now sunk. This is consistent with the fact that we also find that they take considerably fewer resources to produce them, as measured by the man-months invested in the project. Again, we cannot distinguish whether this is the outcome of sheer ability or of previous important discoveries (proxied by citations) for which assets are sunk. Nonetheless, what is key for us is that the inventor past successul experience does not translate into outcomes of greater value but in more outcomes, of lower value, per unit of resource. This result is consistent with Conti et al. (2011) who find that experienced inventors produce patents of slightly lower value, but many more of them per interval of time.

Finally, we find that age, and above all education, raise the man-months invested in the projects and then the value of the patent portfolio – once again largely through the number rather than value of patents. Since older and more educated inventors also have more citations, these effects compensate each other. Thus, more educated inventors with many citations make roughly similar investments in innovation projects, but they are far more efficient (whether because of experience, ability or sunk assets), in that they produce many more patents of similar average value.

The next section relates our work to the literature. Section 3 and 4 present the data and the econometric framework. Section 5 presents the empirical results, and Section 6 provides a concluding discussion.

2. PRIOR THEORETICAL AND EMPIRICAL CONTRIBUTIONS

There are two main streams of the literature on patent value. The first one focuses on the search for indicators of patent value. This starts from the work that uses data on patent renewal payments (e.g., Schankerman and Pakes, 1986), where the working hypothesis is that how long a patent is "kept alive" is an indicator of its value. Trajtenberg (1990) observed that the number of citations received by a patent correlate well with its economic value, thus adding a second indicator to the spectrum. Later work extended the range of indicators further to include, for example, the international coverage of a patent and whether or not a patent is opposed or litigated (Harhoff et al., 2003a). An alternative approach uses indicators based on market transactions instead of those directly taken out of patent databases. In this vein, Serrano (2006) applies data on the sale of patent rights to arrive at an indicator of value. While these contributions provide insights on the distribution of patent value, they do not, by and large, address the question of what determines patent value. What we obtain from these studies is some indicators that can be used to estimate the value of patents ex post.

The second stream of the literature estimates the empirical relationship between the value of patented inventions and the stock market value of the firm (Griliches, 1981; Pakes, 1985; Hall et al., 2005; Bessen, 2009). The focus of this literature is on the correlation between market value and patent counts or citations. In so doing, these papers also provide estimates of the market evaluation of the marginal value of the firm patents.

The literature that addresses the question of the determinants of patent value is scantier, especially when we focus on empirical implementations. Starting with the pioneering work by Hausman et al. (1984), there is a vast literature looking at the relationships between R&D or other measure of investments on patent counts or the various patent indicators, but practically nothing on the relationships between innovation or patent ouput and the resources invested in the project.

At the same time, there is a growing literature suggesting that the role of individual inventors can be quite important. This literature starts with Lotka's (1926) famous observation that researcher performance is very skewed. Ernst et al. (2000) report similar findings for a more recent sample of industrial engineers, based on a patenting indicator. Although this literature, as far as patenting is concerned, has not paid much attention to the issue of what explains differences in inventor performance, its suggestion is that there should be at least some observable characteristics at the inventor level that do this. Gittelman and Kogut (2003) find that how much biotech firms are able to benefit from basic science depends on the decisions they make with regard to which type of scientists to hire. Zucker et al. (2002) find that "star" scientists (as indicated by the number of publications) are important drivers of firm performance in biotech.

The literature has also recognized that many patented inventions do not occur in isolation. Typically, innovation projects lead to many inventions, which are then patented. Note that here we do not mean "equivalent" patents, or patent "families", as they are called. Equivalent patents regard the same invention patented in different patent offices around the world. Here we mean different inventions that are technically related to each other (e.g., complementary, substitutes, or simply flowing from the same project and technically related other in some way). As we shall see, this is what we asked to our inventors in the survey, and what we will use in our analysis. The reason why we need to take these innovations into account is that they may affect the value of the focal patent. In addition, a thorough analysis of the value of individual patents is complemented by the assessment of the total value of a portfolio of technically related patents.

An important area of the literature on the determinants of the value of patents is the one that highlights the distinction between the part of the patent value related to the market protection given by the patent, and the value of the invention to the firm without a patent being issued (Arora et al., 2008; Bessen, 2008). Arora et al. (2008) use the term patent premium to describe the part of the patent value that is related to protection, and estimate that, conditional on having patented an innovation, firms expect to earn, gross of the cost of patent application, between 75% to 125% more than if they had not patented those innovations (the reported variation is industry variation). This means that, depending on the industry, the patent premium can be substantial. One of the important determinants of the patent premium is the degree of competition (with higher levels of competition implying a larger patent premium), but other industry effects, such as the impact of regulation, or the cumulativeness of invention, also enter the patent premium (Harhoff et al., 2003b). In this paper we will not provide a structural estimation that distinguishes the value of the patent premium vis-à-vis the value of the invention as a whole. This is because our sample is only composed of patented innovations and thus we cannot make inferences on the different values of patented and non-patented innovations to retrieve the patent premium. However, we use measures of the degree of competition in producing the technology to control for factors that may affect the patent premium.

3. SAMPLE AND DEPENDENT VARIABLE

3.1 The PatVal-EU Dataset

The PatVal-EU survey collected data on 9550 patents (out of 28,470 submissions). They are patents with priority date 1993-1997 applied for to the European Patent Office, and such that the address of the first inventor listed in the patent is in Denmark, France, Germany, Hungary, Italy, the Netherlands, Spain or the UK. The original 28,470 patents mirrored several characteristics, available from the patent document, of the universe of patents with priority dates 1993-1997 in the countries above. We slightly oversampled "important" patents (i.e., patents with at least one citation and that were opposed by third parties, as allowed for by the European patent legal framework). However, all our analyses below include sample weights to account for the potential peculiarities of the final sample of 9550 responses.

Details of the survey and a descriptive analysis of variables describing the context of invention processes in Europe are presented in Giuri *et al.* (2007).[1] In this paper we focus on the private value of the patents held by firms. Thus, all the samples that we use in our analyses rule out all the PatVal-EU patents held by universities, individuals or non-profit institutions. A significant effort in building this dataset is that we consolidated all the firms according to their ultimate parents by using *Who Owns Whom* (several years).

3.2 Dependent Variable

The dependent variable of the equations estimated in this paper is a measure of the present economic value of patents as an asset. Following Harhoff et al. (2003a), we obtain this variable as the inventor response to this question: "*Suppose that on the day in which this patent was granted, the applicant had all the information about the value of the patent that is available today. In case a potential competitor of the applicant was interested in buying the patent, what would be the minimum price (in Euro) the*

[1] Giuri et al. (2007) only report about patents in six countries. Data about Denmark and Hungary were collected later.

applicant should demand?" The PatVal-EU survey offered a menu of ten interval responses: less than €30K; 30-100K; 100-300K; 300K-1M; 1-3M; 3-10M; 10-30M; 30-100M; 100-300M; more than 300M. We show in Gambardella et al. (2008) that the distribution of this measure is skewed to the left, and in general it conforms to other distributions of the value of patents in the literature (Harhoff et al., 1999; Scherer and Harhoff, 2000; Scherer et al., 2000). Most importantly, as noted, we show that it is highly correlated with some standard indirect indicators of patent value.

In spite of the shape of its distribution, and the consistency of its ordinal ranking with the indirect indicators, we still need to address the problem that a self-reported variable may be correlated with some common factor that also affects the responses to other self-reported measures from the survey that we use as covariates in our regressions. This may produce a spurious correlation with patent value. For example, creative activities, like innovation, often stir up enthusiasm and motivation. More motivated or enthusiastic inventors may then declare both high motivations and higher patent values.

In the next section we therefore retrieve the expected value of our measure from regressing it on a large set of indirect indicators of patent value. Since the indirect indicators are not self-reported, but they are produced and validated independently of the inventor assessments, this expected value captures only the component of patent value correlated with the independent indicators. We use this expected value as the dependent variable of our regressions that estimate the determinants of patent value. As noted in the introduction, this amounts to creating a composite indicator of patent value from an array of indirect indicators, using the inventor assessments as a monetary anchor for measuring this indicator. Our approach is the same as Lanjouw and Schankerman (2004), who develop a composite indicator from some common indirect indicators. Lanjouw and Schankerman, however, do not have a basis to anchor their indicators, and thus they have to estimate it after making some assumptions about the structure of the error terms in the indicator equations.

3.3 A Composite Indicator of Patent Value

Table 1 defines all the variables employed in the analysis of this section. Table 2 presents the corresponding descriptive statistics.

TABLES 1 AND 2 ABOUT HERE

Define our self-reported measure as V^*. By using lower-case letters to denote logs, we assume that $v^* = v + \varepsilon + \eta$, where $v + \varepsilon$ is the "true" value of patent (net of any subjective assessment of the inventors in their questionnaire responses); $v \equiv E(v^* \mid \boldsymbol{x})$, with $\boldsymbol{x}$ being a vector of indirect indicators; ε is a stochastic term independent of any variable in $\boldsymbol{x}$; η is a stochastic term that encompasses several subjective assessments of the inventor, and it is also independent of v and ε. Another concern with our measure is that the inventors may inflate the value of their innovations, which amounts to saying that $E(\eta) > 0$. However, this will just be part of the constant terms of our regressions. As a result, we may be unable to make inferences about the actual magnitudes of our estimated patent values, but we can make inferences about elasticities.

The indicator equation that we estimate has the following form

$$v^* = \boldsymbol{x}\boldsymbol{\beta} + \varepsilon + \eta \qquad (1)$$

where $\boldsymbol{\beta}$ is a vector of parameters to be estimated. The vector $\boldsymbol{x}$ includes the logs of CITES, REFERENCES, CLAIMS, EQUIVALENTS, STATES;[2] the dummies OP, AP, APEX, ACCEX, PCT, OBS3PARTY; the dummies for the 8 countries where the first inventor of the PatVal-EU patents are located; and the dummies for the 30 industry classes in which our patents are classified.[3]

All the indicators above are potential indirect measures of the economic value of the patent. The variables CITES, REFERENCES, CLAIMS, and EQUIVALENTS are the same four indirect indicators of value used by Lanjouw and Schankerman (2004). The rationale for the number of designated STATES is similar to EQUIVALENTS in the context of European patents. The variable EQUIVALENTS is correlated with value because the patent holder will incur the higher costs of patenting the same invention in more administrations only if it is justified by a higher value. In the case of EPO patents applicants can decide in how many European States they want to patent their invention. Since the patent fee increases with the number of States in which they want to be protected, this must reflect a higher expected value from the patented invention. Similarly, there is a growing literature suggesting that patents that stir the attention of rivals, or other parties more generally, are more valuable. This is because the opposition to a patent or any other action has costs, and hence it will not be undertaken if the patent did not have much value. Thus, we expect a higher value of a patent if it was opposed (OP), if there was an appeal after the

[2] Whenever the minimum value of the variable was zero, we employed the log of 1 plus the variable.

[3] We retrieved our technology classes from the ISI-INIPI-OST concordance classification between patent IPC classes and the 30 technology classes elaborated by the German Fraunhofer Institute of Systems and Innovation Research (ISI), the French Patent Office (INIPI) and the Observatoire des Sciences and des Techniques (OST). See Giuri *et al*. (2007) for details. Our 30 technology classes are: Electrical devices, electrical engineering, electrical energy; Audio-visual technology; Telecommunications; Information technology; Semiconductors; Optics; Analysis, measurement, control technology; Medical technology; Organic fine chemistry; Macromolecular chemistry, polymers; Pharmaceuticals, cosmetics; Biotechnology; Materials, metallurgy; Agriculture, food chemistry; Chemical and petrol industry, basic materials chemistry; Chemical engineering; Surface technology, coating; Materials processing, textiles, paper; Thermal processes and apparatus; Environmental technology; Machine tools; Engines, pumps, turbines; Mechanical Elements; Handling, printing; Agricultural and food processing, machinery and apparatus; Transport; Nuclear engineering; Space technology, weapons; Consumer goods and equipment; Civil engineering, building, mining.

opposition (AP) or after the examination (APEX), if third parties provided observations prior to the grant of the patent (OBS3PARTY), if the applicant requested an accelerated examination (ACCEX), or if the patent was a PCT/WO patent (PCT), that is it was applied under the Patent Cooperation Treaty which establishes that the patent is filed with all the Contracting States rather than being a national patent that is then also applied for to the EPO.

Table 3 reports the results of this estimation. Since our PatVal-EU measure is divided in classes, we used an interval regression estimation, which is an ordered probit with known constants, where the constants are the boundaries of the measure. The regressions shows that our PatVal-EU measure of value is highly correlated with practically all the indicators employed in the analysis. From our estimation we retrieved the predicted value $v \equiv E(v^* \mid \boldsymbol{x}) = \boldsymbol{x\beta^*}$, where $\boldsymbol{\beta^*}$ is the set of estimated parameters. This is the dependent variable that we employ in our analysis. We plotted the distribution of this dependent variable which turns out to be skewed and similar to the distributions of patent value in the literature.

TABLE 3 ABOUT HERE

4. ESTIMATED VALUE EQUATION

4.1 Equation and Key Covariates

Table 4 defines all the variables that we employ in this section. Table 5 provides the corresponding descriptive statistics. Compared to Tables 2 and 3, in Table 5 there are missing observations. We could have predicted patent value in the previous section by using the same observations employed here to estimate the determinants of value. However, there is no reason for discarding observations that are helpful to predict our expected value measure.

TABLES 4 AND 5 ABOUT HERE

The value equation that we estimate in this paper is

$$v = \nu_m \cdot m + \nu_n \cdot n + \nu_z \cdot z + controls + error \qquad (2)$$

where lower case letters denote logs, and ν_m, ν_n and ν_z are elasticities to be estimated. Equation (2) highlights the three focal variables of our analysis. First, we are interested in estimating the elasticity of the resources (man-months, M) invested in the project leading to the patented invention. As noted in Section 2, while there is an extensive literature estimating this elasticity at the aggregate firm-level, no study that we know of has estimated it at the patent-level. Second, the variable N measures the number of technically related patents. Since many patented inventions are technically related to other patents, the sole assessment of the value of an individual patent may be incomplete, and we need to better understand the value of the portfolio. Third, we are interested in understanding how the experience or talent of the inventor affects the value of patents, on top of the resources invested in the project and other controls. In (2) we label this variable generically as Z, and as we shall see below we proxy it by the past citations to patents of the inventor. In our regressions we control for the inventor age, academic degree, and the number of years in which the inventor was employed by his current employer. Thus, the variable that we are focussing on here is neither age-related experience, nor academic background or firm-specific experience.

While M and N come from our survey, we proxy for the inventor expertise Z by her average citations to patents with priority three or more years before the focal patent. This was an extensive data collection going beyond PatVal-EU and retrieving all the EPO patents of our inventor before the PatVal-EU patent. We use a three-year lag to reduce as much as possible the links between the patents that we use as a basis for the citations to the inventor and the patents related to the focal patent. For example, patents with priority date closer to the focal patent may be connected patents or part of similar research investments. By contrast, we want to find a measure of the inventor ability or expertise as independent as possible from the focal patent. At the same time, we could not use longer gaps from the priority date (more than 3 years) because our sample includes inventors in the early stages of their career. With longer lags the number of patents may be too correlated with inventor age or the time since she started working. We prefer citations to the number of past patents of the inventors because the number of patents may be affected too much by the inventor availability of resources in the past. By contrast, citations are a better proxy for ability or experience related to successful events. As a robustness check, we run all our regressions using the maximum number of citations received by any patent of the inventor with priority three or more years before the focal patent. The results do not change.

4.2 Instruments

We treat man-months (M), connected patents (N), and the inventor average past citations (Z) as endogenous, and therefore we instrument for them. Man-months is endogenous because firms may increase investments in an innovation project after they obtain initial signals that the project may lead to valuable inventions. The number of connected patents is endogenous because a more valuable patent may produce technically related patents. Average past citations are potentially endogenous for at least two reasons. First, unobserved inventor ability may affect both the value of the current patent and her average past citations. Second, in spite of our caveats to pick a variable as independent as possible from the current patent, even patents with priorities three years or earlier may not be totally unrelated to the focal patent. Sometimes the inventor research trajectories start many years earlier (e.g., during their PhD or

Master thesis), and given the length of the innovation process there can be unobserved characteristics in these past patents that also affect the current patent. Thus, for example, unobserved factors that raise the success of past patents (which are then cited more) may affect the current patent as well.

We employ four instruments: SERENDIPTY, GOVFUND, RD, INTRATECY. Originally, we also tried other instruments, like the total firm sales or a dummy for whether the inventor was a specialized, professional inventor or someone who is normally employed in other activities by her organization. However, our tests for weak instruments or for their over- or under-identification (see next section) suggest that we focus on these four variables.

The variable SERENDIPITY is a natural instrument for our scopes. Unplanned inventions affect both M and N. As noted in the definition of SERENDIPITY in Table 4, the "strike of genius" that leads to serendipituous inventions was not furthered in a formal project. This suggests that it is not correlated with information that may have become available about the invention or about its value. The lack of further investments also makes it unlikely that serendipituous inventions give rise to strategic development of other patents, or to investments in extending the basic invention. At the same time, there is no reason to believe that SERENDIPITY has a direct impact on patent value once we control for inventor characteristics or the other controls that we use in our regressions. Finally, SERENDIPITY may capture the case of an inventor who is not a specialized inventor or someone who does not have systematic access to resources for invention. It may then be correlated with the inventor past citations.

The variable GOVFUND reduces the costs of projects. Subsidies to R&D are common in Europe, both from the European Union and the national or even regional European governments. The availability of these funds stems from policy decisions that depend on macro-economic or industry conditions, or other broad policy goals (e.g., stimulate innovation and growth in a certain area or industry, or for certain types of firms). This makes the creation of these funds exogenous with respect to the individual firms or the value of their patents. Also, in most of the cases, government funding stems from applications to research programs or other more automatic forms (e.g., R&D tax credit). This means that once the government supports are awarded, the monetary sums involved do not change if news about the value of the project becomes available. In this respect, government funding is different from venture capital funding, in which venture capitalists subordinate new injections of capital to previous milestone results. For example, it is rare that firms return the government funds if the project does not produce the expected results, or obtain additional money from the government if it is going better than expected. Government funds are also unlikely to have a direct impact on patent value. Money is fungible, and hence we do not expect that these funds make a difference vis-à-vis the use of other funds. At the same time, the lower project costs associated with these funds translates into higher M or N. Similarly, these funds mirror the fact that the firm in which the inventor is employed or the region in which she is located tend to use them more frequently. Since there is limited mobility among the PatVal-EU inventors (85% have never changed job for the ten years centerd around the priority date of the focal patent), the implied lower project costs may affect their past performance as well.

Total firm RD (and the dummies for missing RD, see Table 4) is correlated with fixed RD costs that encourge both M and the development of connected patents N. For the same reason suggested above, present RD budgets may proxy for the steady flow of RD budgets of the firms over time. Especially in the cross-section, this proxies for firm-level resources that may affect the inventor past citations. At the same time, we posit that total firm RD does not directly affect the specific patent project that constitute our unit of observation. Any firm-level effect on the specific project comes through the resources invested in the project, M, or through N. In particular, N controls for the effects of related activities leading to connected patents, spillovers, fixed assets in related projects, or diseconomies across them. Moreover, we employ other firm-level controls in our regressions. In particular, the specialization index RTA of the company in the technological class of the patent and the dummies for firm size control for broader effects across unrelated projects of the firm.

Finally we employ a 3-year average of the interest rate in the country of the inventor since the priority year of the patent (INTRATECY). This measure varies across countries and over time. A higher interest rate, which is clearly exogenous to the firm, discourages investments in innovation, and therefore both M and N, while it should not have any special effect on Z.

4.3 Controls

We first control for several characteristics of the inventor. Apart from dummies for AGE and EDUCATION, the variable YEARINORG accounts for the impact of the experience of the inventor inside the organization. The variable MOTIVATION proxies for the motivation of the inventor (Cohen and Sauermann, 2010). The dummy MALE controls for gender.

Our controls for the innovation project are the 30 technology class dummies, the dummies for the priority year of the patent, the dummies for the country in which the patent was invented, and a dummy for whether the invention is produced in a city larger than 500K people (BIGCITY). We use dummies for the size of firm consolidated at the level of its ultimate parent: SMALL_PARENT and MEDIUM_PARENT account for ultimate parents with up to 100 and between 100-250 employees, while larger firms are captured by the default dummy. We also control for the relative specialization of the firm in the technology class of the patent. The index RTA is the share of PatVal-EU patents of the consolidated firm in the technology class of the focal patent relative to the share of total patents in that technology class.

Since PatVal-EU mirrors the distribution of patents in the population, the use of the PatVal-EU patents to compute the RTA should not introduce any particular bias.

We employ three controls for the value associated with the patent protection of the invention. As discussed in Section 2 the value of a patent also depends on the extent to which a private protection enhances the benefits accruing to the firm – the so-called "patent premium." Our first control is PROTECTION, which is the sum of two survey questions aimed at assessing the extent to which the invention was patented for reasons related to its protection – see Table 4. The second control is IPC4_NOFIRMS. The rationale for this control is that a higher share of patents held by universities, public research institutions or individuals produces a greater diffusion of knowledge in the technological class, which reduces the value of the patent. In part this is a general effect due to a greater competition faced by each technology holder in the area because academic and non-profit institutions, as well as individuals, have a greater incentives to diffuse the technology. The same effect can be thought of, at least in part, as reducing the patent premium because the greater diffusion of knowledge is likely to impinge on the specific knowledge or technology associated with the patent of each firm. Another way to think about it is that this makes the protection provided by each patent in the class more narrow, which reduces the patent premium, and thus the patent value.

The third control, IPC4_COMP, is the complement to 1 of the Herfindhal index of the share of patents held by different PatVal-EU applicants in the same International Patent Classification (IPC) class at 4-digits. Its rationale is that if the firm is the only patent applicant in the field, the patent premium is small because it is unlikely that other organizations have the assets or the absorptive capacity to copy or to use the unpatented technology. By contrast, if there are quite a few of these firms, the patent shelters from the use of the technology by others. The large sample of PatVal-EU applicants, along with the stratification of the PatVal-EU sample, which reflects the distribution of the population of patents, ensures that this variable is a good proxy of the Herfindhal that we would obtain if we employed the full population. Moreover, following our consolidation of the PatVal-EU applicants, the Herfindhal is computed by considering applicants to be different only if they belong to unaffiliated organizations. Finally, the use of IPC4_NOFIRMS ensures that we control for the extent to which IPC4_COMP is affected by the presence of applicants different from firms.[4]

[4] An alternative to IPC4_COMP would be a measure of the product market competition of the firm, but it is much harder to find information on product market competitors associated to the specific technological class of the patent.

5. EMPIRICAL RESULTS

5.1 Patent Value

Table 6 presents the first stage regressions of our three endogenous variables. In all the regressions in this and the following sections we employ robust standard errors, cluster errors for common parent companies, and weight for the oversampling of important patents in PatVal-EU.[5]

TABLE 6 ABOUT HERE

The table shows that our instruments are significantly correlated with the dependent variables. In particular, SERENDIPITY is negatively correlated with N and Z, suggesting that they are not furthered in bigger projects or that they are not produced by professional inventors or individuals specialized in inventing. The variable GOVFUND is positively correlated with N and M, which is consistent with our interpretation that they make the innovation process less costly. The firm's total RD is positively correlated to N, while, as expected, a higher INTRATECY is negatively related to investments in innovation, N and M. The positive correlation between INTRATECY and Z may stem from the cyclical nature of the macroeconomic fluctuations. A higher INTRATECY today may imply a lower one a few years earlier, which implies more valuable past patents that are then cited.

Tables 7 presents our OLS and GMM regressions for patent value. The OLS regressions produce a positive and significant impact of both N and M, whilst in the GMM regressions the impact of N is insignificant and the impact of M is slightly smaller, though still positive and significant. Thus, without instrumenting for N and M, we overestimate the impact of N and slightly overestimate the impact of M. OLS estimation may for instance accentuate the impact of spillovers from connected projects. The impact of Z is insignificant both in the OLS and GMM equation. Since we provide OLS results only for comparison, below we only discuss the GMM results.

TABLES AND 8 ABOUT HERE

Table 8 also shows that the instruments are relevant and exogenous. The F-statistics of all three first stage regressions are well above 10, the rule of thumb threshold commonly employed in these cases. The Kleigebergen-Paap statistics reject the nul hypothesis that the instruments are underidentified, and the comparison with the Stock-Yogo critical values rejects the nul that the instruments are weak. In addition, the Hansen J statistics does not reject the nul that the instruments are exactly (rather than over-) identified. All this suggests that the instruments that we picked are relevant and exogenous.

[5] The weight is the inverse of the relative shares of important or non-important patents in the sample and in the population of EPO patents with the same priority dates as the PatVal-EU sample.

The elasticity of M with respect to patent value is small, 4%, though quite well measured. Most of the controls in Table 7 are insignificant. The only significant ones are our two proxies of the patent premium, PROTECTION and IPC4_COMP. Thus, after controlling for technologies and countries, patent value is determined by the resources invested in the project, which affect the value of the invention, and the patent premium.

5.2 The Value of a Portfolio of Connected Patents

To understand some of the more complex structural factors that affect the determinants of patent value, we look at the value of the entire portofolio of connected patents, that is $\widetilde{V} = \sum_{i=1}^{N} V_i$, where V_i is the value of the ith patent in this set. Since the specific PatVal-EU patent can be thought of as a randomly chosen patent from this set, we can take its value to be equal to the average of all patents in the set times a random error. This enables us to write the expected value of the whole portfolio as $N{\cdot}V$ where V is the value of the PatVal-EU patent.

We develop a structural estimation of the number of connected patents N. The PatVal-EU survey provides information on the total man-months invested in the whole set of interconnected patents, MF. Therefore, along with M and Z, we introduce MF as an additional determinant of N, and we treat all three variables as endogenous. We employ the same log-log functional form as in (2) and estimate

$$n = n_{mf}{\cdot}mf + n_m{\cdot}m + n_z{\cdot}z + controls + error \qquad (3)$$

where lower-case letters denote logs and n_{mf}, n_m and n_z are elasticities to be estimated. In addition, we estimate two equations for M and MF in the same log-log form as functions of Z and controls, and we treat Z as endogenous. All three equations are estimated by GMM using the same covariates and instruments employed in the value equation, clustered by parent firms. The results of this estimation are in Table 9. Table 10 also shows that for the N-equation, the Kleigebergen-Paap statistics reject the nul hypothesis that the instruments are underidentified, and the Hansen J statistics does not reject the nul that the instruments are exactly identified. The Stock-Yogo critical values rejects the nul that the instruments are weak, though at higher levels of the IV relative bias than the value equation in the previous section. The F-statistics of all three first stage regressions are well above the rule of thumb threshold of 10.

TABLES AND 10 ABOUT HERE

Table 9 shows that N increases with the total resources invested in the set of connected patents MF. It also decreases with M. This is natural because an increase in M given MF implies that the firm invests more resources in the focal patent and thus reduces the investments in related patents. We can then compute the elasticity of labor for the whole value of the set of connected patents; that is, the elasticity of labor when we allow for the number of patents to change as well. This can be written as

$$d\log(NV) = n_m d\log(M) + n_{mf} d\log(MF) + v_n d\log(N) + v_m d\log(M)$$

where the first two terms account for the impact of a change in man-months on *N* and the other two terms account for the impact of a change in man-months on *V*, which is composed in turn of the impact coming through *N* and the direct impact on *V* computed in the previous section. Thus, all the terms in the expression above, but the last one, capture the effects associated to a change in *N*.

As a benchmark, assume that the focal patent proxies for the average investment in man-months in each patent in the set. This implies $MF = N \cdot M$, and thus $d\log(MF) = d\log(M)$. Moreover, $d\log(N)$ is equal to the first two terms of the expression above. We can then write

$$\frac{d\log(NV)}{d\log(MF)} = \left(n_m + n_{mf}\right)\left(1 + v_n\right) + v_m \qquad (4)$$

where again $(n_m + n_{mf})$ captures the effect on *N*, $(n_m + n_{mf})v_n$ captures the total effect of *N*, $(n_m + n_{mf})v_n + v_m$ captures the total effect on *V*, and v_m is the original elasticity of 4% on *V* discussed in the previous section. By using the estimated parameters in Tables 7 and 9 we obtain that $\frac{d\log(NV)}{d\log(MF)} = [(-1.034 + 1.238)\cdot 1.016 + 0.040] = 0.247$. Of this total elasticity, the effect on *N* is $(n_m + n_{mf}) = 0.204$, which is 82.5% of the whole effect. We will discuss the rationale and implications of this difference between the impact on *N* vs *V* in the next section.

Table 9 also reports a negative and significant impact of *Z* on *N*. We will interpret this result together with the results about the impact of Z on *M* and *MF* below. Here we can anticipate that this suggests that inventors with more past citations have achieved some important results in the past. They are then likely to pursue incremental trajectories on their past achievements, which produces some sort of exhaustion effect on the current number of patents. It is also interesting that this effect does not come in the form of a reduction in the average value of their patents, as implied by the fact that both *N* and *Z* are insignificant in the value equation of Table 7, but in the form of fewer patents.

The other controls in the *N*-equation indicate that AGE has a slightly curvilinear effect; it increases *N* up to the 40-50 age category, and then declines. PhDs have a positive and significant effect on *N*. Of the other controls, IPC4_NOFIRMS has a negative and significant effect. This mirrors the fact that when there are many universities and independent inventors there is greater diffusion of knowledge, which makes it harder to patent, or the presence of a culture that discourages private appropriation of knowledge. The specialization of the firm in the field, RTA, and the externalities from larger cities, BIGCITY are positive and significant, which speaks to the importance of these two factors for the patenting productivity of firms.

5.3 Patent Values Produced by Experienced Inventors

Past Citations. The other two columns of Table 9 show that inventors with higher Z invest fewer resources in the production of new patents. We find that inventors with past successful experience: (1) do not produce patents of higher value, as implied by the insignificant impact of Z on V in Table 7; (2) they produce fewer patents, as implied by the negative and significant impact of Z on N in Table 9; (3) they employ far fewer resources than other inventors, as implied by the impacts of Z on M and MF. This suggests that these inventors are continuing trajectories initiated in the past, and for which a good deal of fixed costs are sunk. In other words, they opened a trajectory and are pursuing it, with merely incremental man-months needed to produce new patents. This is not inconsistent with what we know about the innovation process. Many researchers exploit for a long time initial succeses (discoveries, breakthroughs), and sometime these successes date back even to early days of their career.

These results could also be explained by the fact that given important past discoveries, firms produce strategic patents, which call for small additional efforts. However, strategic patents are normally applied jointly and around the same time. One reason for taking citations to patents with priority date three or more years earlier is to avoid such more spurious correlations. Ultimately, however, we cannot really distinguish whether our variable Z is capturing the effects of current strategic patents that enhance the citations of some more basic past invention or of current incremental inventions. What we can say, however, is that they flag the presence of some more basic invention in the past, and additional patents ensuing from that invention require fewer resources to be produced.

In order to weigh the importance of these effects we can compare the impact of differences in the inventor's past citations on the total value of patent portfolios, $N \cdot V$, and on the investment in man-months. As far as the former is concerned, the total elasticity of Z on $N \cdot V$ can be written as

$$\frac{d\log(NV)}{d\log(Z)} = \left(n_m m_z + n_{mf} mf_z + n_z\right)\left(1 + v_z\right) + v_m m_z + v_z$$

where m_z and mf_z are the estimated elasticities of Z on M and MF. By using our estimated elasticities in Tables 7 and 9, $(n_m \cdot m_z + n_{mf} \cdot mf_z + n_z) = -0.664$ and $(n_m \cdot m_z + n_{mf} \cdot mf_z + n_z) \cdot v_z + v_m \cdot m_z + v_z = -0.244$, which produces a total elasticity of –0.908. Note that the effect produced by changes in Z on N is again more pronounced than that on V. The elasticity of Z on MF is much larger than that of Z on $N \cdot V$ in absolute terms, –3.836. In sum, other things being equal, inventors with twice as many past citations will produce portfolio of patents whose value is roughly 90% lower. However, to produce this value they will use almost 400% fewer man-months. This is consistent with our two earlier conjectures: a) inventors with more past citations exploit incremental trajectories that build on previous discoveries; b) this implies lower value, but also fewer resources as the fixed investment for this trajectory is sunk, and new inventions come at low marginal costs.

Another way to see the same thing is that given any two levels of Z, i.e., Z_1 and Z_0, the difference in the values $(NV)_1$ and $(NV)_0$ of the patent portfolios produced by two identical inventors who only differ because of Z_1 and Z_0 is $\frac{(NV)_1}{(NV)_0} = \left(\frac{1+Z_1}{1+Z_0}\right)^{-0.908}$. (Recall that in our estimated equations we use *log*(1+*Z*) because Z can be equal to zero.) The corresponding difference in the total man-months invested in the project, *MF*, is $\frac{(MF)_1}{(MF)_0} = \left(\frac{1+Z_1}{1+Z_0}\right)^{-3.836}$. As an illustration, compare the relative levels of *NV* and *MF* of two inventors who only differ because Z_1 is equal to the 75th, 90th, 95th, or 99th percentile of the sample distribution of *Z*, whereas Z_0 is equal to the median.[6] The relative values of *NV* are equal to, respectively, 55%, 38%, 31%, and 20%, while the relative values of *MF* drop rapidly to zero, i.e., 8%, 2%, 1%, 0%. Thus, very prolific inventors produce smaller values but at practically zero marginal cost. As noted earlier, it is not clear whether these additional patents are genuine incremental inventions or strategic patents. The essence of our finding however does not change. Whether an incremental innovation or a strategic patent, the value produced by a more cited inventor is smaller, but the cost of producing this value is even smaller.[7]

Age and Education. The effects of the AGE and EDUCATION dummies on *M* and *MF* are similar to their effects on *N*. The AGE dummies have a slight curvilinear effects, while higher education degrees, and PhDs in particular, exhibit higher impacts. The impacts of the AGE and EDUCATION dummies on *M* and *MF* are highly significant.

Like in the case of *Z*, we can compare the total value *NV* of the portfolio of patents of two inventors who differ only in their age class or education degree. Since the coefficients of the age and education dummies in the value equation (Table 7) are small and largely insignificant, we set them for simplicity equal to

[6] The median, 75th, 90th, 95th, and 99th percentiles of *Z* are respectively 0, 0.933, 1.875, 2.667, and 5.

[7] Note that our value measure would also reflect the value of a strategic patent. As noted in Section 3.2 we ask for the minimum price at which the owner of the patent would sell the patent at the moment of grant. This is a general measure of value, whether the value is produced by the profits of the invention, the strategic use of the patent, or other reasons.

zero. To assess the differential effects of any two age or education classes we then have to take into account their different impacts on *N* that come through *M* and *MF*, and their different impacts on *V* that come from *N* and *M*. Given the log-log expressions used in our estimations, the difference in $log(NV)$ for our two inventors is $[n_m \cdot (m_1 - m_2) + n_{mf} \cdot (mf_1 - mf_2) + (n_1 - n_2)] \cdot (1+v_n) + v_m \cdot (m_1 - m_2)]$, where n_m and n_{mf} are the estimated elasticities of *M* and *MF* on *N*, v_m and v_n are the estimated elasticities of *M* and *N* on *V*, n_1 and n_2 are the estimated coefficients of the dummies of the two age or education classes 1 and 2 in the *N* equation, and m_1, m_2, mf_1, mf_2, are the coefficients of the same dummies in the *M* and *MF* equations.

We find that other things being equal an inventor with a PhD produces a portfolio of connected patents whose total value $N \cdot V$ is 28% higher than that of an inventor with a college degree, and 61% higher than an inventor with a high-school degree. By comparison, relatively to an inventor younger than 30, an inventor in the age-range 30-40, 40-50, 50-60, and above 60 produces a total value that is, respectively, 2.7%, –0.08%, 5%, and 18% higher. This speaks of the notable premium produced by education. A PhD (or even college education) provides a potential productivity far higher than the one produced by the experience offered by the entire career of an individual. An interpretation of this result is that an inventor with a PhD will gain a lower productivity potential over her entire career than the productivity potential that she gained during the years leading to her PhD degree.

Other Regressors in the M and MF Equations. The other regressors in the *M* and *MF* equations confirm that MOTIVATION is an important determinant of effort. The MOTIVATION variable has a positive and quite significant impact on both *M* and *MF*. We also find, as expected, that an inventor with a longer tenure within the organization exhibits a higher investment in man-month associated to the whole set of connected patents. The experience within the organization encourages the inventor to use inputs and to produce outputs in several related directions and targets. The lower the share of patents in the technological class held by universities or individuals the higher the investments in *M* and *MF*. Like in the case of *N*, a lower IPC4_NOFIRMS may denote fewer external spillovers, which forces firms to substitute such a knowledge with internal efforts. Similarly, the more concentrated is the ownership of patents in the class, the stronger the internal effort, particularly *MF*. Again, this probably mirrors the lack of spillovers because of the limited number of independent sources of knowledge in the field. This is also consistent with the positive and significant impact of BIGCITY on *MF*. The externalities associated with larger metropolitan areas may encourage greater internal efforts for several reasons – for example, the effect of these spillovers is higher for organizations with greater absorptive capacity, or the spillovers encourage additional investments because of the opportunities that they create.

Finally, PROTECTION and RTA are correlated with *MF*. The positive correlation of the former may mirror the fact that protection comes with a greater effort to fence the terrain with related patents. As noted our analysis cannot distinguish whether portfolio of patents stem from genuinely connected inventions or strategic patenting. This particular result would then suggest that we cannot rule out strategic patenting completely. The negative correlation of RTA suggests that the focal PatVal-EU patents may be itself a connected patents of some inventions in other core areas in which the firm is specialized. Thus, the less the firm is specialized in the area of the focal patent, the more this patent is associated to a man-month effort that involve other patents in other domains.

6. DISCUSSION AND CONCLUSIONS

This paper provides three main contributions. First, we show that the resources invested in the innovation process can more effectively raise the number of patents produced with respect to the expected value of the individual patent. We know that the value of innovation is bound by uncertainties. This is largely because the economic value of innovations depends on several factors, many of which out of the control of the inventing firms or organizations. This is a point made long ago by scholars like Nathan Rosenberg. For instance, Rosenberg (1982) noted that the full economic value of innovations depends on the presence of complementary innovations or the vagaries of demand. At the same time, the technical value of the invention only explains a modest part of its economic success. More recently, Audia and Goncalo (2007), Mariani and Romanelli (2007), and Conti et al. (2011 find that firms can better control the quantity of their inventions vis-à-vis their qualities, measured for instance by citations. They also attribute this to the greater difficulties of predicting quality, which – compared to quantity – depends on many factors outside the control of the firm.

This suggests, for example, that it is not an effective innovation strategy to insist on one specific invention by concentrating resources on it. It is better to spread resources across technically related inventions. On the one hand, this is not going to reduce the average value of the patents in the portfolio, as implied by our empirical result that the number of technically connected patents does not have a significant impact on it. On the other hand, this means that the addition of patents to the portfolio raises its total value by the average patent value.

Our second result is that inventors with higher past citations produce portfolios of patents of lower value. We use the average citations to inventor patents with priority date three or more years before the focal patent. This puts enough distance between the focal patent and the past patents of the inventor to limit correlation due to the fact that the focal patent cites the past patents because they are part of the same project. At the same time, the lag cannot be longer because it would then proxy mainly for the age of the inventor. However, even patents that date some years back can be earlier steps of the same trajectory of research of the current patent. For some inventors this trajectory is as long as their whole career. Thus, we treat past citations as endogenous to control for the possibility that the value of the current patent or the number of patents in the portfolio affect the citations of past patents. The subtle issue here is that citations from the current patent to past patents of the inventors may denote both inventor capabilities or persistence of research projects. In fact, even if the research projects are persistent, citations to the inventor past patents are correlated with their quality. In the end, we cannot distinguish the extent to which such citations denote inventor quality or persistence of research. However, since the two factors would have opposite effects, the negative effect that we find suggests that persistence of research dominates. Thus, inventors with more prominent past successes focus on the exploitation of these successes. This implies more incremental innovations of lower value than inventors that have not had similar successes, after controlling for age, education and several other factors.

This interpretation is both reinforced and qualified by our result on the effects of the inventor past citations on the man-months invested in the individual patent and in the whole portfolio of technically related patents. First, we find that while inventors with more past citations produce lower values, they also invest fewer man-months in the project producing the entire patent set. This strengthens our hypotheses about exploitation and incremental innovations. Basically, the big investment in the research trajectory is sunk, and small additional investments produce new patents. We find that inventors with many past citations employ almost no additional man-months to produce new patents, which suggests that some of these new patents may even be small variations on the basic inventions to enhance protection (strategic patenting). Second, we find that the reduction in man-months from highly cited inventors is stronger for the man-months invested in the focal patent than in the whole set. This suggests that highly cited inventors spend very little time on individual patents, while not reducing as much their effort to produce several patents. This is again consistent with the view that highly cited inventors exploit their fundamental knowledge, inventions, or capabilities to produce many patents, whether incremental inventions, strategic patents, or both.

Our third contribution is about the effects of age and education. First, as discussed, we find that the distance in terms of additional value produced between a PhD and a college graduate or between a college and a high school graduate is higher than the difference in value produced by two inventors at the beginning or the end of their career. Second, our first stage regression for past citations indicates that older and more educated inventors are more likely to have more citations. This suggests that age and education compensate the exhaustion effect produced by past citations. In other words, older or educated inventors are likely to produce more incremental innovations (of lower value) because they are more likely to have more past citations. However, age and education imply that they produce higher values, thus compensating the lower values of incremental innovations. Most interestingly, age and education raise the man-months invested by the inventors. This is therefore the mechanism by which they compensate the reduction in value associated with more incremental innovations. Higher man-months raise both the average value and the number of innovations produced along their trajectory.

Even more interestingly, we can predict that older or educated inventors without significant past inventions are most likely to produce higher values. This is because age and education imply that they invest more man-months and the lack of previous breakthroughs means that they spend most of their time in finding a new trajectory than focussing on the incremental innovations along an existing path. As an illustration, high past citations means that inventors have located an ore mine. They will then extract mineral ores of lower value, but more quickly, and if they are older or more educated they spend more time on it. Thus, the combination of higher productivity and intensity of work enable them to extract minerals of higher values and more of them, which makes them ammass quite some value. Older or educated inventors without past citations have not yet found their vein. However, they work harder and are more productive, and therefore we expect that the values that they will extract in the near future will be the highest because they will most likely start to exploit a virgin seam.

Our paper has limitations. First, our measure of the number of technically related patents is self-reported by the inventors. We have a similar problem with our measure of value, however in that case we anchor it to some indirect indicators. Future research may try to develop new measures derived from patent data. The reason why we did not pursue this route is that we want to focus on the set of technically connected patents coming out of a given research project as identified by the inventor. There is no way to identify these patents without some contributions from the inventor in a survey like the one that we have undertaken. In principle, we could have asked the inventor to give us the patent numbers of these technically related patents. But this would have increased dramatically the time and effort of responding to the survey, with several other drawbacks, particularly many more missing responses to this question. A feasible alternative would be to focus on fewer patents or inventors and ask them to identify the patent numbers of technically connected patents. In this research we resolve this trade-off in favor of having a large scale survey and many observations for our regressions. Since to our knowledge this is one of the first times that the problem of the size vs quality of patent portfolios is tackled we prefer to start from the large-scale end of the question, and leave the more refined, smaller scale studies to future research.

The use of all the patents of the inventor within a small interval of time would also introduce biases. For example, prolific inventors may run many projects concurrently, and we would then confound their prolificacy across projects with that within projects. In addition, technically related patents may not be all applied within small intervals of time. Alternatively, we could have looked at patents where our inventor worked with a similar team of inventors or in similar classes or technologies. But again, apart from the effort that this would involve at the scale that we want to preserve in this study, the same teams could work on different projects or technologies; at the same time, the same project could produce patents in different areas (e.g., if technologies are "broad"). In sum, asking the number of technically related patents to the inventors may not have more problems than alternative measures.

A second limitation is that we do not have a measure of the man-months involved in each patented invention of the patent portfolio, nor we have a measure of the value of each connected patents. Moreover, it is not clear at all that each patent correponds to an invention, nor we know how many inventions correspond to the full set of patents in the portfolio. In short, it is hard to say how the total man-months declared by the inventors for the portfolio of patents is divided up among the inventions that constitute the portfolio. The assumption that we make in this paper is that the man-months declared for the focal patent in our analysis is a good proxy for the average man-months for each invention in the set, and similarly we assume that the value of the focal patent is a measure of the average value of the patent in the set up to an error term.

We have a good justification for these assumptions. Most notably, given the design of our survey, the specific patent that we look at is a random patent of the set, and thus we can take its man-months or value as being the average for the set up to an error. More importantly, we need to understand what we miss from being unable to measure the man-months or value of each patent in the portfolio. First, we are unable to test hypotheses about the allocation of the man-months allocated to the set across specific inventions that compose it. We generically find that spreading man-months across more patents tends to be valuable, as the marginal product of labor is higher on the extensive (quantity of patents) than intensive (average value of patents) margin. But we cannot estimate for example the optimal size of the patent portfolio from alternative allocations of a given amount of labor allocated to the whole set. Similarly, we cannot estimate the impacts of our covariates on the value of the marginal patent in the set. In addition, we have no measures of capital assets for producing patents, and we rely on company characteristics for it. It is therefore not clear that our estimates of the elasticities of man-months refer specifically to man-months, or they combine the effects of other assets that are needed to produced patented inventions.

Finally, as noted, we cannot distinguish the extent to which past citations control for some measure of inventor ability or the exhaustion of opportunities. We tried alternative measures or regressors, like past citations net of self-cites or introducing as a covariate a PatVal-EU response to the question whether the patented invention built on other inventions of the inventor developed within the firm. The results do not change and broadly speaking these other variables did not enable us to disentangle inventor ability from a greater focus on incremental trajectories.

Our justification for these limitations is that this is one of the first times that we can address detailed questions about the determinants of patented inventions that go beyond measures of outcome like forward citations or other similar measures. Our proxy for value combines the information contained in several indicators that have been commonly used by the literature to measure patent performance. Moreover, we are one of the first study to address the problem of the value of a set of connected patents, thus moving away from the perspective of one patent being equivalent to one invention. Last but not least, our regressions employ several controls, many of which largely new to this literature. We look forward to future contributions that will resolve the problems that we are unable to address in this research.

References

Arora, A., Ceccagnoli, M., and W.Cohen (2008). "R&D and the Patent Premium," International Journal of Industrial Organization 26 (5), 1153-1179

Audia, P.G. and J.A. Goncalo (2007). "Success and Creativity over Time: A Study of Inventors in the Hard-Disk Drive Industry," Management Science 53(1), 1-15.

Bessen, J. (2008). "The Value of US Patents by Owner and Patent Characteristics", *Research Policy* 37 (5), 932–945.

Bessen, J. (2009). "Estimates of Patent Rents from Firm Market Value," *Research Policy* 38(10), 1604-1616.

Cohen, W. and H. Sauermann (2010). "What Makes Them Thick? Employee Motives and Firm Innovation," *Management Science* 56 (12), 2134-2153.

Conti, R., Gambardella, A., and M. Mariani (2011) "Learning to Be Edison: Individual Inventive Experience and Breakthrough Inventions," draft, Bocconi University.

Ernst, H., Leptien, Ch., and J.Vitt, (2000). *Inventors are Not Alike. The Distribution of Patenting Output Among Industrial R&D Personnel*, IEEE Transactions on Engineering Management, 47 (2), 184-199.

Gambardella, A., Harhoff, D., and B. Verspagen (2008). "The Value of European Patents," *European Management Review* 5(2), 69-84.

Gittelman, M. and B.Kogut (2003). "Does Good Science Lead to Valuable Knowledge? Biotechnology Firms and the Evolutionary Logic of Citation Patterns," *Management Science* 49 (4), 366-382.

Giuri, P., Mariani, M., Brusoni, S., Crespi, G., Francoz, D., Gambardella, A., Garcia-Fontes, W., Geuna, A., Gonzales, R., Harhoff, D., Hoisl, K., Lebas, C., Luzzi, A., Magazzini, L., Nesta, L., Nomaler, O., Palomeras, N., Patel, P., Romanelli, M., and B.Verspagen (2007). "Inventor and Invention Processes in Europe. Results from the PatVal-EU Survey," Research Policy 36 (8), 1107-1127.

Griliches, Z. (1981). "Market Value, R&D and Patents," *Economics Letters* 7, 183-187.

Hall, B., Jaffe, A. and M. Trajtenberg (2005). "Market Value and Patent Citations," *RAND Journal of Economics* 36, 16-38.

Harhoff, D., Scherer, F., and K. Vopel (2003a). "Exploring the Tail of the Patent Value Distribution," in: O. Granstrand (ed.), *Economics, Law and Intellectual Property: Seeking strategies for research and teaching in a developing field*. Kluwer Academic Publisher, Boston/Dordrecht/London, 279-309.

Harhoff, D., Scherer, F., and K. Vopel (2003b). “Citations, Family Size, Opposition and the Value of Patent Rights - Evidence from Germany,” *Research Policy*, 32, 1343-1363.

Harhoff, D., Narin, F, Scherer, F., and K. Vopel (1999). “Citation Frequency and the Value of Patented Innovation,” *Review of Economics and Statistics*, 81 (3), 511-515.

Hausman, J., Hall, B. and Z.Griliches (1984). “Econometric Models for Count Data with an Application to the Patents-R&D Relationship,” *Econometrica*, 52 (July), 909-937.

Lanjouw, J. and M. Schankerman, (2004). “Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators,” *Economic Journal*, 114, 441-465.

Lotka, A.J. (1926). *The Frequency Distribution of Scientific Productivity*, Journal of the Washington Academy of Science, 16 (2), 317-323.

Mariani M. and M. Romanelli (2007). “Stacking and Picking Inventions: The Patenting Behavior of European Inventors,” *Research Policy* 36(8), 1128-1142.

Pakes, A. (1985). “On Patents, R&D, and the Stock Market Rate of Return,” *Journal of Political Economy* 93 (2), 390-409.

Rosenberg, N. (1982). *Inside the Black Box: Technology and Economics*, Cambridge University Press, Cambridge UK.

Schankerman, M. and A.Pakes (1986). “Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period,” *Economic Journal*, 97, 1-25.

Scherer, F. and D. Harhoff (2000). “Policy Implications for a World with Skew-Distributed Returns to Innovation,” *Research Policy,* 29, 559-566.

Scherer, F., Harhoff, D. and J. Kukies (2000). ”Uncertainty and the Size Distribution of Rewards from Technological Innovation,” *Journal of Evolutionary Economics*, 10, pp. 175-2000.

Serrano, C. (2006). “The Market for Intellectual Property: Evidence from the Transfer of Patents.” Working Paper, University of Toronto, 2006.

Trajtenberg, M. (1990). “A Penny for Your Quotes: Patent Citations and the Value of Inventions,” *RAND Journal of Economics*, 21, 172-187.

Ziedonis, R.H. (2004). “Don’t Fence Me In: Fragmented Markets for Technology and the Patent Acquisition Strategies of Firms,” *Management Science* 50(6), 804-820.

Zucker, L. Darby, M. and J. Armstrong (2002). “Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology,” *Management Science*, 48 (1), 138-153.

Table 1: Composite indicator of patent value, definition of variables

Variable	Definition
V*	PatVal-EU index equal to 1-10 for the following classes of patent values: ≤€30K; 30-100K; 100-300K; 300K-1M; 1-3M; 3-10M; 10-30M; 30-100M; 100-300M; ≥300M
CITES	# of patent citations within 5 years
REFERENCES	# of patent references
CLAIMS	# of claims at the date of grant

STATES	# of designated EPC countries
EQUIVALENTS	# of equivalent patents
OP	Dummy = 1 if opposed following grant
AP	Dummy = 1 if appealed after opposition
APEX	Dummy = 1 if appealed after examination
ACCEX	Dummy = 1 if requested for accelerated examination
PCT	Dummy = 1 if PCT/WO application
OBS3PARTY	Dummy = 1 if observations by 3rd parties prior to grant (Art. 115 EPC)
DE	Dummy = 1 if inventor located in Germany
DK	Dummy = 1 if inventor located in Denmark
ES	Dummy = 1 if inventor located in Spain
FR	Dummy = 1 if inventor located in France
IT	Dummy = 1 if inventor located in Italy
NL	Dummy = 1 if inventor located in Netherlands
UK	Dummy = 1 if inventor located in UK
PRIORITY_93	Dummy = 1 if priority year = 1993
PRIORITY_94	Dummy = 1 if priority year = 1994
PRIORITY_95	Dummy = 1 if priority year = 1995
PRIORITY_96	Dummy = 1 if priority year = 1996
PRIORITY_97	Dummy = 1 if priority year = 1997

(+) For the first class mid-point b/w log(1) and log(30); for the last class mid-point b/w log(300,000) and log(1,000,000)

Table 2: Composite indicator of patent value, descriptive statistics

Variable	mean	Sd	min	Med	max	N. obs.
v^{*} (+)	6.018	2.247	1.701	6.306	13.214	6974
CITES	1.518	2.295	0	1	40	6974
REFERENCES	4.398	2.234	0	4	18	6974
CLAIMS	10.648	6.854	1	9	131	6974
STATES	8.64	4.752	1	7	19	6974
EQUIVALENTS	6.569	4.653	1	6	44	6974
OP	0.111	0.314	0	0	1	6974
AP	0.038	0.191	0	0	1	6974
APEX	0.002	0.041	0	0	1	6974
ACCEX	0.056	0.229	0	0	1	6974
PCT	0.351	0.477	0	0	1	6974
OBS3PARTY	0.007	0.081	0	0	1	6974
DE	0.407	0.491	0	0	1	6974
DK	0.006	0.075	0	0	1	6974
ES	0.014	0.119	0	0	1	6974
FR	0.133	0.34	0	0	1	6974
IT	0.139	0.345	0	0	1	6974
NL	0.131	0.338	0	0	1	6974
UK	0.17	0.376	0	0	1	6974
PRIORITY_93	0.287	0.452	0	0	1	6974
PRIORITY_94	0.26	0.438	0	0	1	6974
PRIORITY_95	0.233	0.423	0	0	1	6974

PRIORITY_96	0.166	0.372	0	0	1	6974
PRIORITY_97	0.053	0.225	0	0	1	6974

(+) Mid-point of the log of the two boundaries of the PatVal-EU value class V*

Table 3: Predicting the composite indicator of patent value, interval regression estimation (Dependent variable: interval boundaries = log of the boundaries of V*)

Variable	Estimates	Variable	Estimates	Variable	Estimates
CITES	0.339	AP	0.387	LOG(SIGMA)	0.705
	0.000		0.032		0.000
REFERENCES	0.114	APEX	-0.692		
	0.095		0.098	*Statistics*	
CLAIMS	0.141	ACCEX	0.128	N	6974
	0.004		0.212	LOG-LIK	-6.66E+04
STATES	0.232	PCT	0.167	chi2	591.142
	0.000		0.019		
EQUIVALENTS	0.261	OBS3PARTY	1.014		
	0.000		0.002		
OP	0.072	CONSTANT	4.680		
	0.487		0.000		

All variables (but dummies) in logs; log(1 + variable) if the variable can take value 0; p-values below estimates; equations include country dummies, dummies for technological sectors and patent priority years; observations clustered by ultimate parent company; regression includes weights to adjust the oversampling of important patents in PatVal-EU (important patents = cited at least once or opposed).

Table 4: Value regressions, definition of variables

Variable	Definition
V	Exp(v), where v = predicted value of the interval regression, which is then used as dependent variable of the value regressions
M	Mid-point of the man-month intervals in the PatVal-EU questionnaire required for producing the patented invention (≤1; 1-3; 4-6; 7-12; 13-24; 24-48; 48-72; ≥72)
N	Mid-point of the # of patents that "crucially depend on each other in terms of their value, or in a technical way"(+) (includes the focal patent) (1; 1-2; 3-5; 6-10; 11-20; ≥ 20). This is the # of patents in what we labelled the *portfolio of technically related patents*.
Z	Average # of citations within 5 years to patents of the inventor with priority year 3 or more years before the focal patent
MF	Mid-point of the man-month intervals in the PatVal-EU questionnaire required for producing the entire portfolio of technically related patents (≤1; 1-3; 4-6; 7-12; 13-24; 24-48; 48-72; 72-96; 96-120; ≥144)
SMALL_PARENT	Dummy = 1 if ultimate parent of the applicant firm is a small firm (≤ 100 employees)
MEDIUM_PARENT	Dummy = 1 if ultimate parent of the applicant firm is a medium firm (100-250 employees)

LARGE_PARENT	Dummy = 1 if ultimate parent of the applicant firm is a large firm (≥ 250 employees)
AGE_30	Dummy = 1 if inventor is less than 30 years old
AGE_30-40	Dummy = 1 if inventor is 30-40
AGE_40-50	Dummy = 1 if inventor is 40-50
AGE_50-60	Dummy = 1 if inventor is 50-60
AGE_60	Dummy = 1 if inventor is more than 60 years old
SECONDARY	Dummy = 1 if the inventor has a secondary school degree or lower
HIGH-SCHOOL	Dummy = 1 if the inventor has a high-school degree
BA/MASTER	Dummy = 1 if the inventor has a BA or Master
PHD	Dummy = 1 if the inventor has a PhD
OTHER_UNIV	Dummy = 1 if the inventor has a special university degree (e.g., from Technical Universities, particularly in NL or DE)
MALE	Dummy = 1 if the inventor is a male
MOTIVATION	Sum of the scores (b/w 1-5) to four PatVal-EU questions regarding the extent to which the inventor is motivated by: i) money; ii) career; iii) prestige/reputation; iv) satisfaction with solving the problem
YRINORG	# of years (in 2006) that the inventor has been employed with the applicant organization
IPC4_NOFIRMS	Share of individual or non-profit applicants in the IPC4 class of the patent
IPC4_COMP	1 – Herfindhal index of the share of different applicants in the IPC4 class of the patent
PROTECTION	Sum of the scores (b/w 1-5) to two PatVal-EU questions regarding the applicant's motivation for patenting the invention: i) obtain exclusive rights to commercialize the invention; ii) prevent others from imitation
RTA	Revealed technological advantage of the ultimate parent of the applicant in the technological field (30 technological classes) of the patent. (Share of firm patents in the field over field patents on total patents)
BIGCITY	Dummy = 1 if invention was produced in a city of 500k inhabitants or more
SERENDIPITY	Dummy = 1 if, as stated in the formulation of the question in the survey, "the idea for the invention came from pure inspiration or creativity or from your normal job (which is not inventing), and was not further developed in a (research or development) project (and it was patented without further research or development costs)"
GOVFUND	Dummy = 1 if the funding of the research leading to this patent came from Government research programs or related government funds
RD	R&D expenditures of the firm in 1995 (in 000 euros)
MISS_RD_SF	Dummy = 1 if RD is missing and SMALL_PARENT=1
MISS_RD_MF	Dummy = 1 if RD is missing and MEDIUM_PARENT=1
MISS_RD_LF	Dummy = 1 if RD is missing and LARGE_PARENT=1
INTRATECY	3-yr moving average of the interest rate of the country before the priority date of the patent

Table 5: Value regressions, descriptive statistics

variable	mean	sd	Min	med	max	N. obs.
V	580.416	602.828	68.1	422.357	17254.68	5359
M	11.619	16.829	0.5	4.5	84	5359

N	4.111	6.118	1	1	31	5359
Z	0.58	1.107	0	0	14.583	5359
MF	27.199	36.643	0.5	9	122	5359
SMALL_PARENT	0.078	0.268	0	0	1	5359
MEDIUM_PARENT	0.057	0.232	0	0	1	5359
LARGE_PARENT	0.865	0.342	0	1	1	5359
AGE < 30	0.048	0.213	0	0	1	5359
AGE 30-40	0.344	0.475	0	0	1	5359
AGE 40-50	0.314	0.464	0	0	1	5359
AGE 50-60	0.249	0.432	0	0	1	5359
AGE > 60	0.046	0.209	0	0	1	5359
SECONDARY	0.022	0.147	0	0	1	5359
HIGH-SCHOOL	0.144	0.351	0	0	1	5359
BA/MASTER	0.539	0.499	0	1	1	5359
PHD	0.287	0.452	0	0	1	5359
OTHER_UNIVERSITY	0.008	0.087	0	0	1	5359
MALE	0.977	0.151	0	1	1	5359
MOTIVATION	13.438	3.366	4	14	20	5359
YRINORG	25.325	10.284	1	22	83	5359
IPC4_NOFIRMS	0.096	0.08	0	0.075	0.6	5359
IPC4_COMP	0.934	0.072	0	0.955	0.995	5359
PROTECTION	7.877	2.121	0	8	10	5359
RTA	11.268	13.791	0.024	7.646	185.143	5359
BIGCITY	0.218	0.413	0	0	1	5359
SERENDIPITY	0.111	0.314	0	0	1	5359
GOVFUND	0.068	0.251	0	0	1	5359
RD	1650.498	1637.626	0.912	1274.290	8387.898	2429
MISSING_RD_SF	0.077	0.267	0	0	1	5359
MISSING_RD_MF	0.057	0.232	0	0	1	5359
MISSING_RD_LF	0.412	0.492	0	0	1	5359
INTRATECY	5.742	2.198	1	5.812	11.729	5359

Table 6: First stage regressions of *N*, *M*, *Z*, OLS

Variables	*N*	*M*	*Z*	Variables	*N*	*M*	*Z*
SMALL_PARENT	0.596	0.641	-0.559	IPC4_COMP	-0.657	-0.110	-0.383
	0.000	0.000	0.000		0.177	0.777	0.024
MEDIUM_PARENT	-0.405	-1.080	-0.353	PROTECTION	0.115	0.014	-0.015
	0.026	0.000	0.000		0.068	0.743	0.528
AGE_30-40	0.113	0.028	0.108	RTA	0.002	-0.011	0.005
	0.088	0.638	0.000		0.892	0.290	0.505
AGE_40-50	0.186	0.037	0.216	BIGCITY	0.089	-0.001	0.043
	0.011	0.553	0.000		0.049	0.964	0.035
AGE_50-60	0.276	0.008	0.232	SERENDIPITY	-0.243	-0.042	-0.074
	0.001	0.901	0.000		0.000	0.249	0.000
AGE_60	0.317	0.104	0.163	GOVFUND	0.166	0.234	-0.023
	0.003	0.230	0.000		0.005	0.000	0.489
HIGH-SCHOOL	-0.036	-0.050	0.039	RD	0.051	-0.005	0.002

	0.699	0.474	0.467		0.003	0.754	0.830
BA/MASTER	0.122	0.177	0.137	MISS_RD_SF	-0.375	-0.612	0.483
	0.174	0.006	0.009		0.002	0.000	0.000
PHD	0.328	0.126	0.212	MISS_RD_MF	0.564	1.060	0.254
	0.001	0.059	0.000		0.002	0.000	0.000
OTHER_UNIV	-0.064	0.219	0.172	MISS_RD_LF	0.271	-0.001	-0.031
	0.632	0.012	0.065		0.014	0.992	0.586
MALE	0.183	-0.135	0.088	INTRATECY	-0.337	-5.597	0.068
	0.016	0.055	0.068		0.000	0.000	0.078
MOTIVATION	0.198	0.063	0.030	CONST	0.183	11.928	-0.501
	0.001	0.056	0.202		0.687	0.000	0.010
YRINORG	-0.06	0.028	0.155				
	0.233	0.452	0.000	*Statistics*			
IPC4_NOFIRMS	-0.283	0.353	-0.240	N. obs.	5359	5359	5359
	0.334	0.090	0.073	R2	0.090	0.739	0.136

All variables (but dummies) in logs (including dependent variables); log(1 + variable) if the variable can take value 0; p-values below estimates; equations include country dummies, dummies for technological sectors and patent priority years; observations clustered by ultimate parent company; regression includes weights to adjust the oversampling of important patents in PatVal-EU (important patents = cited at least once or opposed).

Table 7: Patent Value Regression, OLS and GMM
(Dependent variable: v = log of the predicted value from the interval regression in Table 3)

Variable	OLS	GMM	Variable	OLS	GMM
N	0.026	0.016	OTHER_UNIV	0.004	-0.002
	0.000	0.880		0.969	0.986
M	0.053	0.040	MALE	-0.066	-0.059
	0.000	0.002		0.126	0.240
Z	0.023	-0.032	MOTIVATION	0.025	0.030
	0.147	0.,930		0.339	0.348
SMALL_PARENT	0.054	0.051	YRINORG	-0.071	-0.058
	0.048	0.142		0.003	0.390
MEDIUM_PARENT	-0.003	-0.007	IPC4_NOFIRMS	-0.158	-0.134
	0.906	0.853		0.216	0.377
AGE_30-40	0.026	0.033	IPC4_COMP	0.709	0.570
	0.359	0.436		0.003	0.042
AGE_40-50	0.041	0.050	PROTECTION	0.053	0.066
	0.190	0.479		0.014	0.020
AGE_50-60	0.048	0.059	RTA	0.017	0.010
	0.166	0.417		0.050	0.215
AGE_60	0.012	0.027	BIGCITY	0.003	0.001

	0.790	0.653		0.887	0.958
HIGH-SCHOOL	-0.025	-0.024	CONST	5.924	5.960
	0.572	0.610		0.000	0.000
BA/MASTER	-0.000	0.016			
	0.996	0.798	*Statistics*		
PHD	0.013	0.045	N. obs.	5359	5359
	0.782	0.546	LOG-LIK	-2.994.5	-3.012.6

All variables (but dummies) in logs (including dependent variables); log(1 + variable) if the variable can take value 0; p-values below estimates; equations include country dummies, dummies for technological sectors and patent priority years; observations clustered by ultimate parent company; regression includes weights to adjust the oversampling of important patents in PatVal-EU (important patents = cited at least once or opposed). Excluded instruments: SERENDIPTY, GOVFUND, RD, DMISS_RD_SMALL, DMISS_RD_MEDIUM, DMISS_RD_LARGE, INTRATECY

Table 8: Tests for weak and relevant instrument (GMM equation in table 7)

Summary results for first-stage regressions

Variable	Shea Partial R2	Partial R2	F(7, 1649)	P-value
N	0.0064	0.0164	18.91	0.0000
M	0.2737	0.6902	894.59	0.0000
Z	0.0029	0.0049	13.90	0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)

Kleibergen-Paap rk LM statistic Chi-sq(5)=13.41 P-val=0.0198
Kleibergen-Paap rk Wald statistic Chi-sq(5)=134.74 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Kleibergen-Paap Wald rk F statistic 19.01

Stock-Yogo weak ID test critical values:
5% maximal IV relative bias 13.95
10% maximal IV relative bias 8.50
20% maximal IV relative bias 5.56
30% maximal IV relative bias 4.44

Hansen J statistic (overidentification test of all instruments): 4.829
Chi-sq(4) P-val = 0.3053

Table 9: *N, M, MF* GMM regressions

Variable	*N*	*M*	*MF*	Variable	*N*	*M*	*MF*
M	-1.034	--	--	OTHER_UNIV	0.172	1.723	1.409
	0.000				0.265	0.035	0.019
MF	1.238	--	--	MALE	0.044	0.515	0.446
	0.000				0.541	0.192	0.103
Z	-1.103	-5.021	-3.836	MOTIVATION	0.070	0.529	0.734
	0.026	0.002	0.001		0.212	0.005	0.000
SMALL_PARENT	-0.181	-0.208	-0.108	YRINORG	0.125	0.385	0.372
	0.003	0.399	0.549		0.152	0.197	0.075
MEDIUM_PARENT	-0.204	-0.414	-0.311	IPC4_NOFIRMS	-0.668	-2.987	-2.431
	0.002	0.127	0.115		0.023	0.007	0.003
AGE_30-40	0.066	0.741	0.649	IPC4_COMP	0.061	-1.775	-2.987
	0.292	0.011	0.003		0.873	0.268	0.011
AGE_40-50	0.236	1.542	1.243	PROTECTION	-0.069	-0.120	0.215
	0.024	0.000	0.000		0.140	0.503	0.093
AGE_50-60	0.224	1.432	1.254	RTA	0.047	-0.073	-0.069
	0.035	0.002	0.000		0.001	0.160	0.069
AGE_60	0.190	1.213	1.161	BIGCITY	0.075	0.097	0.272
	0.050	0.003	0.000		0.059	0.570	0.018
HIGH-SCHOOL	0.091	0.082	-0.101	CONST	-0.917	-0.077	-0.243
	0.272	0.839	0.746		0.014	0.958	0.829
BA/MASTER	0.130	0.921	0.792				
	0.167	0.036	0.019	*Statistics*			
PHD	0.195	1.365	1.370	N	5359	5359	5359
	0.071	0.009	0.001	LOGLIK	-7281.7	-1.25E+04	-1.20E+04

All variables (but dummies) in logs (including dependent variables); log(1 + variable) if the variable can take value 0; p-values below estimates; equations include country dummies, dummies for technological sectors and patent priority years; observations clustered by ultimate parent company; regression includes weights to adjust the oversampling of important patents in PatVal-EU (important patents = cited at least once or opposed). Excluded instruments: SERENDIPTY, GOVFUND, RD, DMISS_RD_SMALL, DMISS_RD_MEDIUM, DMISS_RD_LARGE, INTRATECY

Table 10: Tests for weak and relevant instrument (GMM equations in table 9)

Summary for first-stage regression

Variable	Shea Partial R2	Partial R2	F(7, 1649)	P-value
M	0.0076	0.6902	894.59	0.0000
MF	0.0041	0.3528	284.20	0.0000
Z	0.0021	0.0049	13.90	0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)
Ha: matrix has rank=K1 (identified)

Kleibergen-Paap rk LM statistic	Chi-sq(5)= 9.84	P-val=0.0798
Kleibergen-Paap rk Wald statistic	Chi-sq(5)= 47.31	P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Kleibergen-Paap Wald rk F statistic	6.67	
Stock-Yogo weak ID test critical values:	5% maximal IV relative bias	13.95
	10% maximal IV relative bias	8.50
	20% maximal IV relative bias	5.56
	30% maximal IV relative bias	4.44

Hansen J statistic (overidentification test of all instruments): 1.362
Chi-sq(4) P-val = 0.8508