Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No.1813

November 2021

Efficient industrial policy for innovation: standing on the shoulders of hidden giants

Charlotte Guillard
Ralf Martin
Pierre Mohnen
Catherine Thomas
Dennis Verhoeven

Abstract
Research and development is underprovided whenever it creates knowledge spillovers that drive a wedge between its total and private economic returns. Heterogeneity in the intensity of this market failure across technological areas provides an argument to vertically target public support for R&D. This paper examines potential welfare gains of such vertical industrial policy for innovation. It develops measures of private and spillover value of patented innovations using global data on patents and their citations. Our new method identifies a large number 'Hidden Giants' – i.e. innovations scoring higher on our new spillover measure than on the traditional forward citation count measure – which are shown to be particularly prevalent among patents applied for by universities. The estimated distributions of private values by technology area are then used to parameterize a structural model of innovation. The model permits estimation of the marginal returns to technology-area-specific subsidies that reduce innovators' R&D costs. Marginal returns are high when knowledge spillovers in the technology area are valuable, when private innovation costs are low, and when private values in a technology sector are densely distributed around the private cost. The results show large variation in the marginal returns to subsidy and suggest that targeted industrial policy would have helped mitigate underprovision of R&D over the time period studied. Variation in the extent to which knowledge spillovers are internalized within countries also makes a compelling case for supranational policy coordination, especially among smaller countries.

Key words: research and development, patented innovations, decoupling, targeted industrial policy

This paper was produced as part of the Centre's Growth Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We acknowledge support from the EU Horizon 2020 Program via the WATSON Project. Verhoeven gratefully acknowledges funding from the Research Foundation Flanders (FWO grant 1241520N) and the H2020 Framework Programme (Marie Sk lodowska-Curie Action DIV INV grant 844254).

Charlotte Guillard, University College London. Ralf Martin, Imperial College London, CEPR and Centre for Economic Performance, London School of Economics. Pierre Mohnen, Maastricht University and UNU_MERIT. Catherine Thomas, London School of Economics and Centre for Economic Performance, LSE. Dennis Verhoeven, Bocconi University, KU Leuven and Centre for Economic Performance, London School of Economics.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

1 Introduction

Public investment in science and technology is the centerpiece of many developed economies' growth strategies. The rationale for these investments is that innovation generates positive externalities in the form of knowledge spillovers. An innovator compares private investment costs with private benefits when deciding the level of investment, but will underinvest if the knowledge embodied in her innovation enables subsequent knowledge creation with private benefits that are captured elsewhere. As such, there is market failure in innovation whenever private incentives reflect only part of its benefits. This logic underpins the national patenting system in most countries, which serves to improve the alignment of private innovation costs and benefits, and has also motivated large-scale public investment to support knowledge creation.

If it were possible to anticipate exactly which innovation activity would create the most valuable knowledge spillovers, policymakers could encourage such activity via targeted financial support to improve alignment in private incentives and thereby achieve welfare gains. This paper develops methods to measure incentive misalignment across areas of innovation activity and uncover the areas with the greatest marginal return to targeted innovation subsidies.

In the first part of the paper, we develop an approach to quantify an innovation's economic value—the sum of its private value (PV) and its external value (EV), where the latter is the value of the knowledge spillovers it creates. We focus on the innovations described in patent families, and infer knowledge spillovers from the information contained in the network of patent citations in the PATSTAT database. Because a patent applicant is required to cite the patents that constitute the prior art in order to establish the extent of any new innovation, patent citations can be viewed as list of the knowledge inputs to innovation production. This 'paper trail' of knowledge flows has become become an important methodological tool in establishing granular spillovers between patents (Trajtenberg (1990), Jaffe et al. (1993)).

The two novel elements to our approach are first to consider both direct and indirect knowledge inputs, as summarized in the citation network, and, second, to assign a value to these flows based on the private values of subsequent innovations. As such, the EV of a given innovation is the sum of shares of the private values of all of the innovations that cite it, either directly or indirectly. This recursive reasoning is formalized in a system of equations that can be solved using a simple iterative algorithm.[1] Our method is inspired by Google's PageRank search algorithm that uses hyperlink networks to rank the web pages returned in search results. In our application, which we call 'Patent Rank' (or P-Rank), a patented innovation plays the role of a web page, and the network of patent citations is analogous to the network of hypertext links in the Web.[2]

An important part of the analysis producing P-Rank is estimating the private value of each patent family. For the subset of patents that are held by firms listed on stock exchanges, estimates of the patents' private values from stock market returns around the time of granting are available (Kogan et al., 2017). Because the data describe patents in some detail, we are able to determine the patent characteristics that are correlated

[1] The solution to the system of equations corresponds to the principal eigenvector of the normalized link matrix of the patent citations.

[2] The reason for describing our application as a ranking of innovations rather than a cardinal ordering based on economic values is that the estimated ranking of patent families by total value is broadly robust across different assumptions about the relative weight placed on direct and indirect citations, whereas the actual values vary with these assumptions.

with estimated private values for the patents held by listed firms.[3] We use the sample of patents from listed firms to predict private values implied by stock market reactions based on characteristics observed for all patents. This predictive model is then applied to to all patents in the population, both for listed firms and other organizations, to construct our measure of private value. [4]

By incorporating the network of citations and attaching values to individual direct and indirect links in the network, our approach uncovers groups of innovations whose total economic values were thus far at least partly obscured. Innovations that appear more eminent in terms of our EV measure than they would based on the number of direct forward citations are dubbed 'Hidden Giants'.[5] Innovations whose direct forward citations outweigh their EV are, instead, 'Illusory Giants'.[6] There are 1.7 Hidden Giants to every Illusory Giant. This is because while most highly-cited innovations also have a high EV, and their 'giantness' is therefore neither Hidden or Illusory, there is a large variance in the EVs of scarcely-cited innovations, and the subset of these with high EVs were Hidden. In fact, Hidden Giants are responsible for 56% of all external value created and are, hence, an important source of knowledge spillovers.

Because the total economic value is measured at the level of the patent family, it can be aggregated to the technology area or to the level of the innovator's country. Furthermore, the knowledge spillovers embodied in EVs can be traced across technology areas and across countries. Some technologies and countries generate more external value outside their own borders than within, a finding that has implications for optimal policy design, as will be shown later in the paper. One insight from the P-Rank measure that serves to validate its relevance is the fact that the patents held by universities are particularly likely to be Hidden Giants and unlikely to be Illusory Giants. That is, the innovations emerging from university research tend to provide indirect knowledge inputs to a larger number of more valuable subsequent innovations. Universities are, hence, performing their knowledge production function more effectively when assessed using the P-Rank measure than when counting forward citations alone.

The second part of the paper estimates the marginal economic return to subsidies targeted to specific technology areas. In this analysis, a subsidy has no impact on the private and external value of innovations that would have been done in the absence of the subsidy, but does increase the innovative activity on the margin in a technology area by lowering private innovation costs. The challenge, then, is to establish how much new activity would result from the subsidy and then calculate the value of this activity. To do this, we put forward a simple structural model of the private costs and benefits of investments in innovation. Potential innovators draw ideas of varying quality from an idea distribution, and choose whether to develop the idea, where development incurs a cost and, with some probability, leads to an innovation. The shape of the idea distribution, the development cost, and the probability of innovation success are all specific to the technology area and time. These model parameters are estimated using the technology-area specific distributions of private values constructed in the first part of the paper. The marginal value of a subsidy is inferred from the marginal impact on the quantity of

[3]The patent characteristics found to be significantly associated with private patent value include fine-grained technological class, application date, patent family size, and number of forward citations.

[4]The patents held by firms listed on US stock markets make up only 3.4% of all patents. Private values from the predictive model correlate reasonably well to stock-market-based estimates (0.53 after taking the logarithm of both values, 0.38 before).

[5]Named in honor of Newton's 1675 quote "If I have seen further it is by standing on the shoulders of giants."

[6]Named in honor of the Herr Turtur in Michael Ende's book "Jim Knopf und Lukas der Lokomotivführer"

innovation in an area and the estimated distributions of PV and EV for the technology area at the marginal quantity. The estimated values allow a ranking of technology areas according to the relative efficiency of targeted subsidies. We call this ranking IStraX, the Industrial Strategy Index.

Some technology areas are higher ranked in IStrax than in the rankings by total EV or in P-Rank, which sums PV and EV. Having low innovation costs or having a high density of private values around the innovation cost threshold amplifies the estimated return to a subsidy coming from a technology area's EV alone. For example, while Organic Fine Chemistry and Pharmaceuticals have the highest EVs, they are ranked lower than 20 other technology areas in IStraX. Wireless technology is top of the IStrax ranking, mostly due to low innovation costs, and Clean Energy jumps 13 places to fourth place in IStraX due to high estimated private value density around the innovation cost threshold.

In the final part of the paper, we turn to the question of how these measures can be used to improve the efficiency of public investment in science and technology. While the estimates are based on innovations from the past, initial investigations show that the P-Rank and IStraX rankings by technology area and country are relatively stable over time. We frame our discussion in terms of revealing where targeted subsidies could have been efficiently employed over the time period studied, but the stability of the rankings suggest that they continue to be useful.

IStraX shows large and statistically significant differences in the marginal social returns to subsidies across technological fields and geographic regions. Technology areas including Wireless, Clean Energy, AI, and Robotics generate returns of above 40%, while Civil Engineering, Machine Tools and Mechanical Elements show return rates of about 15%. Within each country, we show that technology-area IStraX rankings differ considerably from rankings based on the relative intensity of patenting activity in that area.[7]

In another exercise, we estimate how much of the external value generated by targeted innovation subsidies is retained within a country and how much spills over to other countries. This analysis informs the relative efficiency of coordinating industrial policy at a national versus supranational level. We find that large countries, such as the US and China, 'internalize' more than half of the valuable knowledge spillovers that their domestic innovations create. In contrast, although the smaller countries Germany and France generate comparable spillovers on a per capita basis, these countries internalize at most 15% of the value. In addition, IStraX rankings based on global spillovers correlate weakly with those based on national spillovers, particularly for smaller countries. These results suggest that supranational industrial policy would have led to substantial welfare gains, especially for smaller countries.[8]

The paper proceeds as follows: Section 2 presents the data used. Section 3 describes how P-Rank is computed, presents descriptive statistics and compares external values from P-Rank to forward citation counts. Section 4 sets out how IStrax is calculated and reports results from IStraX calibrated on technology areas. Section 5 examines heterogeneity in optimal industrial policy across countries and explores the benefits of supra-national coordination. Section 6 discusses and concludes.

[7] A country's share of patent families in a particular technology area divided by the country's share of patent families in all fields is a measure of 'revealed technological advantage', analogous to the revealed comparative advantage measure of relative intensity of sector-level exports.

[8] The IStraX rankings within country are more positively correlated with the global ranking for large countries than for small countries, and supranational coordination brings country-level rankings closer to global rankings.

2 Data

We rely on the EPO PATSTAT database (version 'Spring 2018') as main source of information. PATSTAT is the most comprehensive collection of patent information and covers documents from patent authorities worldwide[9]. While this worldwide coverage helps to broaden the scope of our analyses, it raises the problem of how to aggregate information from different patent offices. To see this issue, it is important to keep in mind the distinction between an innovation, a patent application, a patent document and a patent family. If we define an innovation[10] to be some improvement upon the state-of-the-art, a patent application is the process of trying to obtain legal protection for that innovation[11]. Each jurisdiction has its own patent authority[12] responsible for the decision to grant legal protection. As such, protecting an innovation in multiple countries requires multiple patent applications. During the patent application process, patent authorities produce legal documents available to the public[13]. It are these 'patent documents' that provide the information present in the PATSTAT (or any other patent) database. We aggregate information from these publications to the level of the innovation using the so-called DOCDB family definition[14] in PATSTAT[15]. To proxy innovations rather than patents, we use this grouping as the main unit of analysis.

We calculate our P-Rank algorithm based on the population of patent families between 2005 and 2014. We timestamp each family using the date at which its first patent application was filed. To construct the innovation network, we use patent families as nodes and citations between patent families as edges. As citations occur at the patent document rather than the patent family level, we drop duplicate links between patent families. In addition, we exclude citations between patent families from the same applicant (often referred to as 'self-citations') because these citations do not reflect knowledge spillovers between different innovators. As applicant names in PATSTAT are not harmonized, spelling variations across patents might lead us miss self-citation links. While we cannot completely avoid this, we try to mitigate this problem by linking applicants

[9]For more information, we refer to https://www.epo.org/searching-for-patents/business/patstat.html in general, and the EPO PATSTAT Data Catalogue in particular.

[10]In the literature there is often a distinction between invention and innovation, where invention is the creation of a novel idea and innovation refers to the commercialisation of an invention. In our data we have no direct data that would allow us to make such a distinction hence we simply use the term innovation throughout for the act of turning and idea into a patent which will lie somewhere between innvention and innovation.

[11]This legal right consists of the monopoly right to commercially exploit the innovation

[12]E.g. the United States Patent and Trademark Office (USPTO) or the European Patent Office (EPO)

[13]This brings about the fact that any innovation can relate to multiple (time-stamped) documents. For instance, an innovation might be linked chronologically to (1) a document filed at the EPO describing the invention, (2) a document describing the outcome of the prior art search from the EPO, (3) a document stating that a patent was filed for at the USPTO, (4) a similar document for the Japan Patent Office, (5) the decision of the patent grant at the EPO, and, (6) the decision of grant at the USPTO. Each of these documents might contain (duplicate) information about the invention, which is not necessarily consistent over time.

[14]This definition uses priority filings to group patent applications into a patent family corresponding to one innovation.

[15]As we are interested in having information about the innovation itself, it is key that we are able to group the information contained in all documents related to an innovation correctly. It is important to note that some decisions need to be made in order to group information from patent documents to the level of the innovation. The clearest example of such a decision is to determine the timing of the innovation. While each of the aforementioned documents has a timestamp, we choose to use the first filing date in the family of documents as the (proxy for) the date the innovation happened. This decision can be less obvious, for instance when multiple applicant or inventor countries are mentioned across different documents (for instance, because the patent applicant is a multinational). Unless mentioned otherwise, we aim to always retain the information available at the earliest time available, because this should reflect the state of information closest to the actual event of an innovation.

to entities present in the Orbis database. The Orbis database uses local registry filings across the world to construct a company database. As this database is linked to patent applicant names in PATSTAT, we can use the Orbis information to infer whether two applicants are indeed the same legal entity. To extrapolate private returns estimated in Kogan et al. (2017), we use PATSTAT information on technological classes, application filing date, patent family size and the number of claims. As for technological classes, we use the IPC codes from all patent applications in a family at the 'main group' level [16].

For the analyses we assign innovations to countries and broad technological fields. For the country information, we combine information on inventors' country code provided in PATSTAT with the country assigned based on geo-located addresses in de Rassenfosse et al. (2019). These two sources of information are complementary because the latter uses address information from local patent offices not present in PATSTAT while the former assigns country codes even to inventors for which address information is insufficient for reliable geo-location. To assign innovations to technological fields, we use PATSTAT's assignment of patent applications to 35 technological fields based on Schmoch (2008). To these fields, we add 6 fields of 'special interest' such Clean Technology and Artificial Intelligence. Each of the resulting 41 fields is based on the technological classes examiners add to patent applications. Appendix B details the definition of our technological fields.[17] For both - countries and fields - one patent family may be a member of multiple groups. In the analyses, we assign such patents fully to each of these groups.

3 Patent Rank

3.1 The Measure

Generally speaking, Patent Rank (or P-Rank) corresponds to the total economic returns produced by an innovation. We assume these returns are made up of (1) private value and (2) external value. The former denotes the returns that are appropriated by the innovator commercially. The latter is the value an innovation creates by reducing the cost of innovating on the part of future innovations because of the knowledge embedded in it. Theoretically, this external value is equal to the reduction in the total private values generated in the future, should the innovation have not been disclosed. To implement this idea, we assign a portion of the total economic returns of any one innovation to all innovations it cites as prior art. This portion is part of the external value of the cited innovation. Therefore, the total external value of any innovation is obtained by summing up this portion over all citing innovations. As a result, the P-Rank of an innovation depends on its own private value, the total value of all innovations citing it, and the portion of value we assign as corresponding to knowledge spillovers. In what follows, we formally describe P-Rank and its parameters.

[16] For more information, we refer to https://www.wipo.int/classifications/ipc/en/preface.html)

[17] Our technology field definitions correspond to IPC and CPC classifications added to patent applications by its examiner. The 'CPC' refers to the Cooperative Patent Classification, which was initiated in 2010 in a joint partnership between the EPO and USPTO in order to harmonize their existing classification scheme. It builds upon and exists next to the IPC classification system and is especially useful for our purposes because it introduces specific classes for 'clean' technologies.

An innovation makes up part of the stock of knowledge and is described in a patent family[18]. Each patent i cites N_i other patents given by set B_i. Patent i may also be one of the patents that make up B_j, i.e. be a citation for patent j, where j is any other patent in the data. The set of all patents that cite i is F_i.[19]

P-Rank is the value of patent i, V_i, and is made up of two parts. First is the private value of patent i, PV_i. Second is the value that patent i has created when used as an input to the production of patent j for all $j \in F_i$, denoted $EV_i = \sum_{j \in F_i} f_{ij}(V_j)$. Therefore:

$$V_i = PV_i + EV_i = PV_i + \sum_{j \in F_i} f_{ij}(V_j) \tag{1}$$

An important aspect of Patent Rank is the nature of the function $f_{ij}(V_j)$. In this version of the paper, we assume that this function contains two parameters that are properties of the innovation production function. Suppose that the value V_j derives from a production function where the inputs include the R&D investment from the firm (RD_j) and also the stock of knowledge embodied in prior patents B_j. For example, if we assume that there exists an innovation value production function and it has a Cobb-Douglas form with efficiency shifter A_j, then:

$$V_j = A_j B_j^{\sigma} RD_j^{1-\sigma} \tag{2}$$

In this formulation, the parameter σ measures the relative contribution of prior knowledge to future innovation, and does not vary with j. By differentiating equation(2) with respect to each cited innovation $i \in B_j$, and then substituting in for V_j, we derive an expression for the marginal contribution of citation i to V_j.

$$\frac{\partial V_j}{\partial i} = V_j \sigma \frac{\partial B_j}{\partial i} \frac{1}{B_j} \tag{3}$$

We denote as ϕ_{ij} the term $\frac{\partial B_j}{\partial i} \frac{1}{B_j}$. This term measures patent i's contribution to the stock of knowledge used in the production of patent j. For now, we assume that each of the N_j patents in the set B_j contributes equally, therefore this term simplifies to $\phi_{ij} = \frac{1}{B_j}$ for each $i \in B_j$ and zero for all $i \neq B_j$.[20] We are now able to write the P-Rank of patent i as:

$$V_i = PV_i + \sigma \sum_{j \in F_i} \phi_{ij} V_j = PV_i + \sigma \sum_{j \in F_i} \frac{1}{N_j} V_j \tag{4}$$

The parameter σ determines what share of patent i's value is attributable to the value of the spillovers that it creates. Because of the recursive nature of the P-Rank measure, it weighs the value of indirect forward citations by σ to the power of the level of indirectness. As such, it can be viewed as a distance decay parameter (e.g. the value of a patent that contains a backward citation of a backward citation of patent i will be weighted by σ^2 in V_i). Below we find that while the overall level of V_i is highly sensitive to σ, it has little impact on the ranking of innovations in terms of value.[21] This is re-assuring as our primary objective is to rank innovation relative to each other rather than come up with an estimate of their value overall.

[18] We use the term patent, patent family, family and innovation interchangeably in the remainder of the paper. Each of these terms refers to an innovation conceptually, and a patent family empirically.

[19] B refers to "backwards citations" and F refers to "forwards citations".

[20] A more general specification would be to assume that the contribution of innovation i to the knowledge used in j is a function of the characteristics of i and j, for example, whether both patents are in a similar technology class. We could denote this $\phi_{ij} = \phi(x_i, x_j)$, where the arguments of the function are patent characteristics.

[21] We present results corresponding to different values of σ, 0.25, 0.50, and 0.75.

3.2 Computing P-Rank

This section describes the computation of the base version of P-Rank. As expression 4 corresponds to a large system of equations, the solution implies inverting an $[N \times N]$ matrix. To avoid such a computationally expensive operation, we make use of an iterative algorithm to solve for P-Rank. In addition, we show how we can use the methodology to calculate both direct and indirect spillovers from and to any area of innovation that can be defined using patent data.

We collect the private values in the vector $\boldsymbol{PV}$, where the number of elements is equal to the total number of patents in the data, N. We also construct the $[N \times N]$ matrix $\boldsymbol{\Phi}$, where the element (i, j) is equal to $\frac{1}{B_j}$ if patent j cites patent i where B_j denotes the number of backward citations of innovation j.

We can write the vector of P-Rank values as V, which is equal to:

$$\boldsymbol{V} = \boldsymbol{PV} + \sigma\boldsymbol{\Phi}\boldsymbol{V}, \tag{5}$$

which can be rearranged as:

$$\boldsymbol{V}^* = (\boldsymbol{I} - \sigma\boldsymbol{\Phi})^{-1}\boldsymbol{PV}. \tag{6}$$

This equation can be estimated using the following recursive procedure: Starting with an arbitrary set of initial values $V_i^{(0)}$[22]. We compute a set of new values $V_i^{(n)}$ as:

$$V_i^{(n)} = PV_i + \sigma \sum_{j \in F_i} \phi_j V_j^{(n-1)} \tag{7}$$

In the appendix we prove that equation 7 has V^* as a fixed point given our assumptions about $\boldsymbol{\Phi}$.[23]

Armed with the P-Rank estimates, $\boldsymbol{V}$, we can find the external value of every patent as the vector:

$$\boldsymbol{EV} = \boldsymbol{V} - \boldsymbol{PV} = \left[(\boldsymbol{I} - \sigma\boldsymbol{\Phi})^{-1} - \boldsymbol{I}\right]\boldsymbol{PV}. \tag{8}$$

The external value of a patent derived here represents the knowledge spillovers that it generates for the benefit of the rest of the innovation network. However, one might be interested in knowledge spillovers originating in or being received by a subset of the network. Simply cutting the network to that subset would, however, ignore many possible higher-degree paths between nodes in this subset. In our application, for instance, we are interested in spillovers originating in a country that are received by innovations within that country. Yet, two innovations in that country might be linked through an innovation in a third country, so that omitting that innovation would discard this indirect spillover. Instead, we propose following slight adaptation to the definition of P-Rank in order to capture spillovers given or received by a subset of innovations in the network.

Suppose we segment all patents into areas, where an area, A, in our application, is a geographical area. The sum of the external values by patents belonging to a given area A constitute the spillovers *generated* by this area.

$$ST_A^{out} = \sum_{i \in A} EV_i \tag{9}$$

[22] A natural choice is $V_i^{(0)} = PV_i$

[23] In our results we compare V^* to a simpler measure of spillovers based on direct linkages only. This is the first iteration of 7; i.e. the estimate of the direct value of innovation i, DV_i, is $DV_i = V_i^{(1)} = V_i + \sum_j \phi_{ij} V_j^{(0)} = V_i + \sigma \sum_j \phi_{ij} V_j$

Furthermore, part of the external value generated by a patent is transmitted to a specific area. To measure the total spillovers an area *receives* from other patents, we need to slightly alter the calculation of value V such that:

$$\tilde{V}_{i,A}^{(n)} = \tilde{PV}_{i,A} + \tilde{EV}_{i,A} = \tilde{PV}_{i,A} + \sigma \sum_{j \in F_i} \frac{V_{i,A}^{(n)}}{N_j} \tag{10}$$

With $\tilde{V}_{i,A}^{(0)} = \tilde{PV}_{i,A}$ and $\tilde{PV}_{i,A} = \begin{cases} PV_i, & \text{if } i \in A, \\ 0, & \text{otherwise.} \end{cases}$

The total spillovers area A receives from other patents can then be easily calculated as follows:

$$ST_A^{in} = \sum_i \tilde{EV}_{i,A}^{(n)} \tag{11}$$

3.3 P-Rank's Parameters

3.3.1 Estimating PV_i

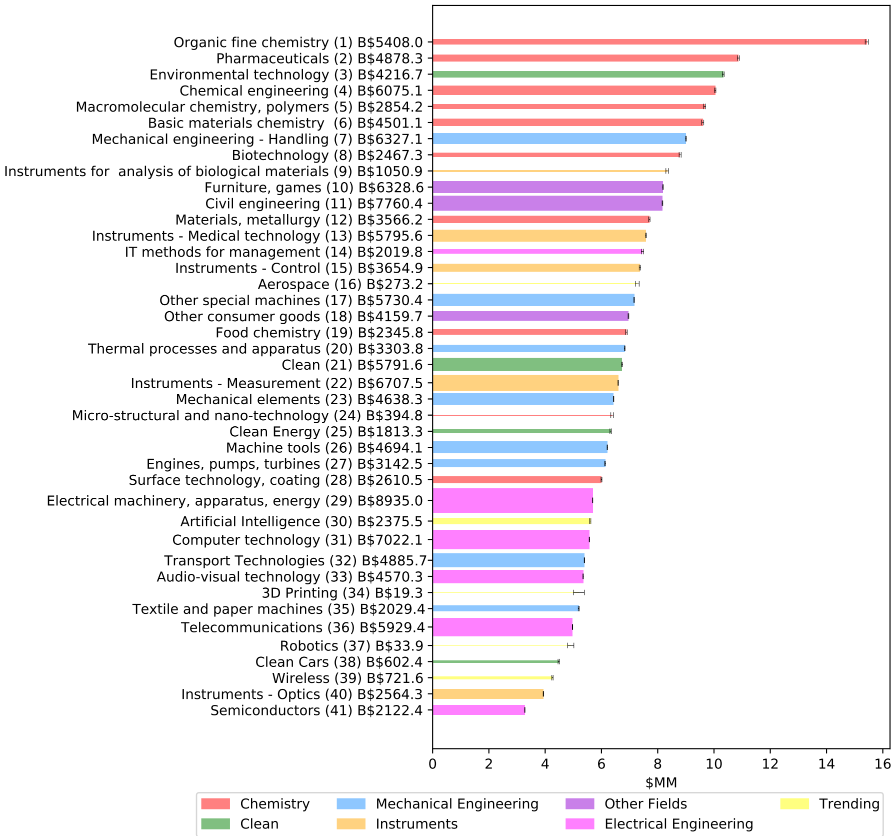
We refer to appendix A for a detailed description of the methodology to estimate private returns to innovations. Generally, our approach is to leverage private return estimates developed in Kogan et al. (2017) – henceforth KPSS – based on the stock market reaction to a patent grant to obtain an approximation of PV_i. KPSS measure the private value of a patent in an event study framework based on the change in the value of the patenting firm's abnormal stock market returns in a 3-day window around the day it is first granted the patent. The main drawback of this approach is that it provides information only for a relatively small set of US publicly-listed firms' patents. These patents amount to only 3.4% of the relevant population for our purposes.

We work around this drawback by extrapolating their estimates to (nearly) the entire population. To do this, we use a set of patent characteristics that are plausible predictors of private value (technology classes, time of application, patent family size and the number of claims). In a first iteration, we define highly detailed discrete categories for all our predictors, and create attribute groups for each combination of these categories. These attribute groups combine all innovations that share (discrete) values for each of the predictors. If an attribute contains at least 30 patent families with a KPSS value estimated, we assign the average of these values to each patent family in that group. Only a fraction of the innovations will be assigned a value in this first iteration because many attribute groups do not contain at least 30 patent families with KPSS values. Therefore, we gradually loosen the bounds on the discrete categories for our predictors in order to assign values to innovations in less populated attribute groups. We iterate this procedure until each patent family for which information is available on at least one of the predictors has received a private value. For patent families belonging to multiple attribute groups in the same iteration, we take the average private value of these groups.

In appendix A we compare the extrapolated private values (PV) to the values obtained by KPSS (ξ) for a test sample that was excluded in the extrapolation process. The correlation between the two measures of private value is 0.38 for the actual values, and 0.53 when taking the logarithm and standardizing the values. This correlation is quite stable (varying between 0.44 and 0.57) across the different extrapolation iterations, with the exception of one iteration representing 0.66% of the population where it is 0.13). Except for in the highest percentiles, the distributions of both measures are similar, but the extrapolated values are more centered around the mean.

Figure 3 examines average private values across technology fields for all innovations for which a patent was filed between 2005 and 2014. The x-axis represents the average private returns in a technology field (in million CPI adjusted 1982 dollars). The width of each bar represents the number of innovations in the technological field. Consequently, the area of each bar represents the total estimated private return in the technological field. This estimate (in billion) is also printed next to each field on the y-axis. Colors of the bars correspond to labels for broader technological domains, such as 'Chemistry' or 'Electrical Engineering'.

Figure 1: Private Returns by Technology - All 2005-2014 Innovations



Notes: Diagram of the average private returns in millions of CPI-adjusted 1982 US dollars (x-axis) by technology field (y-axis). Width of each bar represents the number innovations in the field. Area of each bar (in billions $) represents total private returns in the technology field and is printed next to y-axis labels.

From these results, two interesting patterns emerge. First, there is substantial variation between technologies in terms of private returns to innovation. For technology fields, average private returns range from about 3.5 million to about 16 million dollars per innovation – a range that covers about 50% of the entire distribution of private returns. For countries, the variation is similar (reported in appendix B), where average private returns vary between 2.5 and 10 million dollars. These patterns are reassuring under the plausible assumption that there exist considerable differences between the private returns to innovations across fields and countries.

Second, technological fields with high private returns seem to be those fields with high R&D costs per innovation. This pairs with the economic intuition that organizations only pursue R&D projects when the expected returns from the resulting innovation are larger than the costs. For instance, innovation in the pharmaceutical industry is notoriously expensive (see, for instance, DiMasi et al. (2003)), which suggests that the private returns for the average innovation should be relatively high. Fields where the average project is less costly, on the contrary, need relatively low private returns to convince organizations to pursue an innovation idea. The fact that fields such as 'Computer Technology' and 'Artificial Intelligence' have lower average private returns seems to be consistent with this notion.[24]

3.3.2 Estimating σ

In section 3.1 we derived an interpretation of σ as the marginal effect of spillovers. In other words, a large value for σ assumes that knowledge produced by innovations heavily reduces the cost of R&D by other innovators – i.e. the innovation process is highly cumulative. A low value reflects the assumption that prior knowledge is barely useful to follow-on innovation in that it barely decreases the cost of innovating – i.e. the innovation process is not very cumulative.

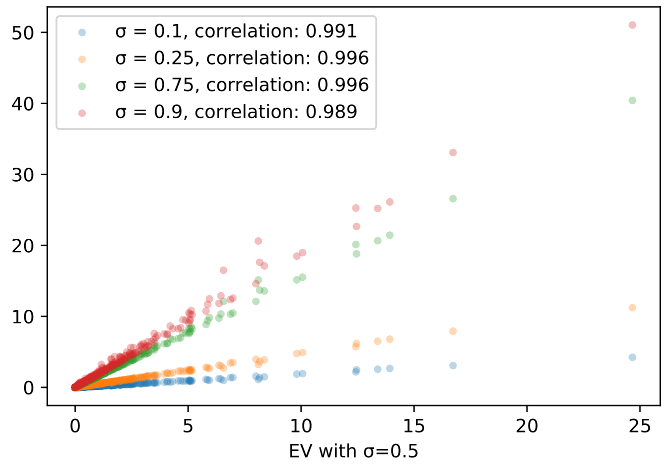
In the remainder of the paper, we assume a value of 0.5 – indicating that, on average, half of an innovation's total value derives from the stock of knowledge it acknowledges as prior art[25]. As this choice is rather arbitrary, we explore the sensitivity of P-Rank to this decision. Because this parameter determines the weighting of the external value to private value in P-Rank, and also determines the importance of indirect spillover values, it has a strong impact on the magnitude of the calculated external values and on P-rank value. For instance, with $\sigma = 25\%$, external values are about half as large as when we set $\sigma = 50\%$. With $\sigma = 75\%$, external values are 50% larger. We note that σ has a disproportionately large impact on indirect spillover value and, hence, on the ratio of indirect to direct spillover value. However, because σ enters the external value part of P-Rank multiplicatively but the external value enters P-rank additively, the value of σ has very little impact on the P-Rank patent ordering. Figure 2 illustrates that there is a near-perfect correlation between Patent Ranks computed under different values for σ. Consequently, the choice of σ has little impact on which innovation is classified as a Hidden Giant or as an Illusory Giant (see section 3.5) or on the share of the value that derives from each of the two groups.

The fact that the choice of σ affects the level of P-Rank but not the ordering has implications for policy design. Specifically, it creates a caveat to statements that quantify the benefits of these policies. Nonetheless, we can be confident that recommendations based on the relative merits of different technologies, sectors, countries or other groupings of innovations are robust to the choice of σ.

[24] A more speculative interpretation of the technology field ranking could be that fields in which more innovations per product are needed to effectively increase monopoly power score lower on private returns. Indeed, when a product consists of many, highly coupled components, an innovation pertaining to one of these components only might not suffice to increase margins much. As such, multiple innovations might be needed to increase profits, resulting in lower private returns per innovation in these fields. Examples of such fields could be 'Clean Cars', '3D printing' and 'Audio-visual Technology'.

[25] In future work, we will explore the possibility to allow σ to vary across technological fields to reflect differences in cumulativeness.

Figure 2: Scatter plot various sigmas



Notes: Scatter plot comparing external values of innovations with varying values for σ. To ease computation of this graph, a random sample of 1000 innovations is used. Correlation figures are based on the full population.

3.4 P-Rank Descriptive Results

Table 1 summarizes the distribution of the different types of innovation-level values we estimate. The first column ('PV') shows the extrapolated private values for all innovations between 2005 and 2014. The second column ('EV global') shows estimated external values resulting from our baseline specification of P-Rank. It counts spillovers generated from and to any geographic area without restrictions. The third and fourth column show distributions for 'direct' and 'indirect' external values. 'EV direct' takes into account only the spillovers generated to innovations one degree away in the network (that is, innovations directly citing the focal innovation), while 'EV indirect' only takes into account spillovers generated to innovations at least two degrees away in the network.[26]

Consistent with the notion that most innovations are of little value while few innovations are highly valuable, we see that each of these distributions is left-skewed with a high mass on zero. For private values, zeroes are those patent families for which none of its members was granted a patent.[27] For external value, zeroes result from families that did not receive any citations within the time frame considered. It is important to note here that innovations late in the time window considered had only limited time to receive citations. Our external value estimates should be interpreted as the spillovers generated *as of* the end of 2014. Furthermore, we find that indirect spillovers only constitute about one tenth of the total spillovers generated. This can be explained by the rather short time window employed, and by the fact that spillovers from indirect network linkages are exponentially discounted.

[26] 'EV global' is the sum of 'EV direct' and 'EV indirect'.

[27] Before February 2018, the end of the PATSTAT version we use.

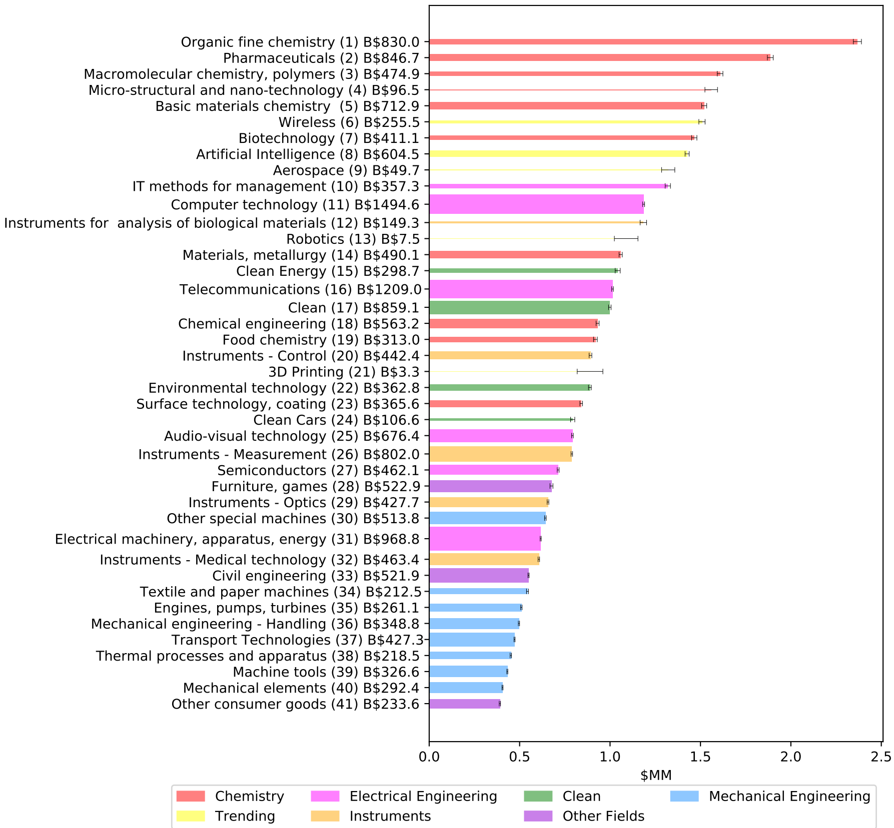
Table 1: Distributions Private (PV) and External Value (EV)

	PV	EV	EV direct	EV indirect
mean	6.87	0.68	0.61	0.071
min	0	0	0	0
p1	0	0	0	0
p5	0	0	0	0
p10	0	0	0	0
p25	0	0	0	0
p50	6.80	0	0	0
p75	10.4	0.41	0.37	0
p90	14.2	1.88	1.73	0.10
p95	18.5	3.49	3.17	0.31
p99	33.0	9.60	8.48	1.41
max	184.9	960.8	947.1	139.9
count	15068373	15068373	15068373	15068373

Notes: All values are in million CPI adjusted 1982 US dollars. The first column shows the distribution of private values. The second column shows the distribution of EV (spillovers to all geographic areas) for all innovations in the 2005-2014 period. The third column shows the distribution of spillovers created to innovations that directly cite the focal innovation. The fourth column shows the distribution of spillovers created to innovations that are at least 2 degrees away in the citation network.

Figure 3 shows the average external value by technological field. Fields at the top of the list are notably chemistry- and IT-related, and produce about 3 times more spillovers on average than fields at the bottom, which are often related to mechanical engineering. A comparison to private values (see Figure 3) reveals that fields with high private values generally generate high spillover value. This is unsurprising because spillovers derive from private returns obtained by others and are often localized within technological fields. However, fields such as Wireless, AI and Aerospace rank notably higher for external than for private value. This could be explained by the 'general purpose nature' of these fields, where they generate relatively high amounts of spillovers to distant fields.

Figure 3: Global External Value by Technology - All 2005-2014 Innovations



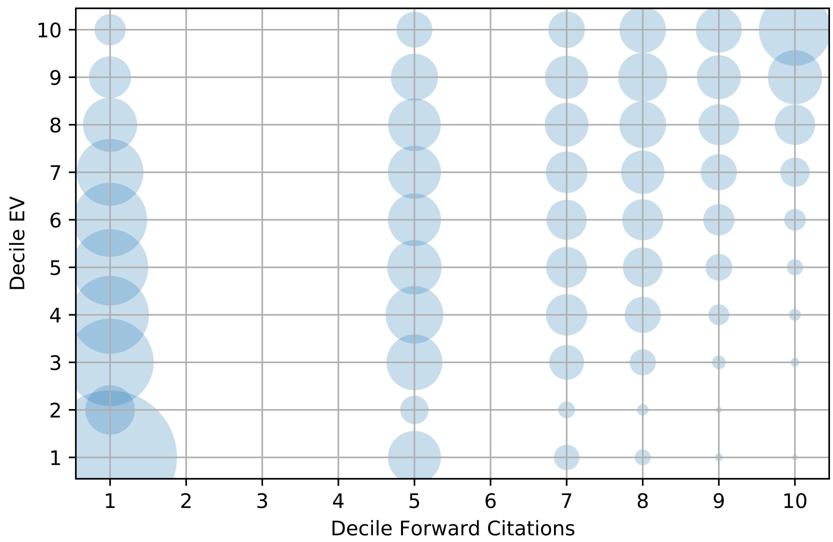
Notes: Diagrams of the global external value of innovations in millions of CPI-adjusted 1982 US dollars (x-axis) by technology field (y-axis). Width of each bar represents the number innovations in the field. Area of each bar (in billions) represents total external value in the technology field and is printed next to y-axis labels.

3.5 On Hidden and Illusory Giants

In this section, we explore the difference between P-Rank and forward citation counts – the most wide-spread measure of the economic returns to inventions. If the correlation between these measures is high, the added value of P-Rank – even if deemed superior to other methods – is low. If the reverse is true, P-Rank might add relevant information.

Figure 4 examines the joint distribution of EV as calculated using P-Rank, and the number of forward citations received. It plots a cross-tabulation between the decile bins of both indicators, where the size of the circle represents the number of families. 39.9% of all observations receive no forward citations in our time window (and hence produce zero EV) and are omitted from this graph. The figure shows that, while there is a correlation of 0.47 between the measures, off-diagonals are very prevalent. Especially for inventions that received few citations, the variation in terms of EV spans nearly the entire spectrum and is very large. The two measures seem to agree considerably more for high values of citation counts. High-scoring inventions with respect to EV, however, are to be found nearly uniformly across the citation distribution. If one were to see EV as the 'true' spillover measure, using forward citation counts to proxy spillovers would result in relatively few false positives, but would miss many high-spillover inventions (many false negatives).

Figure 4: Joint distribution External Value (EV) – Forward citation count



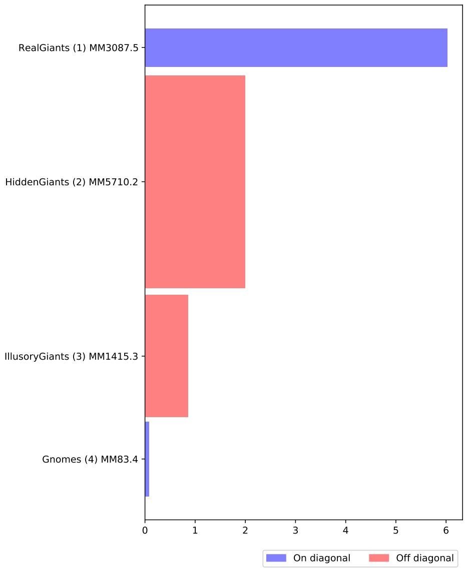
Notes: Visual cross-tabulation of decile bins for EV and forward citation counts. The y-axis and x-axis represent decile bins (higher decile denotes higher score) for global External Value (EV) and forward citation counts within the time window of our analyses. Size of the circles represent the number of innovations. Before assigning decile bins, we exclude all patent families (39.9% of our population) that were not cited (resulting in a value of zero for both measures).

The presence of many 'off-diagonals' suggests that our new measure could be bringing relevant new information to the table. To further investigate how EV compares to forward citation counts, we introduce 4 categories of innovation. 'Hidden Giants' are inventions that seem less unimportant (Hidden) when using traditional forward-citation-based measures, but are important (Giant) spillover generators when using P-Rank. We implement this by classifying innovations above the diagonal of decile groups in Figure 4 as Hidden Giants. Conversely, inventions below the diagonal – those that rank higher in terms of forward citations counts than EV – are coined 'Illusory Giants'. They appear to be giants in knowledge creation, but have relatively little spillover value according to the P-Rank approach.[28] 'Gnomes' and 'Real Giants' are inventions for which both measures agree – i.e. they are on the diagonal of decile groups – and are defined as belonging to decile 1 to 5 and decile 6 to 10 respectively for both measures of spillovers.

Figure 5 compares these 4 groups in terms of prevalence (width of bars) and average EV (height of bars). This figure confirms the presence of many off-diagonals. Only 25.2% of all innovations are either Real Gnome (16.7%) or Real Giant (8.5%). The average values in this figure follow expectations. As they are in the lowest, respectively highest, decile groups by definition, Gnomes and Real Giants display very low, respectively very high average external value. Hidden Giants and Illusory Giants do not necessarily belong to higher or lower decile groups, and therefore their average EV is modest. More interestingly, we see that most of the overall EV created, springs from Hidden Giants (55.5%) while Illusory Giants account for only 13.7% of all spillover value. This suggests that the amount of spillovers generated by Hidden Giants is not modest. They account for almost 4 times as much EV as compared to Illusory Giants. This suggests that being able to identify these Hidden Giants when supporting innovation has the potential to increase social value created by knowledge spillovers by a large margin.

[28] To learn more about Illusory giants see Michael Ende, "Jim Knopf und Lukas der Lokomotivführer".

Figure 5: Global EV - Hidden Giants groups



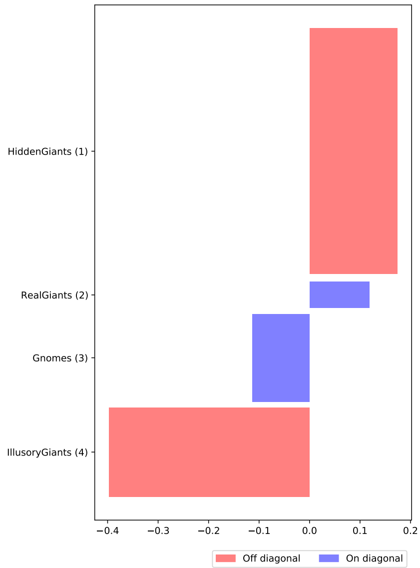
Notes: Diagram of average global External Value (EV) in millions of CPI-adjusted 1982 US dollars (x-axis) for 'Hidden Giants', 'Illusory Giants', 'Gnomes' and 'Real Giants'. Hidden Giants (resp. Illusory Giants) are defined as inventions above (resp. below) the diagonal of decile bins in Figure 4. Gnomes (resp. Real Giants) are inventions on the diagonal belonging to decile bins 1-5 (resp. 6-10) for both measures. The area of the bars represent the total global EV generated by the category (in billions), and is plotted next to the y-axis labels.

One potential caveat to this conclusion might be that P-Rank does not necessarily measure spillovers equally well for Hidden Giants as for Gnomes. To rule out this caveat, we design a validity check as follows. We assume that universities create inventions that have higher spillovers on average than private companies. This assumption seems plausible because university research exists precisely to mitigate the market failure for basic knowledge creation due to knowledge spillovers. Hence, to evaluate whether the EV generated by Hidden Giants is at least as well measured as that generated by Real Giants, we compare the relative advantage of universities for each of the 4 groups. We first assign patent families to companies and universities based on the 'sector allocation' given by PATSTAT. We only retain the patent families that are unambiguously assigned only to companies or universities.[29] Figure 6 shows the result of this exercise. The width of the bars represents the size of each of our 4 groups. The x-axis shows the relative share of universities in each group. This is calculated as the share of university patents in the group, divided by the share of this group in all patents. We subtract one from this so that positive (resp. negative) values indicate that universities are overrepresented (resp. underrepresented) in a particular group. This graph shows that universities are overrepresented in both Real Giants and Hidden Giants, and underrepresented in Gnomes and (particularly) Illusory Giants. We interpret this as evidence that the validity of EV as a measure of spillovers is no different for Hidden Giants than for on-diagonal inventions. Furthermore, the stark underrepresentation of universities in Illusory Giants could suggest that spillovers from

[29] We drop cases for which the sector allocation in PATSTAT contains ambiguous categories such as 'Company-Government', as well as cases where a patent family belongs to multiple different categories, for instance because it is the result of a company-university collaboration.

inventions that *only* score high on citation counts are overstated based on this measure. Together with the importance of Hidden Giants in overall spillover creation, this suggests that our new measure is likely to add relevant information to policy makers by being able to identifying spillovers missed when looking at forward citation counts only.

Figure 6: Relative share university - Hidden Giants groups



Notes: Diagram of relative share of university patents (x-axis) for 'Hidden Giants', 'Illusory Giants', 'Gnomes' and 'Real Giants'. Hidden Giants (resp. Illusory Giants) are defined as inventions above (resp. below) the diagonal of decile bins in Figure 4. Gnomes (resp. Real Giants) are inventions on the diagonal belonging to decile bins 1-5 (resp. 6-10) for both measures. The relative share is calculated as (share university patents in group)/(share group in total) - 1.

4 Industrial Strategy Index

This section describes how P-Rank can inform policy design. We develop IStraX (Industrial Strategy Index), a measure for the expected economic returns to subsidies – i.e. the sum of the private value and the externality generated by spillovers – that can be used to efficiently allocate innovation support to different areas of innovative activity. In our application, these areas correspond to combinations of technological field and country, but the framework can be applied to any grouping of (patented) innovation that is relevant to policy makers (for instance, regions, cities or industrial sectors).

Suppose a policy maker in a certain country is interested in allocating subsidies to a number of technological fields so as to maximize expected economic returns. We assume that a certain subsidy S to a technological field is used to decrease the field-specific fixed cost of developing an innovative idea into an innovation. The question, then, is how should any subsidy amount S be distributed? To answer this question, we need to quantify the marginal impact of a small amount of subsidy in every area. This marginal impact will depend on the private and external value that an average innovation in an area generates given an additional amount of subsidy. In the previous section, we quantified the private and spillover returns of innovations in an area absent this subsidy. However, the subsidy's effect depends on the idea value distribution in a field. In addition, the effect of a subsidy also depends on the field-specific cost of developing an innovation, as this will determine how many additional ideas can be pursued with a certain subsidy amount S. Neither the idea value distribution, nor the cost of developing an idea in a field is observed. To work around this, IStraX relies on simple model of innovation to estimate the cost of an innovation and the shape of the idea value distribution based on the observed value distributions of realized innovations in the past.

4.1 A Model of Innovation

In this section, we develop a model of innovation to estimate the shape of the idea value distribution and the fixed cost of developing an idea into an innovation for a specific area of innovation. Our model imposes structural assumptions on the idea arrival rate and the cost of developing an innovation. This allows us to derive theoretical quantile values on the observed distribution of private values of innovations. We match these quantile values to quantile values observed in the data to obtain estimates of the parameters in our model.

Assume that a new innovation first requires an idea. Ideas in a given technology class are heterogeneous in quality δ, and follow a Pareto distribution with the following probability density function (pdf):

$$f(\delta) = \begin{cases} \frac{\alpha\mu^\alpha}{\delta^{\alpha+1}} & \text{if } \delta > \mu \\ 0 & \text{if otherwise} \end{cases} \tag{12}$$

The support of this quality distribution is $[\mu, \infty)$. α is a parameter that determines the curvature of the idea distribution (with higher values leading to more ideas of low quality). An inventor that has an idea will try to innovate using the idea if it generates private financial gain for her. Her payoff at the time of deciding whether to pursue the idea includes a fixed cost c and also takes into account that the outcome is uncertain. For simplicity, we assume that the probability of innovation success is independent of the idea quality and is a draw from a uniform distribution on the interval $[0, \kappa)$ where $\kappa < 1$. The expected private benefit from innovating conditional on having an idea of quality δ is $PV = \epsilon \times \delta$. Which, because the expected value of ϵ is $\frac{\kappa}{2}$, gives $E\{PV|\delta\} = \frac{\kappa}{2}\delta$.

An inventor chooses to innovate if:

$$E\{PV|\delta\} \geq c$$

Consequently, she will only pursue ideas where $\frac{\kappa}{2}\delta \geq c$. We define λ as the lowest quality idea that will be developed where

$$\lambda = \frac{2c}{\kappa} \tag{13}$$

We are interested in the distribution of idea quality conditional on idea development, which can be written as:

$$f(\delta_i|\delta > \lambda) = \frac{f(\delta_i)}{P(\delta > \lambda)} = \begin{cases} \frac{\alpha\lambda^\alpha}{\delta_i^{\alpha+1}} & \text{if } \delta > \lambda \\ 0 & \text{if otherwise} \end{cases}$$

where $P(\delta > \lambda)$ is the likelihood that any new idea is above the minimum quality required to be developed:

$$P(\delta > \lambda) = \int_\lambda^\infty \frac{\alpha\mu^\alpha}{\delta_i^{\alpha+1}} d\delta_i = \frac{\mu^\alpha}{\lambda^\alpha} \tag{14}$$

We can write the distribution of the private values of ideas that will be developed - i.e. the values we can observe - as:

$$P(PV_i = v|\delta > \lambda) = \int \phi(PV_i = v|\delta) f(\delta|\delta > \lambda) d\delta$$

where

$$\phi(PV_i = v|\delta) = \begin{cases} \frac{1}{\delta\kappa} & \text{if } v < \kappa\delta \\ 0 & \text{if } v > \kappa\delta \end{cases} \tag{15}$$

is the density of PV conditional on δ. Together, these expressions yield:[30]

$$P(PV_i = v|\delta > \lambda) = \int_{\max\{\lambda, \frac{v}{\kappa}\}}^{\infty} \frac{f(\delta)}{\delta\kappa} d\delta = \int_{\max\{\lambda, \frac{v}{\kappa}\}}^{\infty} \frac{\alpha\lambda^{\alpha}}{\kappa\delta^{\alpha+2}} d\delta$$

Consequently

$$P(PV_i = v|\delta > \lambda) = \left[-\frac{\alpha\lambda^{\alpha}}{(\alpha+1)\kappa\delta^{\alpha+1}}\right]_{\max\{\lambda, \frac{v}{\kappa}\}}^{\infty} = \begin{cases} \frac{\alpha}{(\alpha+1)\kappa\lambda} & \text{if } \lambda > \frac{v}{\kappa} \\ \frac{\alpha\lambda^{\alpha}\kappa^{\alpha}}{(\alpha+1)v^{\alpha+1}} & \text{if } \lambda < \frac{v}{\kappa} \end{cases} = \begin{cases} \frac{\alpha}{(\alpha+1)2c} & \text{if } 2c > v \\ \frac{\alpha 2^{\alpha} c^{\alpha}}{(\alpha+1)v^{\alpha+1}} & \text{if } 2c < v \end{cases} \tag{16}$$

where the last equality follows from equation 13.

Notice that the density of PV given $\delta > \lambda$ depends only on c and α. This is because c is a sufficient statistic for the combined effect of κ and μ on the density. Because equation 16 describes the observed innovations , we can estimate parameters α and c by fitting it to the observed distribution of private values PV_i in Section 3.3.1.[31] We can also work out the expected value of the distribution of conditional private values. This is:

$$E\{PV_i|\delta > \lambda\} = \left[\frac{\alpha}{\alpha+1}\frac{v^2}{4c}\right]_0^{2c} + \left[-\frac{\alpha 2^{\alpha} c^{\alpha}}{(\alpha+1)(\alpha-1)v^{\alpha-1}}\right]_{2c}^{\infty} = \frac{\alpha c}{\alpha+1} + \frac{\alpha 2c}{(\alpha-1)(\alpha+1)} = \frac{\alpha c}{\alpha-1} \tag{17}$$

The cumulative density is given by:

$$P(PV_i \leq v|\delta > \lambda) = \begin{cases} \frac{\alpha v}{(\alpha+1)2c} & \text{if } 2c > v \\ \frac{\alpha}{(\alpha+1)} + \int_{2c}^{v} \frac{2^{\alpha}\alpha c^{\alpha}}{(\alpha+1)w^{\alpha+1}} dw & \text{if } 2c < v \end{cases}$$

We note that

$$\int_{2c}^{v} \frac{2^{\alpha}\alpha c^{\alpha}}{(\alpha+1)w^{\alpha+1}} dw = \left[-\frac{2^{\alpha}c^{\alpha}}{(\alpha+1)w^{\alpha}}\right]_{2c}^{v} = \frac{1}{(\alpha+1)} - \frac{2^{\alpha}c^{\alpha}}{(\alpha+1)v^{\alpha}},$$

which means that the cumulative density is

$$P(PV_i \leq v|\delta > \lambda) = \Phi^{PV}(v) = \begin{cases} \frac{\alpha v}{(\alpha+1)2c} & \text{if } 2c > v \\ 1 - \frac{2^{\alpha}c^{\alpha}}{(\alpha+1)v^{\alpha}} & \text{if } 2c < v \end{cases}$$

We can invert this to find quantiles of the distribution. Note that $\Phi^{PV}(2c) = \frac{\alpha}{(\alpha+1)}$. Hence, the p quantile is given by:

$$Q^{PV}(p) = \begin{cases} p\frac{(\alpha+1)2c}{\alpha} & \text{if } \frac{\alpha}{(\alpha+1)} > p \\ \frac{2c}{(\alpha+1)^{\frac{1}{\alpha}}(1-p)^{\frac{1}{\alpha}}} & \text{if } \frac{\alpha}{(\alpha+1)} < p \end{cases}$$

[30] Note that $f(\delta_i|\delta > \lambda)$=0 if $\delta < \lambda$ from 12. However, we also have that $\phi(PV_i = v|\delta) = 0$ if $\delta < \frac{v}{\kappa}$. This means that $\phi(PV_i = v|\delta)f(\delta|\delta > \lambda)$ will be zero if either of those conditions is binding. This is the reason for the $\max\{\lambda, \frac{v}{\kappa}\}$ expression that is the lower bound of integration.

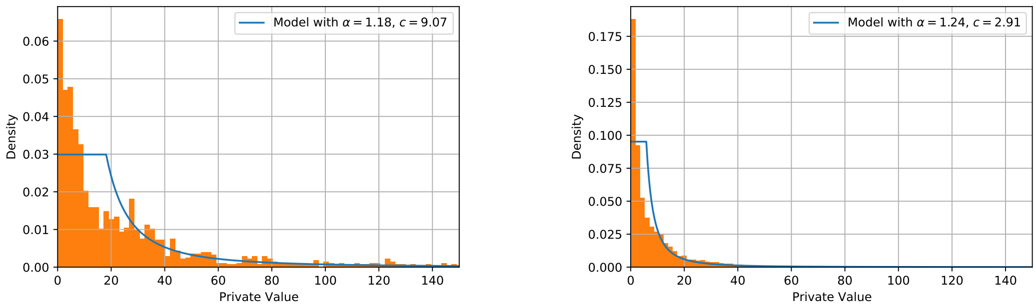
[31] Because 16 won't be differentiable in c and α, we rely on a genetic algorithm to fit the model quantiles to observed ones.

The data give us a p-quantile value for every technology class. Hence, we can estimate parameter values for α and c by matching the model quantiles with the data. While this could be done at any level of technology grouping, we implement the estimation at the level of IPC subclasses and year, using estimates for private values developed in KPSS. This results in time-varying estimates for our parameters that can be grouped to any area of innovation by taking a weighted average across innovations belonging to that area.

To illustrate our parameter estimation, Figure 7 shows the modeled and actual distributions for two prevalent IPC subclasses[32] – one with a high and one with a low estimated cost – for the year 2010. The estimated parameter values for this area produce the blue lines in the graphs. The histograms present the actual data.

Figure 7: Actual vs modeled PV distributions

(a) A61K (b) G06F



Notes: Comparison of actual and modeled private value distributions for two prevalent IPC subclasses. Histogram plots actual private value distribution in the class, blue line shows the modeled density. Private values are based on those estimated in KPSS.

4.2 The return to R&D subsidies

We now turn to constructing our policy instrument IStraX. We can think of a unit of subsidy as changing the private cost of idea development. Let s be the amount with which the private cost of idea development decreases in response to the subsidy. Hence, s directed to an area will affect the quality distribution of ideas that are developed. We let $c' = c - s < c$, be the subsidized cost of idea development. This gives us a new minimum quality threshold, $\lambda' < \lambda$ so that more ideas will be developed. We develop a measure of the social return of an increase in s to the policy maker. We assume[33] that a policy maker cares about the effect of a subsidy on the sum of private and external value, minus idea development costs which we can write as follows:

$$E\{V\} = E\{PV - c + EV|\delta > \lambda\}\, P(\delta > \lambda) \times N \tag{18}$$

[32] A61K: 'Preparations for medical, dental, or toilet purposes' and G06F: 'Electric digital data processing'

[33] We can easily adapt this analysis to other assumptions about what the policy maker cares about.

where N is the number of entrepreneurs that are active in a particular technology area. To examine the effect of (further) subsidies we need to quantify the marginal impact of a change in s

$$\begin{aligned}
\frac{\partial E\{V\}}{\partial s} &= \frac{\partial E\{V\}}{\partial c}\frac{\partial c}{\partial s} \\
&= \left[\left(\frac{\partial E\{PV|\delta > \lambda\}}{\partial c} + \frac{\partial E\{EV|\delta > \lambda\}}{\partial c} - 1\right) P(\delta > \lambda)\right. \\
&\quad \left. + E\{PV - c + EV|\delta > \lambda\}\frac{\partial P(\delta > \lambda)}{\partial c}\right]\frac{\partial c}{\partial s} \qquad (19)
\end{aligned}$$

Given our assumptions about the innovation process it turns out that this can be re-written in a rather compact way as suggested in the following proposition. This will make it easy to quantify a social returns measure from available data as we will discuss in detail below:

Proposition 4.1. *Derivative of Expected Social Value*

$$\frac{\partial E\{V\}}{\partial s} = E\{c + EV(\alpha - \alpha \times \mathbb{I}\{v > 2c\} + \mathbb{I}\{v < 2c\})\,|\delta > \lambda\}\frac{P(\delta > \lambda)}{c} \times N \qquad (20)$$

Appendix D contains the proof of this proposition. Equation 20 only reports the marginal benefit of increasing the policy threshold. We note that the same change of the cost threshold level s will require different amounts of actual support depending on the likelihood of worthwhile ideas emerging. Hence, ex ante expected government costs S of a hypothetical ex post (after idea generation) support level of s will amount to

$$E\{S\} = P(\delta > \lambda) \times s \times N$$

with

$$\frac{\partial E\{S\}}{\partial s} = \left[P(\delta > \lambda) + \frac{\partial c}{\partial s}\frac{\alpha}{c}P(\delta > \lambda)s\right] \times N = \left[P(\delta > \lambda) - \frac{\alpha}{c}P(\delta > \lambda)s\right] \times N$$

where we are using the result in equation 37 in the Appendix. Note that this depends on the level of support already granted. If existing support is non-existing it simplifies to $\frac{\partial E\{S\}}{\partial s} = P(\delta > \lambda)N$

This allows us to work out a measure of the expected net benefit of a fixed amount of government spending across different technology areas, and this is the Industrial Strategy Index:

$$\begin{aligned}
IStraX &= \left(\frac{\partial E\{V\}}{\partial s} - \frac{\partial E\{S\}}{\partial s}\right) \times \left(\frac{\partial E\{S\}}{\partial s}\right)^{-1} \qquad (21) \\
&= \frac{1 + \frac{1}{c}E\{EV(\alpha - \alpha \times \mathbb{I}\{v > 2c\} + \mathbb{I}\{v < 2c\})\,|\delta > \lambda\}}{1 - \frac{\alpha}{c}s} - 1 \qquad (22)
\end{aligned}$$

This expression allows us to calculate subsidy returns at the level of specific technology areas. We assume that c and α are fixed at the level of technology area a. Section 4.1 illustrated how to derive estimates of α_a and c_a. We can estimate $E\{EV(\alpha - \alpha \times \mathbb{I}\{v > 2c\} + \mathbb{I}\{v < 2c\})\,|\delta > \lambda\}$ at technology class a as

$$\begin{aligned}
&E\{EV(\alpha - \alpha \times \mathbb{I}\{v > \hat{2}c\} + \mathbb{I}\{v < 2c\})\,|\delta > \lambda\}\Big|_{a,\kappa} \\
&\qquad = \frac{1}{\#A}\sum_{i \in A} EV_i \times (\alpha_a - \alpha_a \times \mathbb{I}\{v_i > 2c_a\} + \mathbb{I}\{v_i < 2c_a\}) \qquad (23)
\end{aligned}$$

where A is the set of innovations assigned to technology a and $\#A$ denotes the size of that set. We can then compute IStraX as

$$IStraX_a = \frac{1 + \frac{1}{c_a}\, E\left\{EV\left(\alpha - \alpha \times \mathbb{I}\{v > \hat{2c}\} + \mathbb{I}\{v < 2c\}\right)|\delta > \lambda\}\right|_{a,\kappa}}{1 - \frac{\alpha_a}{c_a} s} - 1$$

$$= \frac{1}{\#A} \sum_{i \in A} \frac{1 + \frac{1}{c_a} EV_i \times \left(\alpha_a - \alpha_a \times \mathbb{I}\{v_i > 2c_a\} + \mathbb{I}\{v_i < 2c_a\}\right)}{1 - \frac{\alpha_a}{c_a} s} - 1 \quad (24)$$

In the previous section, we estimated α and c at the level of IPC subclass $\times$ year combinations. Using these estimates thus results in one IStraX estimate for each of these groupings. For most policy purposes, however, we are interested in broader technology groupings consisting of many IPC subclasses (for the analyses in this paper we are interested in broad fields such Clean Energy, Computer Technology, Biotechnology etc.). In addition, or alternatively, we might only be interested in a subset of innovations (e.g. innovations by inventors from a particular country, innovations by universities etc.). To avoid having to estimate α and c for each grouping, we define IStraX at the level of any arbitrary grouping R of innovation-technology tuples with elements (i, a) as:

$$IStraX_R \;=\; \sum_{(i,a) \in R} w_{i,a} \frac{1 + \frac{1}{c_a} EV_i \times \left(\alpha_a - \alpha_a \times \mathbb{I}\{v_i > 2c_a\} + \mathbb{I}\{v_i < 2c_a\}\right)}{1 - \frac{\alpha_a}{c_a} s} \;-\; 1 \quad (25)$$

where $w_{i,a}$ are tuple specific weights. Specifically, as a given innovation is potentially assigned to multiple technology subclasses, we first average across all subclasses of an innovation and subsequently across innovations that fall in a set R so that

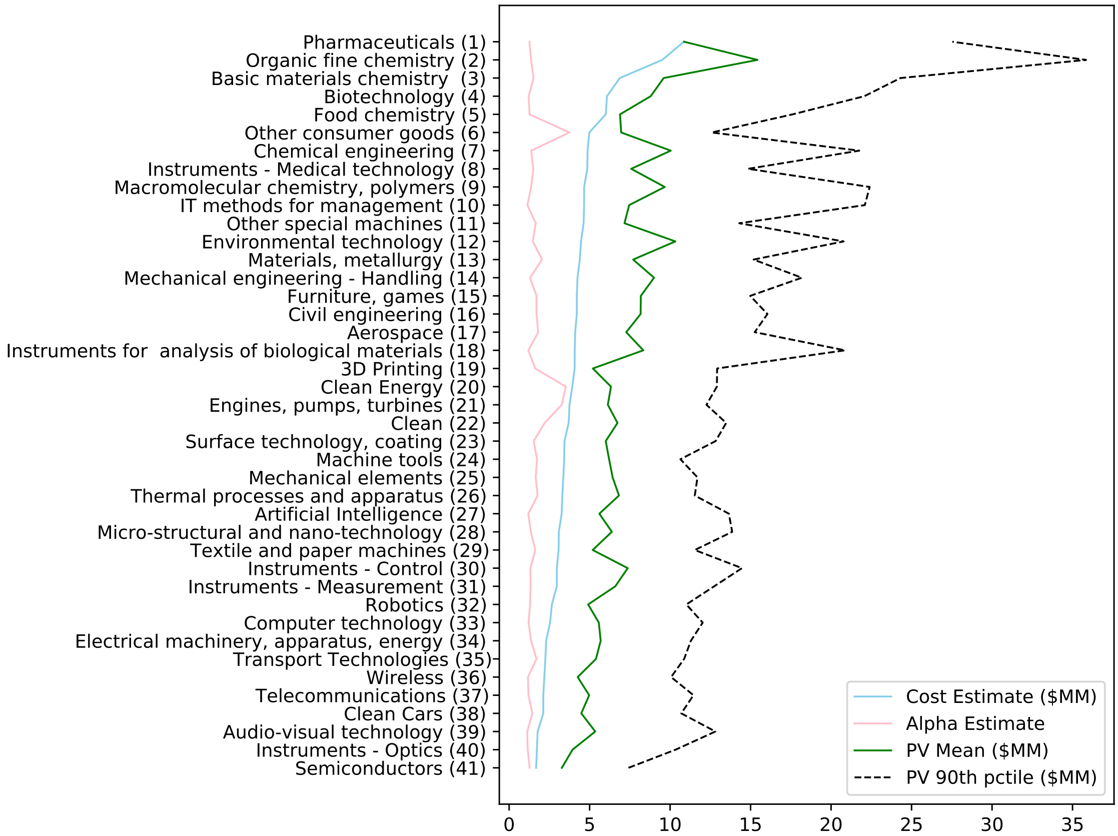
$$w_{i,a} = \frac{1}{\sum_{j \in R} \#a(j)}$$

where $\#a(j)$ is the number of sub-classes innovation j is assigned to.

4.3 IStraX Descriptive Results

IStraX combines the external value information from the first part of the paper with estimates of the responsiveness of innovation to governments subsidies. The latter depends on the R&D costs (c) and the curvature of the innovation private value distribution (α) across technology fields. In this section, we briefly illustrate how our IStraX methodology can be used to rank our broadly defined technological fields by expected returns to subsidies. For this illustration, we estimate IStraX using equation 25, where R groups the innovations in one of our 41 broad technology fields for the time windonw 2005-2014. Figure 8 shows the weighted average of c and α at the level of technological fields in our time window. As explained above, we derive these from a simple structural model of innovation that we fit to the distribution of observed private values in IPC subclass X year combinations. To show the relationship between the parameter estimates and the private value distribution, Figure 8 also reports the mean and 90th percentile of these private value distributions. The figure orders technologies by the estimated R&D cost of a research project. It shows that there is a positive relationship between cost and average private value. In particular, pharmaceutical and chemical sectors rank highly. The relationship between costs and private value is far from monotone, however, because the curvature of the private value distribution varies by field.

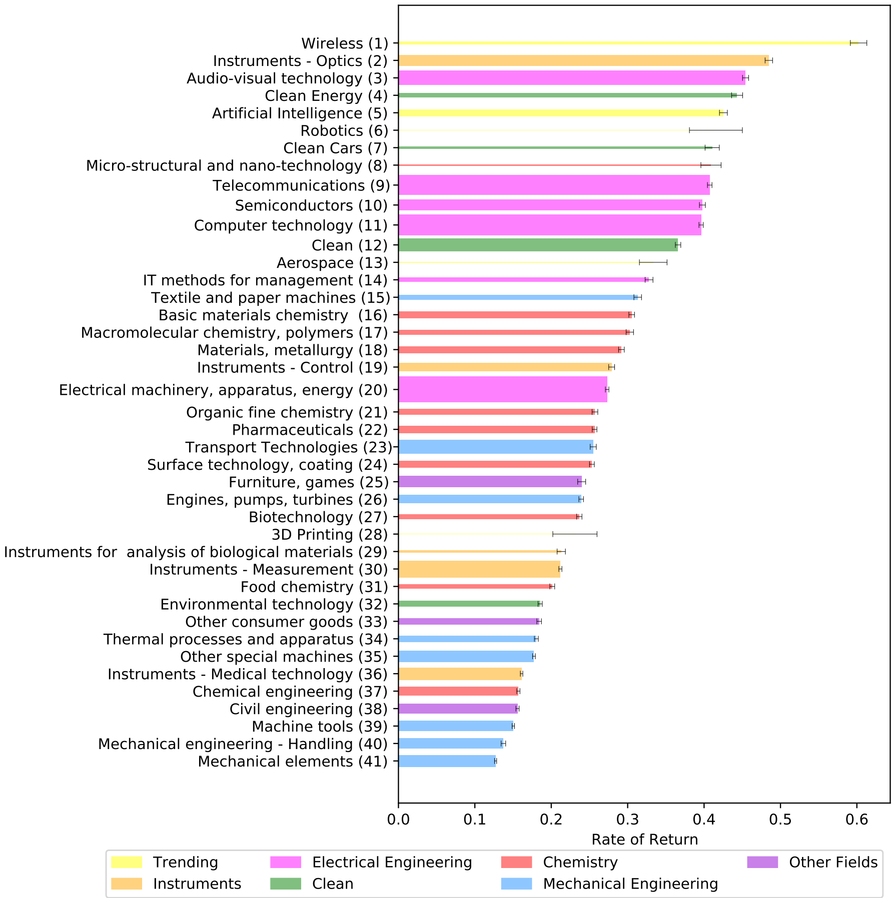
Figure 8: Cost Diagram – Categories – All 2005-2014 innovations



Notes: Diagram of various estimates relevant to the calculation of the IStraX indicator. The blue line shows the estimated cost (in millions of CPI-adjusted 1982 US dollars) of pursuing an innovation idea by technology field (y-axis). The green and dashed lines show the mean and 90th percentile of the estimated private returns distribution in each field (in in millions of CPI-adjusted 1982 US dollars). The pink line shows the estimate of α for each technology field. Calculations are based on innovations for which a patent application was filed in the period 2005-2014.

Figure 9 reports calculations of IStraX across all innovations by technology field. We use EV global as the measure for external value of innovations (in the next section we examine IStraX when accounting for national or supra-national spillovers only). We observe a rather different ranking from either private or external value figures. Categories such as Wireless, Optics, Clean Energy and AI are now at the top with returns to subsidies of more than 40%. For AI and Wireless, these high IStraX values are thanks to relatively low costs (they are ranked 27th and 36th in Figure 8) in combination with high external value (they are ranked 8th and 6th in terms of Global EV). Clean Energy ranks around the median in terms of Global EV (ranked 15th) and costs (ranked 20th), but has among the highest values for α. This implies a large probability mass around the cost threshold, which makes the marginal subsidy relatively effective at increasing innovation.

Figure 9: Global IStraX by Technology - All 2005-2014 Innovations



Notes: Diagram of the rate of return to a subsidy as estimated by IStraX based on EV global (y-axis). Width of each bar represents the number innovations in the field.

5 Results Industrial Policy

5.1 The Case for Targeted Industrial Policy

This section uses IStraX to explore the potential of targeting industrial policy based on the extent of incentive misalignment. To do so, we change our perspective from looking at *global* rates of returns – i.e. IStraX based upon spillovers created globally – to *internal* rates of returns – i.e. restricting spillovers induced within geographic regions. In a first step, we examine the heterogeneity in internal rates of return in technological fields overall[34], and unpack country-level variation in these rates of returns. For both these analyses, larger amounts of variation point at larger amounts of potential welfare increases from targeting R&D support according market failure intensity. In addition, the presence of country-level heterogeneity advises us on whether targeted industrial policy should look different for different countries. In a second step, we examine how different industrial policy based on IStraX would be as compared to other measures that may be used to target fields. To do so, we first compare IStraX to average external and private

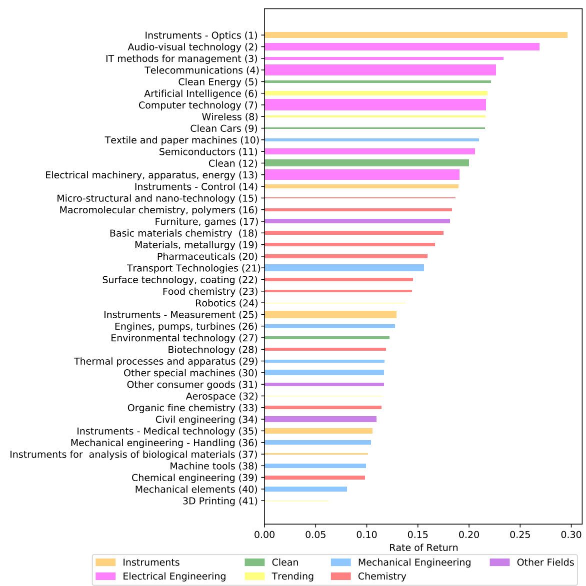
[34]This teaches us whether the heterogeneity across fields from Figure 9 holds when considering internal IStraX for the country of origin.

value across technological fields. Next, we compare the difference between IStraX and 'Revealed Technological Advantage' across technological field for various countries. These analyses are instructive because they examine how R&D support strategies based on our new framework differ from those based on more 'standard' criteria.

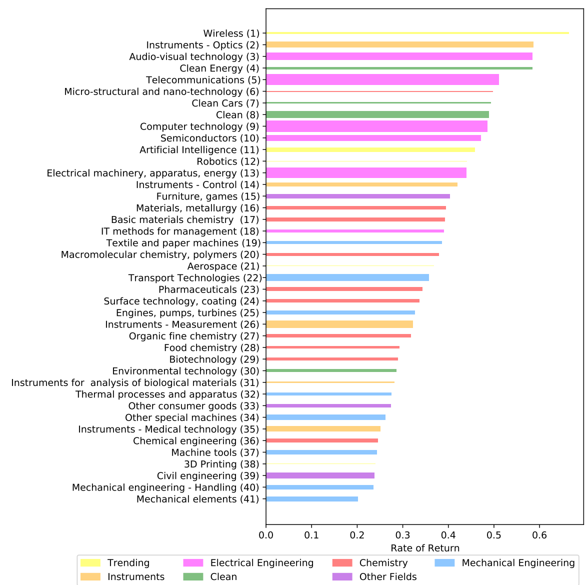
The left-hand side panel of Figure 10 shows internal IStraX for technological fields. It is the result of using only the portion of external value created by each innovation within the country of origin. As such, this figure shows the rate of return to subsidies that are realized within the country of origin in our time window. Our focus here on internal spillovers restricts us to patent families for which at least one inventor could be linked to a country. To allow for comparison, the right-hand side plot shows global IStraX for the same sample. Unsurprisingly, internal rates of returns are much lower – about half across the entire distribution – than global rates of return because only a portion of all spillovers are retained within a country's border. The rankings of fields only change modestly. In addition, we again see considerable variation between technological fields. While low-scoring fields produce return rates around 10%, high-scoring fields have returns between 25% and 30%. This suggests that, no matter whether one is concerned with global or internal spillovers, targeting industrial policy according to the extent of market failure has considerable potential to increase welfare.

Figure 10: IStraX by Technology – All 2005-2014 Innovations

(a) Internal IStraX



(b) Global IStraX



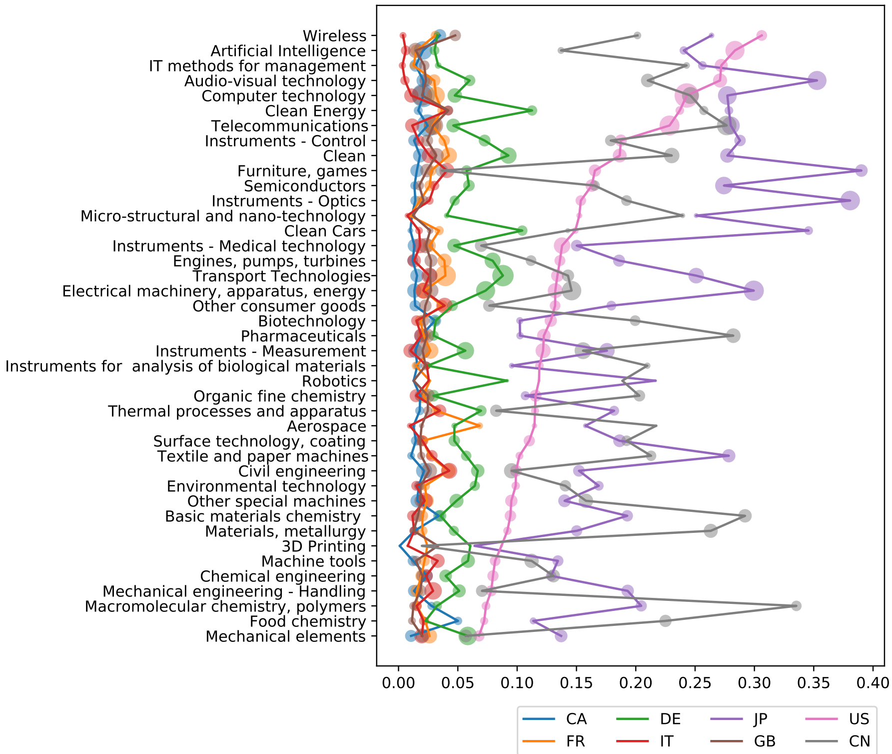
Notes: Diagram of the rate of return to a subsidy as estimated by IStraX (y-axis) based on EV internal (left) and global (right). Width of each bar represents the number innovations in the field.

Figure 11 unpacks country-level variation in IStraX across technological fields. It summarizes IStraX for technology fields for 8 different countries (G7+China). Each line represents a country, and the x-axis shows IStraX for national EV.[35] The size of the circles represents the share of the technology field in a country's total innovation output (equivalent to the width of the bars in previous plots). The graph shows two interesting

[35] Referring to equation 25, a group R here combines all innovations i from a technology field and country combination. PV_i is the private value of innovation i and EV_i is the external value realized within the country of origin. c_a and α_a are calculated for IPC subclass X year combinations.

patterns. First, the average rates of returns vary strongly by country. While Japan, the United States and China show return rates well above 10%, France, Italy, Great Britain (and to a lesser extent Germany) show low return rates barely surpassing 5%. Second, the correlation of IStraX in technology fields between countries is very low (see also Table 2). In fact, many of these correlations are negative, and their (absolute) value rarely exceeds 0.5. These results suggest substantial benefits of tailoring industrial policy by country and leave little hope for a 'one-size-fits-all' approach.

Figure 11: Internal IStraX by Technology and Country - All 2005-2014 Innovations



Notes: Diagram summarizing internal rates of return to a subsidy as estimated by IStraX (x-axis) for 8 countries (G7+China). Size of the circles represents the share of innovations in the field within the country. Fields on the y-axis are ordered by IStraX values of the United States.

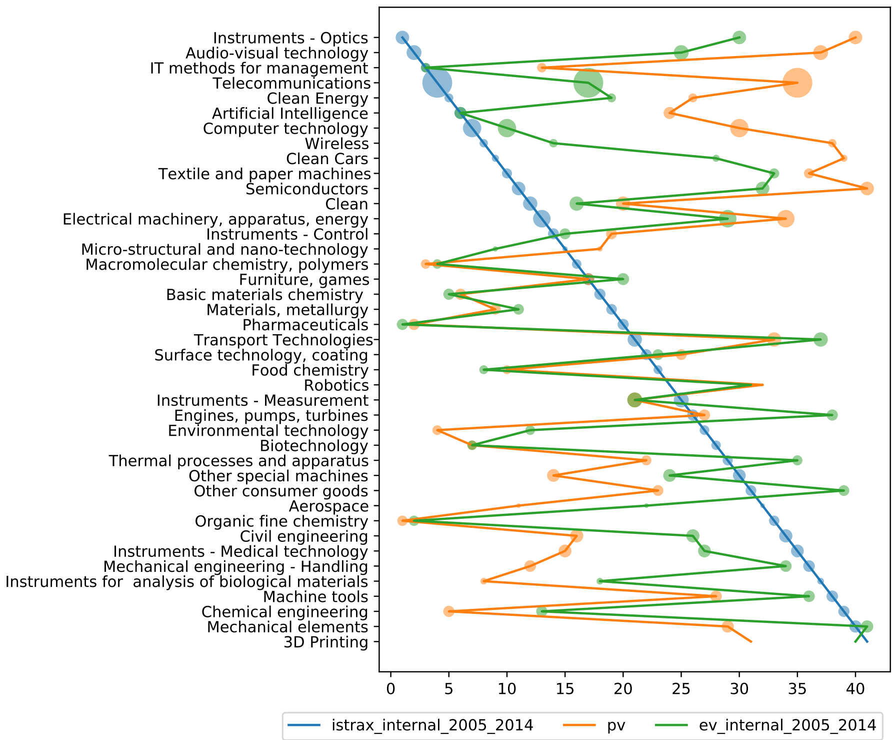
Table 2: Correlation internal IStraX countries

	CA	FR	DE	IT	JP	GB	US	CN
CA	1.00	-0.12	-0.45	-0.13	-0.10	-0.04	0.11	0.44
FR	-0.12	1.00	0.49	0.19	0.13	0.38	0.10	-0.12
DE	-0.45	0.49	1.00	0.39	0.40	0.36	0.06	-0.25
IT	-0.13	0.19	0.39	1.00	0.06	0.06	-0.32	-0.36
JP	-0.10	0.13	0.40	0.06	1.00	0.21	0.60	0.12
GB	-0.04	0.38	0.36	0.06	0.21	1.00	0.47	-0.12
US	0.11	0.10	0.06	-0.32	0.60	0.47	1.00	0.24
CN	0.44	-0.12	-0.25	-0.36	0.12	-0.12	0.24	1.00

Notes: Correlations between the internal IStraX of technological fields of 8 countries (G7+China). Each correlation figure is based on 41 observations, one for each technological field.

Figure 12 uses all[36] innovations in our time window to compare field rankings based on internal IStraX to those based on private returns and internal spillovers. The x-axis displays the rank of technological fields using the different indicators (in descending order of the indicator). To make comparison easy, fields on the x-axis are sorted by the internal IStraX rank. It becomes clear that IStraX implies a very different ranking as compared to private and spillover returns. In fact, the correlation to the rank[37] of IStraX is -0.44 for PV and 0.31 for EV. These results suggest that industrial policy based on the rate of return to subsidies looks substantially different from basing oneself on measures of average private returns and average spillover value created in technology fields.

Figure 12: Comparison internal IStraX, PV and internal EV



Notes: Comparison of technological field rankings (x-axis) according to internal IStraX, PV and internal EV using all innovations in time window 2005-2014. Size of the circles represents the number innovations in the field. Fields on the x-axis are ordered by the ranking based on internal IStraX.

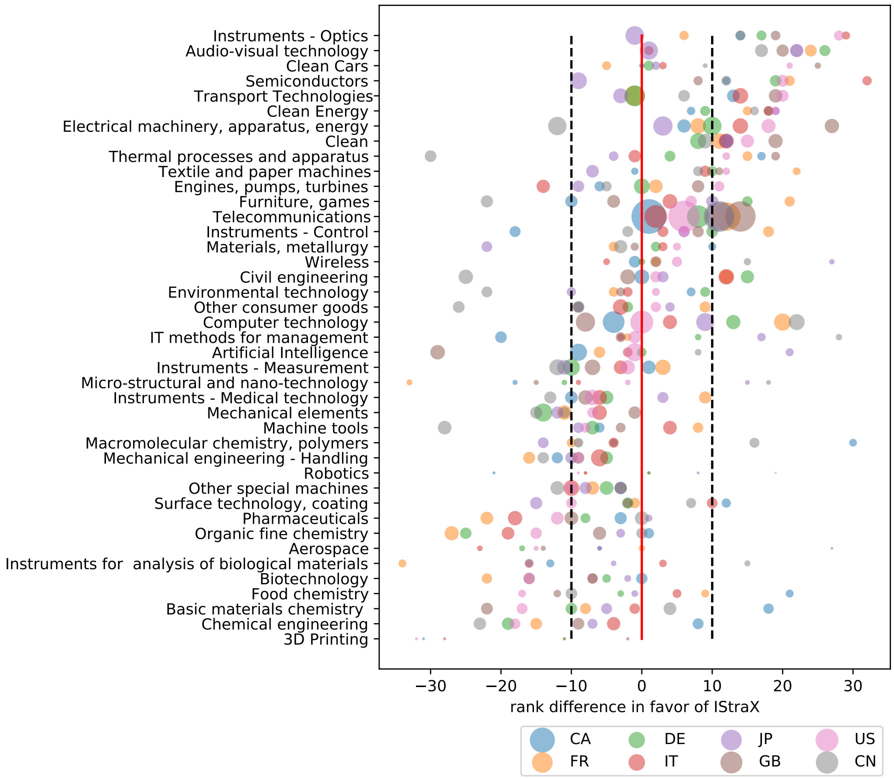
Figure 13 compares rankings based on internal IStraX to those based on Revealed Technological Advantage (RTA) for several countries. RTA is an indicator of relative technological specialization of a country in a certain area of innovation. It is calculated as the ratio between a country's share of patent families in a particular technology field and the country's share in all fields. For each country, we rank technology fields based in internal IStraX and RTA, and we plot the difference in ranks between RTA and IStraX so that a positive (negative) difference implies a higher (lower) rank based on IStraX than based on RTA. We see that rankings considerably differ between the two criteria, with 48% of the 328 field-country combinations displaying a difference in rankings of 10 or higher. Table 3 shows the correlation between the values of RTA and IStraX for technological

[36] Just like in figure 10 we are restricted to including all innovations for which we can assign a country

[37] Using the correlation of the actual values gives a similar picture, with a correlation of -0.38 between IStraX and PV, and 0.22 between IStraX and PV

fields by country. Correlations vary between 0.33 and 0.61 for all countries, except for China which has a correlation of 0.17 only. These results show that ranking fields based on RTA are likely to give substantially different returns as compared to using the IStraX index.

Figure 13: Comparison internal IStraX and RTA



Notes: Comparison of technological field rankings (y-axis) according to internal IStraX and RTA using all innovations in time window 2005-2014. Values on the x-axis are the difference between rankings according to RTA and IStraX. Values above (below) 0 indicate that the field in the country ranks higher (lower) based on IStraX than based on RTA. Size of the circles represents the share of innovations in the field within the country. Fields on the x-axis are ordered by the rank difference of the United States.

Table 3: Correlation internal IStraX and RTA by country

CA	FR	DE	IT	JP	GB	US	CN
0.33	0.4	0.55	0.51	0.61	0.46	0.48	0.17

Notes: Correlations between the internal IStraX and RTA of technolgical fields for 8 countries (G7+China). Each correlation figure is based on 41 observations, one for each technological field.

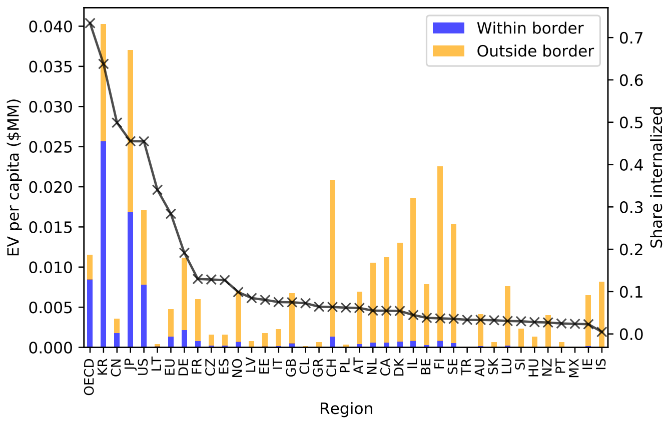
5.2 The Case for Supra-National Industrial Policy

In a final set of analyses we explore the potential welfare benefits of coordinating industrial policy at the supra-national level. As we have seen before, P-Rank allows to restrict the analysis of spillovers to those that are retained within any area of innovation. We have already seen that the returns to subsidies are about double when counting global, as opposed to internal spillovers only. However, this does not imply that coordination is necessarily beneficial, because *rankings* based on internal returns may align with those based on global returns. If that is the case, industrial policy based on what is best locally may produce the highest possible returns globally as well. This section, therefore, examines the extent to which policies that care about internal returns align with those that care about supra-national and global returns.

We kick off with an analysis of the extent to which different regions 'internalize' the spillovers they create. This can inform us about whether there are large differences between countries in terms of the 'ability' to keep spillover returns within their boundary. Then, we look at whether the differences between countries in the internal rates of return hold when looking at global rates of return. This is interesting because it sheds light on what an industrial policy with global interests in mind would look like. Finally, we compare rankings implied by internal as opposed to global IStraX for different countries. This analysis is interesting because it shows how different a globally oriented industrial policy looks from a nationally oriented one, and hence what is the potential value of supra-national coordination.

Figure 14 shows the average within- and across-border spillovers per capita for different countries and regions. It is sorted by the internalization rate – within-border spillovers divided by global spillovers – which is shown by the black line. What stands out is that large countries and regions show much higher internalization rates. Smaller (European) countries create considerable amounts of per capita spillovers, with Finland, Switzerland, Israel and Sweden well above the OECD average. However, their internalization rate is much smaller (around 5%) compared to Korea, China, Japan and US (with rates around 50-60%).

Figure 14: National and international spillovers by country



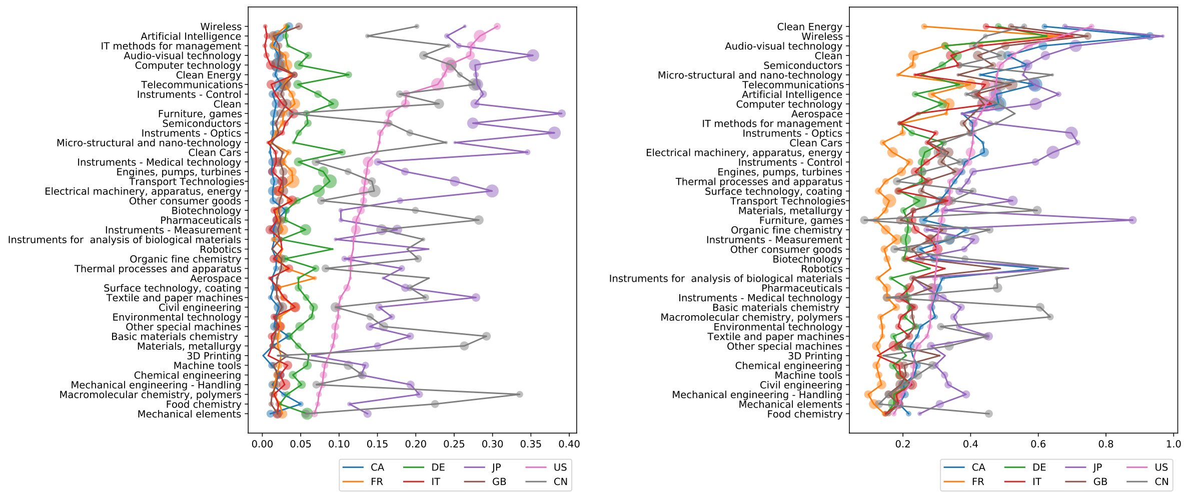
Notes: Comparison between countries of per capita within-border (in blue) and across-border (in yellow) external value (left y-axis) and share of external value internalized (right y-axis) using all innovations in time window 2005-2014. Regions on the x-axis are ranked by internalization rate.

Figure 15 compares the rate of returns for different technology fields and countries. The right-hand panel does so for global IStraX. The left-hand panel includes figure 11 for comparison. Two interesting patterns emerge. First, the global rates of return are much closer to one another for different countries than the internal ones. While global rates of return increase significantly for all countries, they do so by a larger margin for smaller countries like Italy and France than for large countries like the US and China. This is not surprising, because these larger countries are able to internalize much more of the spillovers they create. Second, the correlation between countries in terms of global IStraX is considerably larger than for internal IStraX (see also Table 4). This implies that, from a global perspective, industrial policy needs less tailoring by country.

Figure 15: IStraX by Technology and country

(a) Internal IStraX

(b) Global IStraX



Notes: Diagram summarizing internal (left) and global (right) rates of return to a subsidy as estimated by IStraX (x-axis) for 8 countries (G7+China). Size of the circles represents the share of innovations in the field within the country. Fields on the y-axis are ordered by IStraX values of the United States.

Table 4: Correlation global IStraX countries

	CA	FR	DE	IT	JP	GB	US	CN
CA	1.00	0.84	0.89	0.89	0.75	0.96	0.87	0.43
FR	0.84	1.00	0.83	0.87	0.60	0.87	0.75	0.30
DE	0.89	0.83	1.00	0.88	0.72	0.86	0.85	0.25
IT	0.89	0.87	0.88	1.00	0.74	0.87	0.81	0.26
JP	0.75	0.60	0.72	0.74	1.00	0.70	0.72	0.11
GB	0.96	0.87	0.86	0.87	0.70	1.00	0.86	0.41
US	0.87	0.75	0.85	0.81	0.72	0.86	1.00	0.38
CN	0.43	0.30	0.25	0.26	0.11	0.41	0.38	1.00

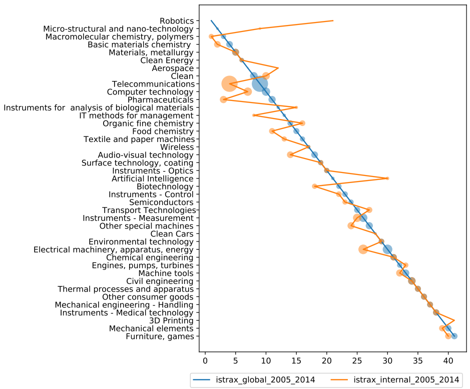
Notes: Correlations between the global IStraX of technological fields of 8 countries (G7+China). Each correlation figure is based on 41 observations, one for each technological field.

Figures 16 and 17 show the difference in rankings implied by global and internal IStraX for 7 countries and the EU. For larger countries (Figure 16) we see that these rankings line up pretty well. The figures are displayed in descending order of the correlation between actual internal and global IStraX values. These correlations vary between 0.92 for China and 0.83 for the US. For smaller countries (Figure 17), we see a much different picture. The rankings of internal and global returns strongly differ, with correlations of 0.53, 0.36 and 0.31 for the UK, Germany and France respectively. The EU (considered as one country) shows a correlation of 0.54, which is between the individual countries and the US.

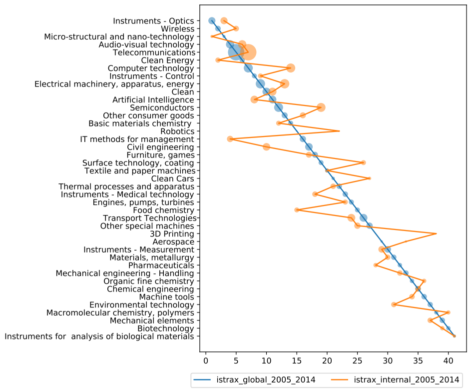
Taken together, the results in this section make a clear case for supra-national coordination of industrial policy – especially for smaller countries. From a global perspective, having individual countries behave according to their optimal industrial policy results in sub-optimal global return rates. This is especially the case for (a number of innovation-intensive) countries that internalize few of their spillovers.

Figure 16: Rankings global and internal IStraX by Technology – Large countries

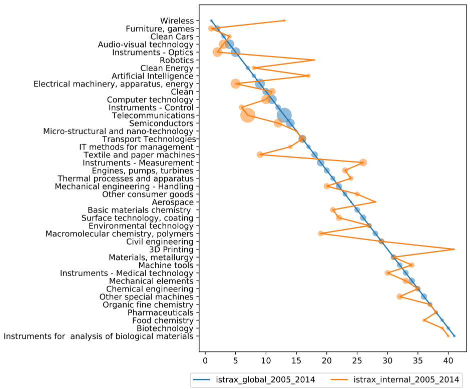
(a) China



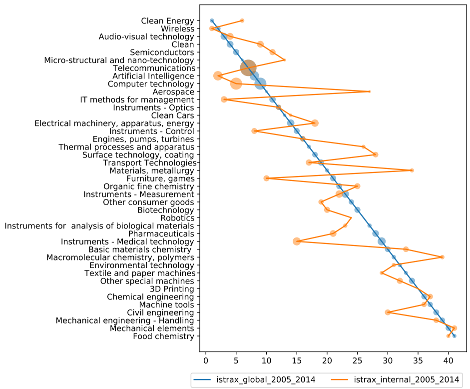
(b) Korea



(c) Japan



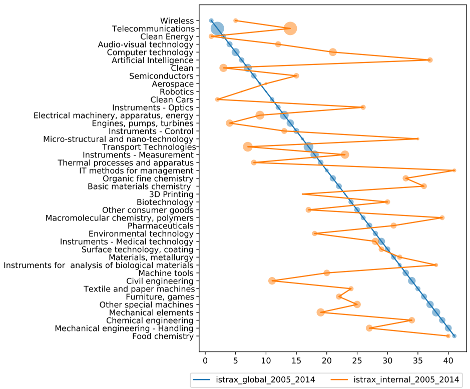
(d) United States



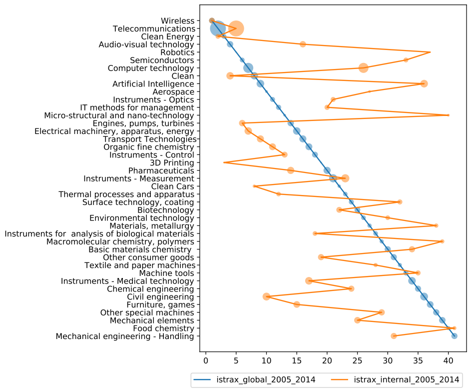
Notes: Comparison of technological field rankings (y-axis) according to internal and global IStraX for large countries based on innovations in time window 2005-2014. Size of the circles represents the number innovations in the field. Fields on the x-axis are ordered by the ranking based on global IStraX.

Figure 17: Rankings global and internal IStraX by Technology – EU27 and small countries

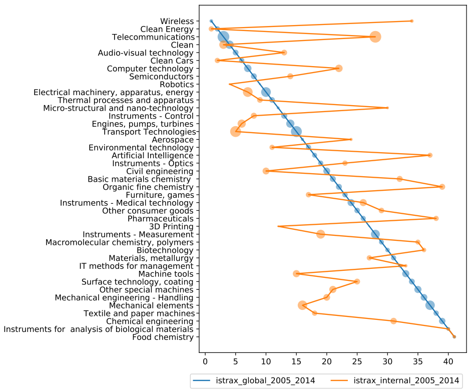
(a) EU27



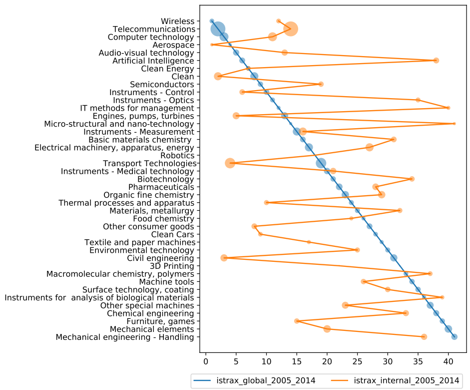
(b) United Kingdom



(c) Germany



(d) France



Notes: Comparison of technological field rankings (y-axis) according to internal and global IStraX for the EU as a whole and a number of smaller countries based on innovations in time window 2005-2014. Size of the circles represents the number innovations in the field. Fields on the x-axis are ordered by the ranking based on global IStraX.

6 Discussion & Conclusion

This paper develops a new framework to measure both direct and indirect knowledge spillovers from patent data, resulting in a new way to quantify the monetary value of spillovers. Overall, the benefits of this methodology include the objectivity coming from data combined with the theoretical properties of knowledge as an input to further innovation. Different versions of our new measure can be used to differentiate between global and national knowledge flows. Moreover, we use estimates of private and spillover returns to develop a new methodology to assess the marginal returns to government subsidies in different sectors or technology areas of an economy.

We apply our methodology to the question of vertically differentiated industrial policy; i.e. targeted support by government for specific sectors or technologies. Varying degrees of knowledge spillovers between different sectors can in principle justify such policies. However, when such differences exist, identifying the set of sectors that deserve special government attention is an empirical question. This paper shows that there is substantial and statistically significant variation in the social returns to government support.

We also show that the set of technologies or sectors that should be supported varies greatly from country to country. It also depends greatly on the desired level of internalization of externalities. Hence, our results provide an interesting starting point for a discussion about specific national, supra-national or indeed sub-national designs of industrial policy. The framework also suggests a set of indicators that could continuously be computed to monitor ongoing efforts by governments.

While these first result stem us positive on the potential of this framework to inspire more successful industrial strategy, a number of caveats warrant attention. First, regarding the way the measure is implemented, it is critical to note that much of the knowledge stock is not embodied in patents. Moreover, we capture only the value of spillovers that are ultimately reflected in private values. Other externalities, such the effect on inequality or the impact on the environment, are muted out in our approach. [38] In addition, private value might reflect monopoly power rather than productivity increases. Another caveat is that these estimates of private and external values are subject to the effects of differences in local institutions. Patent systems are themselves distorting, and some national systems may permit more or less private capture of benefits. Also, factors such as local industry structure and local regulation of prices affect patents' private values. Furthermore, the IStraX policy instrument implicitly assumes that there are no subsidies, made in the absence of any institutional change, that affects existing incentives for innovation.[39] The structural estimates of σ, ϕ, α, δ, c, are all open to debate. Finally, if we use past patterns of spillovers as a guide for current policy, we require those to remain sufficiently stable over time. We have collected some experimental evidence that computed P-Rank values for countries and technological areas are very stable over time. Future research will require further exploration of this pattern.

Future research will mitigate some of the concerns outlined above: First, we aim to improve private value estimates using more sophisticated extrapolation methods including additional indicators of value. Second, we plan to improve our estimate of σ by having the data inform us about the marginal effect of spillovers across areas of innovative activity. Third, we aim to provide further validation of our parameters and the resulting indicators. Finally, we will investigate the extent to which current R&D support is in line with what our methodology suggests in order to better quantify the potential of this framework.

References

Campbell, J. Y., Champbell, J. J., Campbell, J. W., Lo, A. W., Lo, A. W., and MacKinlay, A. C. (1997). *The econometrics of financial markets*. princeton University press.

de Rassenfosse, G., Kozak, J., and Seliger, F. (2019). Geocoding of worldwide patent data. *Scientific Data 2019 6:1*, 6(1):1–15.

DiMasi, J. A., Hansen, R. W., and Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *Journal of health economics*, 22(2):151–185.

Harhoff, D., Scherer, F. M., and Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. *Research Policy*, 32(8):1343–1363.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3):577–598.

[38] That being said, the framework is flexible enough to incorporate such externalities to the extent they can be quantified on the relevant level of analysis

[39] That said, we could build in pre-existing subsidies if such data is available.

Kogan, L., Papanikolaou, D., Seru, A., and Stoffman, N. (2017). Technological Innovation, Resource Allocation, and Growth. *The Quarterly Journal of Economics*, 132(2):665–712.

Lanjouw, J. O. and Schankerman, M. (2004). Patent Quality and research productivity: Measuring innovation with multiple indicators. *The Economic Journal*, 114(April):441–465.

Schmoch, U. (2008). Concept of a technology classification for country comparisons. *Final report to the world intellectual property organisation (wipo), WIPO*.

Tong, X. and Frame, J. D. (1994). Measuring national technological performance with patent claims data. *Research Policy*, 23(2):133–141.

Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. Technical Report 1.

Appendix

A Estimating Private Returns to Innovation

This appendix describes the methodology used to estimate private returns to innovation. To infer the private value of an innovation, we rely on the extrapolation of estimates described in Kogan et al. (2017) – henceforth, KPSS – to our population of inventions. To do this, we use information on technological classification, time, the number of patent applications filed in relation to the innovation and the number of claims in its first patent grant. As this information is available for (nearly) all patented innovations, this approach allows to extrapolate from stock-market-based estimates. Results suggest that our extrapolation models – whilst leaving room for improvement – capture a considerable amount of heterogeneity in the private returns to innovation.

A.1 Estimating private returns using stock market reactions

We use the methodology developed in KPSS to measure the private value of an innovation. In this section, we give an overview of the event study design KPSS employ to obtain these estimates, but refer to their paper for a more detailed description.

Suppose PV_i captures the monopoly rents from exploiting the innovation patented in i. In the absence of any other news, the stock market reaction to the patent grant event is equal to

$$\Delta W_i = (1 - \pi_i) PV_i, \tag{26}$$

where ΔW_i is equal to the difference of a firm's value before and after the moment patent i is granted. π_i is the ex ante probability of any patent being granted conditional on it being public knowledge that the patent application has been made. This expression reflects the assumption that the market knows the value of patent i prior to granting. The day the patent is granted the firm's market value increases by the fraction of the inherent value already known by the market corresponding to the relaxation of the probability that the patent would not be granted. Expression (26) allows to calculate $PV_i = \frac{\Delta W_i}{(1-\pi_i)}$ given an assumption on the ex-ante grant probability. π_i is assumed to be 56% for all patents i, which is the grant rate of US patents between 1991-2001.

This approach to estimating PV_i is subject to the fact that the observed stock market return of any firm might incorporate general movements of the market and unrelated events that might affect stock market returns of the patenting firm. To isolate firm-specific returns that are due to the patent grant, a 'market-adjusted-return model' is used as in Campbell et al. (1997). It specifies the firm's idiosyncratic return R_i (i.e. a firm's return around the event minus the return on the market portfolio), as:

$$R_i = v_i + e_i, \tag{27}$$

where v_i is the portion of the return associated to the patent grant event and e_i is the return's component due to unrelated news around the event date. Replacing ΔW_i with the product of the expected value of W_i conditional on the observed R_i and the market capitalization M_i of the firm on the day prior to the event, expression (26) is rewritten as

$$PV_i = (1 - \bar{\pi})^{-1} E[v_i | R_i] M_i. \tag{28}$$

In their preferred specification, KPSS assume a normal distribution for e_i and a normal distribution truncated at zero for v_i. The variance of e_i, as well as the signal-to-noise ratio (the variance of the distribution of v_i divided by the sum of the variances of v_i and e_i) is estimated from the data (the former is allowed to vary by firm; the latter is assumed constant). These parameter estimates allow to calculate private values for a set of $1,801,879$ patent grants published at the USPTO.

A.2 Extrapolating to the population

As we are interested in the population of innovations, we extrapolate the private value estimates from stock market reactions to all patents in our population – including those for which a KPSS-value is available. To do this, we employ a set of patent characteristics that are plausible predictors of private value.

We start by downloading the KPSS set of private values (version 4, September 6, 2020) and linking them to PATSTAT patent family identifiers using the patent publication number. Some patent families contain multiple US patents in the KPSS set, which means they receive multiple estimates for their private value. About 11.2% of the patents in KPSS belong to such a family. We obtain a family-level estimates by taking the average for these cases. In a second step, we collect information on the predictors for our population of patent families (all families applied for between 2005 and 2014 which can be assigned to a technological field). Specifically, we obtain each family's IPC classes[40], the application filing year of its first application, the number of patent applications in its family[41], and the number of claims for its first publication where it is available. This results in a data set of 15,068,373 patent families, 584,429 of which have been assigned a KPSS private value.

The use of our particular set of predictors will result in eventual estimates of private values that depend on the technological classes an innovation was assigned, its timing, the number of patent applications in its family and the number of claims. While it is quite clear why an innovation's private returns is a function of it's technological content reflected by classes, and of its timing, the latter two predictors might require more context. An innovation might result in multiple patent applications when the innovator is interested

[40]That is, all IPC classes any of its patent applications was assigned to.

[41]Using PATSTAT's DOCDB family definition. This definition assigns patent applications to families based on whether they share the same set of priority application filings.

in obtaining patent protection across different jurisdictions.[42] Because each patent application incurs a cost, more valuable innovations can be argued to result in more patent applications. Therefore, the number of applications associated to an innovation is a value measure available for each innovation in the population (Harhoff et al., 2003). The number of patent claims has been argued to correlate to an invention's breadth (and therefore potential profitability) and found to correlate to its private returns (Tong and Frame, 1994; Lanjouw and Schankerman, 2004).

We assume that the KPSS value provides a noisy but unbiased estimate of the true value of an innovation. Our task, then, is to impose a functional form on the set of predictors that closely reflects the KPSS estimate. Ideally, we do not impose overly stringent assumptions on the shape of the direct and interacted correlations of the predictors to the KPSS estimates. One approach, therefore, would be run a regression of KPSS values on multiple levels of (interacted) fixed effects of highly detailed binned versions of the predictors. Not each of these bins, however, would have (enough) KPSS patents to (reliably) make a prediction. Making the bins large enough so that each family can be assigned a prediction, however, would result in crowded bins (and less precises estimates) for innovation types well-covered in KPSS.

To circumvent having to specify (broad) bins in a fixed-effects regression framework, we design a simple iterative algorithm to extrapolate KPSS values to our population. The principle is to use a more detailed binning for our predictors where the data allow to do so, and gradually decrease the granularity of the bins in order to cover an increasingly comprehensive set of innovations. We start with assigning innovations to bins defined by the combination of IPC groups (8212), filing year (10), family size quantiles (10) and number of claims quantiles (6). This results in 4,927,200 potential bins assigned to innovations. In case a bin has at least 30 KPSS values, we assign the weighted[43] average of these KPSS values to the bin. In the second iteration we decrease the granularity of the bins, and repeat the process. We do so until each innovation only belongs to bins that have KPSS values assigned. Table 5 shows the bin definitions and the fraction of the innovations belonging to that bin. In the final step, we assign private values to innovations based on the most detailed bin it belongs to. If an innovation has multiple IPC classes (hence multiple bins), we take the average value of the most detailed bin for each class. Innovations that were not granted any patent (but filed for an application) are assigned a private value of zero. Note that, for each bin definition, we use all the KPSS values available to compute the weighted average for that bin. By doing this, we use all the information available for each bin definition.

[42]Geographical scope of patent protection is the most important reason one invention results in multiple patent applications. Additionally, one invention might be related to multiple patent applications filed for at one and the same patent authority because of so-called continuing applications (e.g. continuations, continuations-in-part or divisional applications).

[43]If a KPSS patent i has N_i classes, its KPSS value PV_i will count for N_i bins. We calculate the weighted average for bin B as $\frac{\sum_{i \in B} \frac{1}{N_i} PV_i}{\sum_{i \in B} \frac{1}{N_i}}$.

Table 5: Overview bin definitions extrapolation

	Definition bin	Fraction families in bin	Correlation to KPSS values	Correlation (in logs) to KPSS values
(1)	IPC Group X Year X Family Size X Claims Count	5.09%	0.39	0.57
(2)	IPC Group X 2 Years X Family Size X Claims Count	1.49%	0.49	0.59
(3)	IPC Group X 5 Years X Family Size X Claims Count	2.17%	0.46	0.58
(4)	IPC Subclass X Year X Family Size X Claims Count	1.09%	0.50	0.57
(5)	IPC Subclass X 2 Years X Family Size X Claims Count	1.17%	0.51	0.56
(6)	IPC Subclass X 5 Years X Family Size X Claims Count	1.38%	0.50	0.54
(7)	IPC Group X Year X Family Size	25.08%	0.44	0.49
(8)	IPC Group X 2 Years X Family Size	7.96%	0.53	0.53
(9)	IPC Group X 5 Years X Family Size	14.11%	0.45	0.53
(10)	IPC Subclass X Year X Family Size	13.74%	0.49	0.48
(11)	IPC Subclass X 2 Years X Family Size	6.34%	0.49	0.44
(12)	IPC Subclass X 5 Years X Family Size	8.9%	0.42	0.45
(13)	Year X Family Size X Claims Count	0.48%	0.04	0.13
(14)	2 Years X Family Size X Claims Count	0%	N/A	N/A
(15)	5 Years X Family Size X Claims Count	0%	N/A	N/A
(16)	Year X Family Size	11.01%	N/A	N/A

Notes: Overview of the various bin definitions used in the extrapolation process, ordered by decreasing granularity. These bin definitions correspond to different iterations of the extrapolation algorithm. In each iteration, bins with 30 or more KPSS values are assigned an extrapolated private value using the weighted average of KPSS values in that bin. Private values are assigned to innovations based on the most granular bin (which has been assigned a value) they belong to. If an innovation belongs to multiple bins (because of having multiple IPC classes), the average of the (most granular) bins is assigned as private value. The first column defines the bin as a combination of the categorical predictors (cfr. a set of fully interacted fixed effects). The second column shows the fraction of all innovations in our sample that were assigned a value based on the bin definition. In case of multiple bins, we count the innovation with the most granular bin. Columns 3 and 4 show the correlation between KPSS values and extrapolated values (logged in case of column 4) for innovations in the bin that have a KPSS value. For the last 3 bin definitions, no such families exist.

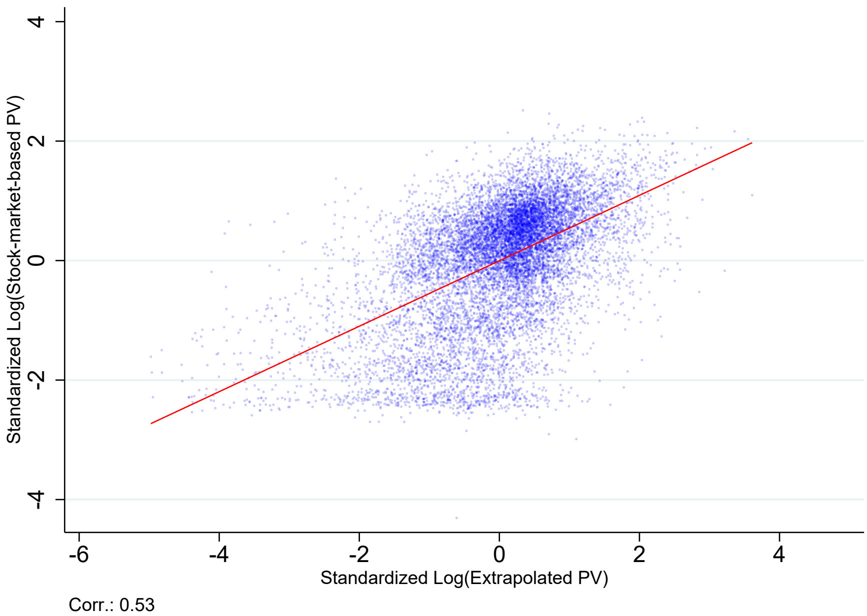
A.3 Extrapolation Descriptive Results

In this section, we examine the fit between our extrapolated private values and those in KPSS. It is important to note here that the KPSS values do not present us with a 'ground-truth set' of private returns to innovation. Indeed, KPSS filter out the expected value of the patent grant based on distributional assumptions of both the true value of the innovation and the value of unrelated news around the patent grant. While this expected value is plausibly close to the true value for any group of innovations, it is likely to contain significant noise at the level of individual innovations. This was also the primary reason for us to only assign the KPSS values to bins with at least 30 observations. That being said, it is instructive to compare the extrapolated values to KPSS in order to ensure that our predictors are able to capture variation present in KPSS estimates.

In a first step, we examine correlations between extrapolated values and KPSS values for innovations that have both (584,429 in our time window). The overall correlation amounts to 0.44 for the 'raw' values, and 0.57 when first taking the natural logarithm. The latter is less sensitive to extreme outliers (potentially based on extremely valuable alternative news) in KPSS. Table 5 examines these correlations for each binning definition. Reassuringly, the bulk of the innovations is assigned a private value based on bins for which this correlation is between 0.39 and 0.53 (0.44 and 0.59 when taking logs first). Only one category, representing less than 0.5% of the population is in a category with a poor correlation of 0.04 (0.13 after logging). We take this as evidence that the fit between the extrapolated value and actual KPSS values does not depend strongly on which binning definition we use. In addition the overall fit seems satisfactory, especially when considering that KPSS values may contain considerable amounts of (white) noise.

To rule out that these correlations are artificially high because of over-fitting, we excluded a test sample of 11,885 from the population of KPSS values before performing the extrapolation procedure. The correlation in this test sample is 0.38 (0.53 when logging first), which is very close to the correlations in the overall sample, and therefore suggests over-fitting is not a problem. Figure 18 uses this test sample and shows a scatter plot of the two (z-standardized) measures of private value. This figure confirms a clear positive (but noisy) relationship between the two indicators.

Figure 18: Comparing extrapolated to stock-market-based estimates of private value



Notes: Correlation between private return estimates based on stock market reaction as in KPSS (y-axis) and estimates using the extrapolation approach. Values are log-transformed and z-standardized. This figure is based on a sample of 11,885 innovations that was excluded from the extrapolation exercise to avoid over-fitting when comparing the extrapolation results to the stock-market-based estimates.

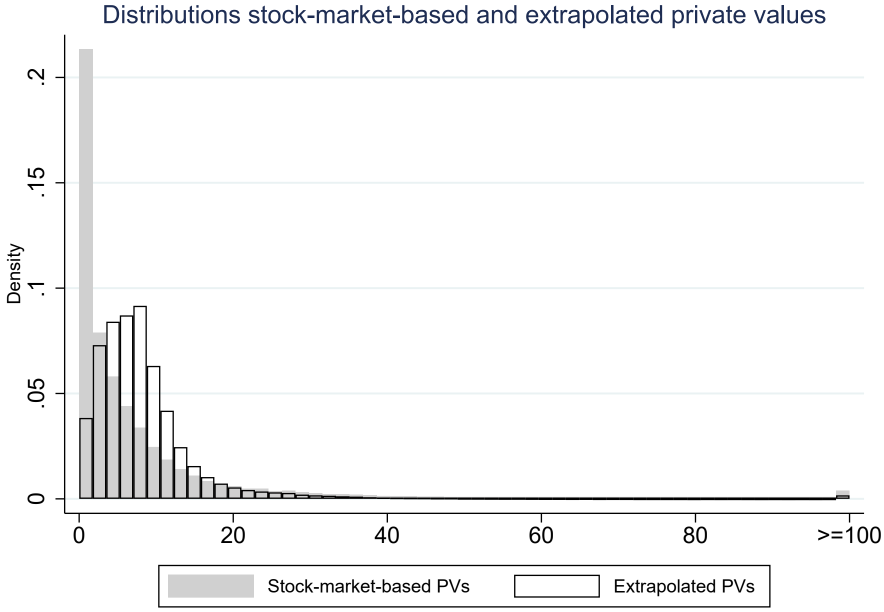
Finally, we compare the distributions of the two measures. Table 6 shows the distribution of KPSS values, as well as the distribution of PV for our entire sample and for the sample for which also a KPSS value exists. Figure 19 shows the distribution for the latter visually. We see that private values for the entire sample are nearly 1.5 million lower than private values of patent families that belong to a US stock-listed firm. To a large extent, this can be explained by the fact that the full sample contains non-granted patents as well, while those in KPSS (by definition) are granted. We also see that the distribution of extrapolated private values are more centered around the mean. This is natural, as they result from taking a weighted average of KPSS values for at least 30 innovations in a bin. Reassuringly, though, we see still plenty of variation in this measure. This indicates that our predictors can capture a great deal of variation present in the KPSS values. This variation is also present when we compare average private returns by country (see Figure 20, which range between 2.5 and 10 million dollars.

Table 6: Comparing distributions of PV and ξ (from KPSS)

	PV	PV (in KPSS)	ξ (from KPSS)
mean	6.87	8.32	8.41
min	0	0.013	0.00022
p1	0	0.32	0.012
p5	0	1.42	0.023
p10	0	2.27	0.060
p25	0	4.17	0.66
p50	6.80	7.06	3.32
p75	10.4	10.2	8.73
p90	14.2	14.8	19.5
p95	18.5	20.3	31.9
p99	33.0	35.8	81.7
max	184.9	166.8	1640.1
count	15068373	584429	584429

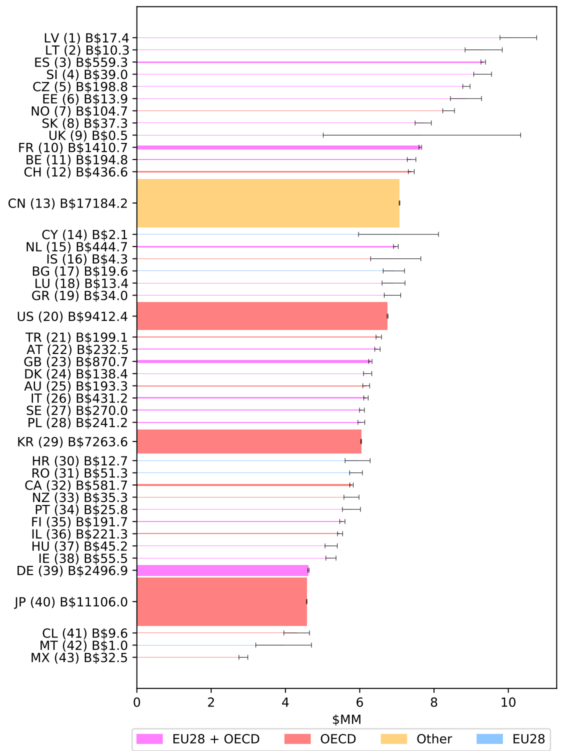
Notes: All values are in million CPI adjusted 1982 US dollars. The first column shows the distribution of extrapolated private values for the entire sample. The second column shows these private values for the sample of innovations that also have a KPSS value. The third column shows the distribution of ξ reported in Kogan et al. (2017) for our sample of patent families.

Figure 19: Comparing extrapolated to stock-market-based estimates of private value



Notes: Distribution of PV and ξ for the sample of innovations that has a KPSS estimate.

Figure 20: Private Returns by Country - All 2005-2014 Innovations



Notes: Diagram of the average private returns in millions of CPI-adjusted 1982 US dollars (x-axis) by country (y-axis). Width of each bar represents the number innovations in the field. Area of each bar (in billions $) represents total private returns in the technology field and is printed next to y-axis labels.

A.4 Concluding remarks

Undoubtedly, improvements to our approach here are possible, but the results strengthen our belief that our approach generates sensible estimates. Results suggest that much of the heterogeneity in stock-market-based estimates of private values can be captured using our extrapolation algorithm with simple, observable predictors.

The estimates produce heterogeneity between countries and–most notably–technological fields (see Figure 3). Given the distinct underlying conditions of innovative activity across technological and geographic domains, this is an expected result. Further analysis of some descriptive patterns suggests that cost factors might be one of the important drivers of private returns. This has inspired the construction of IStraX, the second main indicator constructed in this paper. Indeed, to formulate policy recommendations based upon the value of knowledge spillovers, one should clearly take into account such cost conditions as they do not only affect private, but also public spending on R&D. In other words, if there are important differences in the costs of R&D across technological sectors (which based on these results is likely), any amount of support should be corrected for the cost of one additional innovation project pursued. We take such correction into account when formulating an index that could be used for industrial policy in the next part of this paper. Next to serving as important conceptual input into the remainder of this paper, the work performed here has resulted in an arguably useful side product: a database of private returns that goes beyond stock-listed firms. This is useful because much of the innovative activity targeted in policy agendas involves start-ups and universities for which such estimates are important.

A number of current limitations motivate follow-on work. First, future work can improve upon the extrapolation approach by including additional innovation features that relate to private values of innovations – for instance, renewal information – or by exploring supervised machine learning approaches. Second, the confidence we can place upon our results would increase with more intensive validation of the private returns estimates we have produced. For instance, one could compare the estimates to inventor-given estimates of value, or renewal decisions by firms that reflect the value of an innovation. Finally, the analyses would strongly benefit from a link of the private returns data to information on firms size and organization type in order to further scrutinize mechanisms driving private returns.

B Defining technological fields

Table 7: Concordance between technological fields and IPC/CPC classes

Label	Field	Classes	Scheme
Electrical Engineering	Electrical machinery, apparatus, energy	F21H, F21K, F21L, F21S, F21V, F21W, F21Y, H01B, H01C, H01F, H01G, H01H, H01J, H01K, H01M, H01R, H01T, H02B, H02G, H02H, H02J, H02K, H02M, H02N, H02P, H02S, H05B, H05C, H05F, H99Z	IPC
Electrical Engineering	Audio-visual technology	G09F, G09G, G11B, H04N 3, H04N 5, H04N 7, H04N 9, H04N 11, H04N 13, H04N 15, H04N 17, H04N 19, H04N 101, H04R, H04S, H05K	IPC
Electrical Engineering	Telecommunications	G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N 1, H04Q, H04L, H04N 21, H04W, H03B, H03C, H03D, H03F, H03G, H03H, H03J, H03K, H03L, H03M	IPC
Electrical Engineering	Computer technology	G06C, G06D, G06E, G06F, G06G, G06J, G06K, G06M, G06N, G06T, G10L, G11C	IPC
Electrical Engineering	IT methods for management	G06Q	IPC
Electrical Engineering	Semiconductors	H01L	IPC
Instruments	Instruments - Optics	G02B, G02C, G02F, G03B, G03C, G03D, G03F, G03G, G03H, H01S	IPC
Instruments	Instruments - Measurement	G01B, G01C, G01D, G01F, G01G, G01H, G01J, G01K, G01L, G01M, G01N 1, G01N 3, G01N 5, G01N 7, G01N 9, G01N 11, G01N 13, G01N 15, G01N 17, G01N 19, G01N 21, G01N 22, G01N 23, G01N 24, G01N 25, G01N 27, G01N 29, G01N 30, G01N 31, G01N 35, G01N 37, G01P, G01Q, G01R, G01S, G01V, G01W, G04B, G04C, G04D, G04F, G04G, G04R, G12B, G99Z	IPC
Instruments	Instruments for analysis of biological materials	G01N 33	IPC
Instruments	Instruments - Control	G05B, G05D, G05F, G07B, G07C, G07D, G07F, G07G, G08B, G08G, G09B, G09C, G09D	IPC
Instruments	Instruments - Medical technology	A61B, A61C, A61D, A61F, A61G, A61H, A61J, A61L, A61M, A61N, G16H, H05G	IPC
Chemistry	Organic fine chemistry	A61K 8, A61Q, C07B, C07C, C07D, C07F, C07H, C07J, C40B	IPC
Chemistry	Biotechnology	C07G, C07K, C12M, C12N, C12P, C12Q, C12R, C12S	IPC
Chemistry	Pharmaceuticals	A61K 6, A61K 9, A61K 31, A61K 33, A61K 35, A61K 36, A61K 38, A61K 39, A61K 41, A61K 45, A61K 47, A61K 48, A61K 49, A61K 50, A61K 51, A61K 101, A61K 103, A61K 125, A61K 127, A61K 129, A61K 131, A61K 133, A61K 135, A61P	IPC
Chemistry	Macromolecular chemistry, polymers	C08B, C08C, C08F, C08G, C08H, C08K, C08L	IPC
Chemistry	Food chemistry	A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, C12C, C12F, C12G, C12H, C12J, C13B 10, C13B 20, C13B 30, C13B 35, C13B 40, C13B 50, C13B 99, C13D, C13F, C13J, C13K	IPC
Chemistry	Basic materials chemistry	A01N, A01P, C05B, C05C, C05D, C05F, C05G, C06B, C06C, C06D, C06F, C09B, C09C, C09D, C09F, C09G, C09H, C09J, C09K, C10B, C10C, C10F, C10G, C10H, C10J, C10K, C10L, C10M, C10N, C11B, C11C, C11D, C99Z	IPC
Chemistry	Materials, metallurgy	B22C, B22D, B22F, C01B, C01C, C01D, C01F, C01G, C03C, C04B, C21B, C21C, C21D, C22B, C22C, C22F	IPC
Chemistry	Surface technology, coating	B05C, B05D, B32B, C23C, C23D, C23F, C23G, C25B, C25C, C25D, C25F, C30B	IPC
Chemistry	Micro-structural and nano-technology	B81B, B81C, B82B, B82Y	IPC
Chemistry	Chemical engineering	B01B, B01D 1, B01D 3, B01D 5, B01D 7, B01D 8, B01D 9, B01D 11, B01D 12, B01D 15, B01D 17, B01D 19, B01D 21, B01D 24, B01D 25, B01D 27, B01D 29, B01D 33, B01D 35, B01D 36, B01D 37, B01D 39, B01D 41, B01D 43, B01D 57, B01D 59, B01D 61, B01D 63, B01D 65, B01D 67, B01D 69, B01D 71, B01F, B01J, B01L, B02C, B03B, B03C, B03D, B04B, B04C, B05B, B06B, B07B, B07C, B08B, C14C, D06B, D06C, D06L, F25J, F26B	IPC

Clean	Environmental technology	A62C, B01D 45, B01D 46, B01D 47, B01D 49, B01D 50, B01D 51, B01D 52, B01D 53, B09B, B09C, B65F, C02F, E01F 8, F01N, F23G, F23J, G01T	IPC
Mechanical Engineering	Mechanical engineering - Handling	B25J, B65B, B65C, B65D, B65G, B65H, B66B, B66C, B66D, B66F, B67B, B67C, B67D	IPC
Mechanical Engineering	Machine tools	A62D, B21B, B21C, B21D, B21F, B21G, B21H, B21J, B21K, B21L, B23B, B23C, B23D, B23F, B23G, B23H, B23K, B23P, B23Q, B24B, B24C, B24D, B25B, B25C, B25D, B25F, B25G, B25H, B26B, B26D, B26F, B27B, B27C, B27D, B27F, B27G, B27H, B27J, B27K, B27L, B27M, B27N, B30B	IPC
Mechanical Engineering	Engines, pumps, turbines	F01B, F01C, F01D, F01K, F01L, F01M, F01P, F02B, F02C, F02D, F02F, F02G, F02K, F02M, F02N, F02P, F03B, F03C, F03D, F03G, F03H, F04B, F04C, F04D, F04F, F23R, F99Z, G21B, G21C, G21D, G21F, G21G, G21H, G21J, G21K	IPC
Mechanical Engineering	Textile and paper machines	A41H, A43D, A46D, B31B, B31C, B31D, B31F, B41B, B41C, B41D, B41F, B41G, B41J, B41K, B41L, B41M, B41N, C14B, D01B, D01C, D01D, D01F, D01G, D01H, D02G, D02H, D02J, D03C, D03D, D03J, D04B, D04C, D04G, D04H, D05B, D05C, D06G, D06H, D06J, D06M, D06P, D06Q, D21B, D21C, D21D, D21F, D21G, D21H, D21J, D99Z	IPC
Mechanical Engineering	Other special machines	A01B, A01C, A01D, A01F, A01G, A01J, A01K, A01L, A01M, A21B, A21C, A22B, A22C, A23N, A23P, B02B, B28B, B28C, B28D, B29B, B29C, B29D, B29K, B29L, B33Y, B99Z, C03B, C08J, C12L, C13B 5, C13B 15, C13B 25, C13B 45, C13C, C13G, C13H, F41A, F41B, F41C, F41F, F41G, F41H, F41J, F42B, F42C, F42D	IPC
Mechanical Engineering	Thermal processes and apparatus	F22B, F22D, F22G, F23B, F23C, F23D, F23H, F23K, F23L, F23M, F23N, F23Q, F24B, F24C, F24D, F24F, F24H, F24J, F24S, F24T, F24V, F25B, F25C, F27B, F27D, F28B, F28C, F28D, F28F, F28G	IPC
Mechanical Engineering	Mechanical elements	F15B, F15C, F15D, F16B, F16C, F16D, F16F, F16G, F16H, F16J, F16K, F16L, F16M, F16N, F16P, F16S, F16T, F17B, F17C, F17D, G05G	IPC
Mechanical Engineering	Transport Technologies	B60B, B60C, B60D, B60F, B60G, B60H, B60J, B60K, B60L, B60M, B60N, B60P, B60Q, B60R, B60S, B60T, B60V, B60W, B61B, B61C, B61D, B61F, B61G, B61H, B61J, B61K, B61L, B62B, B62C, B62D, B62H, B62J, B62K, B62L, B62M, B63B, B63C, B63G, B63H, B63J, B64B, B64C, B64D, B64F, B64G	IPC
Other Fields	Furniture, games	A47B, A47C, A47D, A47F, A47G, A47H, A47J, A47K, A47L, A63B, A63C, A63D, A63F, A63G, A63H, A63J, A63K	IPC
Other Fields	Other consumer goods	A24B, A24C, A24D, A24F, A41B, A41C, A41D, A41F, A41G, A42B, A42C, A43B, A43C, A44B, A44C, A45B, A45C, A45D, A45F, A46B, A62B, A99Z, B42B, B42C, B42D, B42F, B43K, B43L, B43M, B44B, B44C, B44D, B44F, B68B, B68C, B68F, B68G, D04D, D06F, D06N, D07B, F25D, G10B, G10C, G10D, G10F, G10G, G10H, G10K	IPC
Other Fields	Civil engineering	E01B, E01C, E01D, E01F 1, E01F 3, E01F 5, E01F 7, E01F 9, E01F 11, E01F 13, E01F 15, E01H, E02B, E02C, E02D, E02F, E03B, E03C, E03D, E03F, E04B, E04C, E04D, E04F, E04G, E04H, E05B, E05C, E05D, E05F, E05G, E06B, E06C, E21B, E21C, E21D, E21F, E99Z	IPC
Trending	Robotics	B25J 9	CPC
Trending	Wireless	H04W	CPC
Trending	3D Printing	B29C64	CPC
Trending	Artificial Intelligence	G06F17, F06N5, G06N3, G10L15, G06F3, G06Q10, G06Q30, G06F9, G06Q50	CPC
Trending	Aerospace	C22F1, C08K3, B64G1, C08G59, C22C21, B64C1, C22C1, C08G73	CPC
Clean	Clean Energy	Y02E	CPC
Clean	Clean Cars	Y02T10	CPC
Clean	Clean	Y02	CPC

C Computing Patent Rank Recursively

To compute patent rank in practice we need to rely on a recursive procedure rather than inverting outright as in equation 6. Here we show that our recursive procedure converges to the actual solution. We can write the vector of social values at a given iteration as

$$V^{(n)} = V^* + \Delta^{(n)} \tag{29}$$

where $\Delta^{(n)}$ captures the difference of the social value at iteration n relative to the actual solution. Using 7 we can write

$$\boldsymbol{V}^* + \Delta^{(n)} = (\boldsymbol{V}^* + \boldsymbol{\Delta}^{(n-1)})\sigma\boldsymbol{\Phi} + PV = \sigma\boldsymbol{\Phi}V^* + \boldsymbol{PV} + \boldsymbol{\Delta}^{(n-1)}\sigma\boldsymbol{\Phi} \tag{30}$$

Because $\boldsymbol{V}^*$ is the actual solution of the equation system we have $\boldsymbol{V}^* = \boldsymbol{V}^*\sigma\boldsymbol{\Phi} + \boldsymbol{PV}$, hence we can re-write equation 30 as

$$\boldsymbol{\Delta}^{(n)} = \boldsymbol{\Delta}^{(n-1)}\sigma\boldsymbol{\Phi} \tag{31}$$

We now need to show that 31 is a contraction mapping. For that consider the characteristic element

$$\Delta_i^{(n)} = \sigma \sum_j \phi_{ij}\Delta_j^{(n-1)} \tag{32}$$

Note that all elements ϕ_{ij} are positive but some elements $\Delta_j^{(n-1)}$ could be negative. Hence

$$|\Delta_i^{(n)}| \leq \sigma \sum_j \phi_{ij}|\Delta_j^{(n-1)}| \tag{33}$$

Summing over all innovations yields

$$\sum_i |\Delta_i^{(n)}| \leq \sigma \sum_i \sum_j \phi_{ij}|\Delta_j^{(n-1)}| = \sigma \sum_j |\Delta_j^{(n-1)}| \sum_i \phi_{ij} \tag{34}$$

Note that $\sum_i \phi_{ij} = 1$ so that we can write

$$\sum_i |\Delta_i^{(n)}| \leq \sigma \sum_j |\Delta_j^{(n-1)}| \tag{35}$$

We can think of the LHS of this as an index of the total error we are making in using the n-th iteration of the the algorithm rather than the actual solution V^*. The equation consequently suggests that the error at iteration n is smaller than at iteration $n-1$ by virtue of $\sigma < 1$. Hence, the error will tend exponentially to zero.

D Proof of Proposition 4.1

Proposition 4.1 provided a simple expression for the marginal effect of changes in government innovation support on S on innovation value V. Here we derive this expression.

Proof. We look at three elements of equation 19 in turn:

Marginal effect on probability of having worthwile idea $\frac{\partial P(\delta>\lambda)}{\partial c}$*:*

Note that

$$P(\delta > \lambda) = \frac{\mu^\alpha}{\lambda^\alpha} = \frac{(\kappa\mu)^\alpha}{(2c)^\alpha} \tag{36}$$

Hence
$$\frac{\partial P(\delta > \lambda)}{\partial c} = -\alpha \frac{(\kappa\mu)^{\alpha}}{2^{\alpha} c^{\alpha+1}} = -\frac{\alpha}{c} P(\delta > \lambda) \tag{37}$$
Hence the derivative is equal to the probability of having a good idea times the ratio between α and c.

Marginal effect on expected private value profits $\frac{\partial E\{PV|\delta>\lambda\}}{\partial c}$*:*

$$\frac{\partial E\{PV|\delta > \lambda\}}{\partial c} = \frac{\alpha}{\alpha - 1} = \frac{1}{c} E\{PV|\delta > \lambda\}$$

Marginal effect on expected external value $\frac{\partial E\{EV|\delta>\lambda\}}{\partial c}$*:*

Firstly note that

$$E\{EV|\delta > \lambda\} = E\{E\{EV|PV\}|\delta > \lambda\} = \int E\{EV|v\} P(v|\delta > \lambda) dv$$

i.e. we can compute the expected value of EV via iterated expectation. Moreover

$$E\{EV|PV, c\} = E\{EV|PV\}$$

i.e. external values depend on the cost threshold c via private values. We can consequently write

$$E\{EV|\delta > \lambda\} = \int_0^{2c} E\{EV|v\} P(v|\delta > \lambda) dv + \int_{2c}^{\infty} E\{EV|v\} P(v|\delta > \lambda) dv \tag{38}$$

Let's look at the derivative of each integral in 38 in turn. For the first integral we get

$$\frac{\partial}{\partial c}\left[\int_0^{2c} E\{EV|v\} P(v|\delta > \lambda) dv\right] = E\{EV|2c\} P(2c|\delta > \lambda) \times 2 + \int_0^{2c} E\{EV|v\} \frac{\partial P(v|\delta > \lambda)}{\partial c} dv$$

using the Leibniz Rule.

Recall from equation 16 that $P(v|\delta > \lambda) = \frac{\alpha}{(\alpha+1)2c}$if $2c > v$, hence

$$\frac{\partial P(v|\delta > \lambda)}{\partial c} = -\frac{P(v|\delta > \lambda)}{c}$$

consequently

$$\begin{aligned}
\int_0^{2c} E\{EV|v\} \frac{\partial P(v|\delta > \lambda)}{\partial c} dv &= -\frac{1}{c} \int_0^{2c} E\{EV|v\} P(v|\delta > \lambda) dv \\
&= -\frac{1}{c} E\{EV|\delta > \lambda, v < 2c\} P(v < 2c|\delta > \lambda) \\
&= -\frac{1}{c} E\{EV \times \mathbb{I}\{v < 2c\}|\delta > \lambda\}
\end{aligned}$$

where the first $\mathbb{I}\{\cdot\}$ is the indicator function. Put differently, we can estimate this integral by averaging over the observed $EV \times \mathbb{I}\{v < 2c\}$ and dividing by c.[44]

For the second integral we get

[44] Note that we could estimate $P(v < 2c|\delta > \lambda)$ as the share of observed innovations with private value smaller than $2c$, which is equivalent to the average of $\mathbb{I}\{v < 2c\}$

$$\frac{\partial}{\partial c}\left[\int_{2c}^{\infty} E\left\{EV|v\right\} P(v|\delta > \lambda)dv\right] = -E\left\{EV|2c\right\} P(2c|\delta > \lambda)\times 2 + \frac{\alpha}{c}\int_{2c}^{\infty} E\left\{EV|v\right\} P(v|\delta > \lambda)dv$$

exploiting once more the Leibniz Rule and recalling from equation 16 that $P(v|\delta > \lambda) = \frac{\alpha 2^{\alpha} c^{\alpha}}{(\alpha+1)v^{\alpha+1}}$if $2c < v$, hence

$$\frac{\partial P(v|\delta > \lambda)}{\partial c} = \alpha\frac{P(v|\delta > \lambda)}{c}$$

Note that

$$\int_{2c}^{\infty} E\left\{EV|v\right\} P(v|\delta > \lambda)dv = E\left\{EV \times \mathbb{I}\{v > 2c\}|\delta > \lambda\right\}$$

i.e. we can estimate this integral by averaging over $EV \times \mathbb{I}\{v > 2c\}$

Combining these results yields

$$\frac{\partial E\left\{EV|\delta > \lambda\right\}}{\partial c} = \frac{\alpha}{c}E\left\{EV \times \mathbb{I}\{PV > 2c\}|\delta > \lambda\right\} - \frac{1}{c}E\left\{EV \times \mathbb{I}\{PV < 2c\}|\delta > \lambda\right\}$$

Consequently we can write the marginal effect on the total innovation value as

$$\begin{aligned}
\frac{\partial E\left\{V\right\}}{\partial c} &= \left[E\left\{PV|\delta > \lambda\right\} - c + \alpha E\left\{EV \times \mathbb{I}\{v > 2c\}|\delta > \lambda\right\} - E\left\{EV \times \mathbb{I}\{v < 2c\}|\delta > \lambda\right\}\right. \\
&\quad \left. -\alpha E\left\{PV + EV - c|\delta > \lambda\right\}\right] \times \frac{P(\delta > \lambda)}{c} \\
&= E\left\{PV - c + EV\left(\alpha \times \mathbb{I}\{v > 2c\} - \mathbb{I}\{v < 2c\}\right) - \alpha\left(PV + EV - c\right)|\delta > \lambda\right\}\frac{P(\delta > \lambda)}{c} \\
&= E\left\{(1-\alpha)\left(PV - c\right) + EV\left(\alpha \times \mathbb{I}\{v > 2c\} - \mathbb{I}\{v < 2c\} - \alpha\right)|\delta > \lambda\right\}\frac{P(\delta > \lambda)}{c} \\
&= E\left\{-c + EV\left(\alpha \times \mathbb{I}\{v > 2c\} - \mathbb{I}\{v < 2c\} - \alpha\right)|\delta > \lambda\right\}\frac{P(\delta > \lambda)}{c}
\end{aligned}$$

where the last equation follows from $E\{PV|\delta > \lambda\} = \frac{\alpha c}{\alpha - 1}$ (see equation 17).Finally becase $\frac{\partial c}{\partial s} = -1$ we get the expression in 20:

$$\frac{\partial E\left\{V\right\}}{\partial s} = E\left\{c + EV\left(\alpha - \alpha \times \mathbb{I}\{v > 2c\} + \mathbb{I}\{v < 2c\}\right)|\delta > \lambda\right\}\frac{P(\delta > \lambda)}{c}$$

$\square$

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1812	Andreas Teichgräber John Van Reenen	Have productivity and pay decoupled in the UK?
1811	Martin Beraja Andrew Kao David Y. Yang Noam Yuchtman	AI-tocracy
1810	Ambre Nicolle Christos Genakos Tobias Kretschmer	Strategic confusopoly evidence from the UK mobile market
1809	Matthew D. Adler Paul Dolan Amanda Henwood Georgios Kavetsos	"Better the devil you know": are stated preferences over health and happiness determined by how healthy and happy people are?
1808	Raphael Calel Jonathan Colmer Antoine Dechezleprêtre Matthieu Glachant	Do carbon offsets offset carbon?
1807	Luke Milsom Isabelle Roland	Minimum wages and the China syndrome: causal evidence from US local labor markets
1806	Matej Bajgar Chiara Criscuolo Jonathan Timmis	Intangibles and industry concentration: supersize me
1805	Luis Bosshart Jeremiah Dittmar	Pandemic shock and economic divergence: political economy before and after the black death
1804	Carlo Altomonte Gianmarco Ottaviano Armando Rungi Tommaso Sonno	Business groups as knowledge-based hierarchies of firms

1803	Nikhil Datta Stephen Machin	Living wages and discontinuities for low-wage workers
1802	Holger Breinlich Dennis Novy J.M.C. Santos Silva	Trade, gravity and aggregation
1801	Ari Bronsoler Joseph Doyle John Van Reenen	The impact of healthcare IT on clinical quality, productivity and workers
1800	Italo Colantone Gianmarco Ottaviano Piero Stanig	The backlash of globalization
1799	Jose Maria Barrero Nicholas Bloom Steven J. Davis	Internet access and its implications for productivity, inequality and resilience
1798	Nicholas Bloom Tarek Alexander Hassan Aakash Kalyani Josh Lerner Ahmed Tahoun	The diffusion of disruptive technologies
1797	Joe Fuller Stephen Hansen Tejas Ramdas Raffaella Sadun	The demand for executive skills
1796	Stephen Michael Impink Andrea Prat Raffaella Sadun	Communication within firms: evidence from CEO turnovers
1795	Katarzyna Bilicka Daniela Scur	Organizational capacity and profit shifting
1794	Monica Langella Alan Manning	Income and the desire to migrate

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE