Universidad del Desarrollo
Facultad de Economía y Negocios

SERIE WORKING PAPERS

A simple index of innovation with complexity

José Fernández Donoso*
Universidad del Desarrollo

Enero 2016 Working Paper 04

A simple index of innovation with complexity

Jose Fernandez Donoso*

Universidad del Desarrollo

Abstract

Patents are the main source of data on innovation. Since most of the innovative activity happens outside of the patenting system, and since patents –and innovations- have different quality, complexity, and impact on each market, unweighted sums of patents and proxies are a bad indicator of a country's innovative activity. I generate a very simple index of innovation that weights patents and exports by a complexity measure. Country rankings using this measure are consistent with market size, GDP per capita, and technological development of each country.

1. Introduction

Patents have become the standard measure for innovation in most disciplines, mostly because it is public and available information. There are, however, numerous concerns that patent counts may be a biased and imperfect measure of innovation. For example, simply adding patents without any measure of the quality of the invention (e.g. inventive step covered by a patent), inflates the measure of innovation for countries where most patents are just small inventive steps from previous inventions. Similarly, the unweighted sum of patents ignores the sophistication and complexity of each innovation, and just assumes that all patents have the same innovative content and impact.

* Facultad de Economía y Negocios. Universidad del Desarrollo. Av. Plaza 680, San Carlos de Apoquindo, Las Condes, Santiago, Chile. E-mail: jgfernandez@udd.cl.

Moreover, most inventive activity happens outside of the patenting system (Moser 2013). Keeping an innovation as a secret can be a dominant strategy over patenting when the cost of secrecy is lower than the risk of "inventing around" by imitators when the innovation is disclosed. There is empirical evidence suggesting that the complexity of the invention is actually a deterrent for imitators, as the cost of copying the new idea (e.g. reverse engineer) increases with complexity (Fernandez Donoso 2014).

Along history, innovation metrics have evolved consistently from input measures of innovation, such as R&D expenditure, to output measures, such as patent counts, and then to composite indicators. The awareness of patents being a biased measure of innovation made composite indices and rankings popular, even though these indices rely heavily on patent counts, and do not take into account the differences in inventive steps across patents. Moreover, these indices use a large number of proxies to account for different types of innovation, and how much innovation these proxies account for is questionable. For example, the Global Innovation Index (2013) counts Wikipedia entries as part of the innovation output sub-index.

How can we accurately measure innovation when most of it stays outside of the formal intellectual property rights system? How does one generate a measure of innovation that incorporates complexity or sophistication differences across inventions? This paper offers a simple, computable and comparable metric to compare innovation across economies, without using large sets of proxies, such as Wikipedia entries, or number of LinkedIn profiles on the web.

Using a very simple method, I generate a normalized index of innovation that incorporates differences in the complexity at the industry level for patents and exports. Though the index is improvable, the rankings of computing the index are consistent with intuitive results, such as the correlation with technological development or the total GDP of the country.

This paper is organized as follows. Next section discusses different measures of innovation used along history and their limitations. Section 3 analyzes the limitations of current innovation metrics, in particular available composite indicators. Section 4 develops an index of innovation with complexity. Final section concludes.

2. Overview of innovation metrics

The first generation of innovation measures, mostly based on input indicators, date from the late 1950s to mid 1960s (e.g. National Science Foundation surveys in the US). Input measures such as R&D investment, S&T personnel, or university graduates in science were typically used as proxies to innovation metrics. Cross-countries R&D comparisons were based on such measures, ignoring the limitations of the definitions of such measures, and the evident endogenous role of governments in using these type of metrics to compare public policies to other countries (e.g. R&D in socialist economies and OECD in the 70s and 80s). The limitations of such measures are self-evident, nonetheless have not been completely ruled out, as there are no available output measures of R&D in such sectors as health or education.

Many contributions intended to accurately measure those activities in R&D that do matter to innovation and technology change, and to develop international standards for R&D measurement. Among them, the Frascani Manual (1981) theoretically breaks up activities that should be excluded from R&D measurement by splitting functions between novelty and routine. If a given activity *"follows an established routine pattern,"* it should be excluded from R&D, while if it *"departs from routine and breaks new ground, it should be qualified as R&D."* As example, collecting weather data should be excluded, while investigating new methods to analyze the data for forecast should be included in R&D measurement.

While this distinction between novelty and routine activities helps to construct an accurate measure of R&D, it does not provide a clear statement of what constitute an innovation, and how to measure it at the firm, industry, and country level. The reason for this lies in the fact that not all innovative activities are developed in specialized laboratories or plants with full-time qualified staff. Measures of R&D are a good statistic to infer professional R&D activity, but they fail to account for important inventions made by private inventors, production engineers, or creative firm staff. Moreover, if this type of "informal" R&D was somehow negatively correlated with the technological complexity of the industry, then R&D measures would underestimate the amount of innovation input for many industries, and particularly for poor and middle-income countries, as their technological development is lower (Fieler 2011).

The second generation measures (1970s-1980s) focused on innovation outputs, such as patent applications, publications, or licensing, among others. Though patenting a new product variety, input, or process requires a fixed cost, depending on the legal system of the jurisdiction where the patent is granted, the inventor would earn a legal monopoly right over its invention. If the monopoly profits over the time of the patent exceed the fixed cost of the patent, one would expect that all profitable innovations ought to be patented.

Consequently, the fact that since 1900 the share of individual patents have declined, while corporate patents have increased their share (Freeman and Soete 2009), means that most innovative activity happens within the boundaries of specialized R&D laboratories and departments of firms, government, and academia. If the patenting story holds, something does not add. According to the 2008 U.S. Census R&D and Innovation Survey (NRDIS), for 85% of surveyed firms, trademarks are not important. Moreover, for 96% of surveyed firms utility patents are not important, and for 95% of them design patents are not important for business. Only by splitting the sample and selecting those firms that engage in formal R&D activity, these numbers decrease (though 67% consider design patents as not important, and 85% thinks of them as not or somewhat important).

In fact, patents have shown to be an imperfect proxy for innovation. First, not all innovations can be patented, as States have exclusions for some innovations. Second, the enforcement of the patent is private, which means that if the patent is imitated without the owner's consent, the owner must take action at nonzero cost, i.e. legal costs and uncertain outcome. If the outcome probabilities depend on the legal costs (e.g. more qualified and expensive lawyers), it is straightforward that smaller firms will patent less than the big players. Third, firms may engage in strategic patenting if the size of a patent portfolio affects bargaining power in patent disputes (Noel and Schankerman 2013), or if it affects the ability of other firms to develop a similar patentable innovation (Stiglitz 2014). Third, if there is a fixed cost of imitation, i.e. product complexity (Fernandez Donoso 2014) or the timing of shorter product cycles (Bilir 2013), there is no incentive to patent an innovation, since the cost of imitation for a potential rival exceeds the profits of imitating. Finally, only "successful" innovations can be patented, meaning that all trial and error are omitted from the measure.

These limitations of patent counts as an output statistic were at the origin of the development of innovation output indicators, many of them based on innovation surveys, within the framework of the Oslo manual (1992). The manual defines innovation as follows: "An innovation is the implementation of a new or significantly improved product (good or service), a new process, a new marketing method, or new organizational method in business practices, workplace organization, or external relations." Even though national innovation surveys are informative of micro-evidence on how firms perceive and fund their innovative activity, the data generated by these surveys is hardly useful for comparative purposes between countries. On one side, not every country administers these surveys on a yearly frequency, while others have never surveyed their firms on their innovative activity. Moreover, surveys differ in questions across countries, and respondents' idea of what constitutes an innovation varies across countries.

The third generation of indexes are super indexes, also known as composite or multidimensional indices. These type of metrics combine different pillars of input and output measures of innovation. The weight of each component depends on the metric. Input measures include institutions, human capital, and market performance. For most of these indices, innovation output measures include formal intellectual property applications, such as patents and trademarks. In addition to intellectual property, output measures include a variety of other statistics, such as published academic papers, ISO 9001 certificates, or license receipts.

3. Limitations of current metrics

Most indices today are complex. This means that several statistics are summed using different weights, and then sorted to present country rankings of innovation. Whether the inclusion and the weight of each measure on the index is questionable, there are two important limitation of these indices: (i) the strong relation with formal intellectual property rights, and (ii) they do not take into account the complexity of each innovation, or the industry where the innovative activity is taking place.

Even though patents and innovation are not perfectly related in these type of indices, most of the output components of these indices rely on innovators formally registering their ideas. As an example, the output components of the Global Innovation Index (GII) include domestic resident patents, trademark and utility models, PCT resident patents and utility models, licensing receipts. Other measures of output in the GII are not necessarily pure innovation output: scientific papers –could be thought as innovation input rather than output-, computer software spending, or FDI outflows as percentage of GDP.

Historical evidence suggests that most innovative activity does not take place inside the formal intellectual property rights system (Fernandez Donoso 2014). Moreover, recent findings suggest that innovations in some industries have shown similar patent rates in countries with very different intellectual property rights regimes (Moser 2013).

As a rule, innovation indices, and in particular the output measures of innovation, do not take into account the complexity of the industry where the innovative activity is taking place. For example, a patent for a simple invention, such as a breastfeeding shirt to avoid cold stomach in the winter, has the same impact on the national innovation metric than devices and methods for transferring data through a human body. This limitation is important, as countries may show higher patenting rates because of strategic reasons (e.g. patent thickets), and with most innovative activity taking place in industries of low complexity, and yet be ranked as more innovative than countries with little patenting rates, but leading exports and drastic innovative activity in highly complex industries.

Furthermore, complexity and the decision of using formal IP are also connected. Indeed, complex inventions need less patent protection, as complexity itself generates additional costs for potential imitators. As inventions are more complex, there are additional learning costs (e.g. reverse engineer) when the innovation is kept in secret instead of made public through patents (Fernandez Donoso 2014).

4. A simple index of innovation with complexity

I propose an indicator that considers the predisposition of innovators to not using formal intellectual property rights and in particular to not using patents, according to the complexity of the industry where the innovative activity is taking place. More explicitly, the index of innovation should take into account three potential problems that current indices do not control. First, the index should account for complexity, either of the industry where the innovation is happening, or the innovation itself. Second, the index should account for innovations taking place outside of the formal intellectual property rights system. Finally, the index should be simple and comparable between countries.

4.1. Complexity weights

There is no unique definition of complexity. Complex systems consist of a large number of elements with no centralized control. In brief, a complex system is a "non-simple" system. In economics, complexity is related to the diversification and sophistication of large economic systems (Hidalgo and Hausmann 2009; Hausmann, Hidalgo et al. 2012). The production of a given country becomes more complex as the sophistication of the products it produces, and the number of country destinations of its exports are larger. This definition is useful to analyze large economic systems, such as countries, using holistic measures of production characteristics. However, it does not say much about the complexity of each product or service.

An ideal measure of industry level complexity would take into account both the number of inputs used to produce a specific product (Hidalgo and Hausmann 2009; Nunn 2007), as well as the complexity of the tasks involved to produce it (Naghavi, Spies, and Toubal 2015). For illustration purposes, in this paper I use the normalized index of Naghavi, Spies and Toubal (2015) based on labor statistics. The index uses survey data for 809 occupations collected by the U.S. Bureau of Labor's Occupational Information Network (O*Net), and industry occupations from the U.S. Bureau of Labor Statistics' Occupational Employment Statistics (OED). As in Costinot et al. (2011), it assumes that all countries have access to the same production technology.

4.2. Patents and exports complexity

An important limitation when analyzing patents, and probably one of the reasons to simplify the measures of innovation to unweighted sum of patents, is the lack of a unique accepted correspondence between patent classifications and product classifications. There are currently different published attempts that take into account the fact that one patent may be useful in different industries (Schmoch et al. 2003; Lybbert and Zolas 2014). For illustration purposes, I use a very simple concordance (Fernandez Donoso 2014) based on the similarities of each title (e.g. patents for "tobacco; cigars; cigarettes; smokers' requisites" were matched to the industry "tobacco products").

4.3. An example of innovation index with complexity

As an example of complexity weighting, I generate an index of innovation based only on innovation outputs. The innovation output sub-index of the Global Innovation Index is comprised of two pillars: knowledge and technology outputs (unweighted patents and utility models, and published articles in peer-reviewed journals), and creative outputs (trademarks and other proxies such as newspapers' circulation, printing output, or Wikipedia entries). In this example, I restrict the innovation output to two main variables: complex inventions with formal IP (patents), and production of complex goods.

For the complexity weights, I use the normalized complexity index by Naghavi, Spies, and Toubal (2015).[1] Then, I generate a complexity-weighted sum of patents and exports, and I normalize the two sums to a [0,1] scale using the min-max method. Finally, I compute the unweighted average of these two normalized measures. As a robustness exercise, I also generate a per capita index, which follows the same calculations but using patents per capita and exports per capita. Nevertheless, the per capita index is not suitable to analyze the overall innovative output of each country.[2]

Since I use 2010 patents' data, and 2011 exports data, the results are comparable to the 2013 Global Innovation Index. The rankings with complexity for the 63 computed countries are presented in Table 1. The numbers in parenthesis are each country's ranking position in the 2013 Global Innovation Index.

Table 1: Country Innovation Ranking (with complexity)

Rank	**Country**	**Rank**	**Country**	**Rank**	**Country**
1 (5)	United States	22 (66)	India	43 (58)	South Africa
2 (35)	China	23 (28)	Czech R.	44 (37)	Croatia
3 (22)	Japan	24 (19)	Australia	45 (41)	Bulgaria

[1] See Appendix for details.
[2] For innovation per capita index see Appendix.

4 (15)	Germany	25 (31)	Hungary	46 (25)	Estonia
5 (20)	France	26 (49)	Poland	47 (48)	Romania
6 (3)	Great Britain	27 (64)	Brazil	48 (46)	Chile
7 (18)	Korea	28 (32)	Malaysia	49 (13)	Iceland
8 (11)	Canada	29 (16)	Norway	50 (83)	Ecuador
9 (4)	Netherlands	30 (7)	Hong Kong	51 (85)	Indonesia
10 (29)	Italy	31 (57)	Thailand	52 (92)	Morocco
11 (1)	Switzerland	32 (62)	Russia	53 (108)	Egypt
12 (21)	Belgium	33 (68)	Turkey	54 (33)	Latvia
13 (23)	Austria	34 (17)	New Zealand	55 (54)	Serbia
14 (8)	Singapore	35 (60)	Colombia	56 (84)	Kazakhstan
15 (2)	Sweden	36 (30)	Slovenia	57	Cuba
16 (26)	Spain	37 (12)	Luxembourg	58 (77)	Belarus
17 (6)	Finland	38 (34)	Portugal	59 (73)	Georgia
18 (63)	Mexico	39 (36)	Slovakia	60 (70)	Tunisia
19 (14)	Israel	40 (55)	Greece	61 (65)	Bosnia and H.
20 (10)	Ireland	41 (90)	Philippines	62 (79)	Dominican R.
21 (9)	Denmark	42 (71)	Ukraine	63 (59)	Armenia

Since this is not a per capita index, there should be a strong correlation between the market size, or total GDP, and the capacity to generate innovation outputs. The correlation between these two variables is 0.97. This importance of size is not trivial. Using the Global Innovation Index methodology, Switzerland or Sweden score higher than the United States, suggesting that these countries generate more innovative outputs than the U.S. The result is at least controversial. Moreover, China scores extremely low (ranked 35, below Latvia, Malta, or Slovenia), which seems unlikely for the country of companies such as Alibaba, Lenovo, or Huawei.

Table 2 shows the results of a simple linear regression between the innovation index and total GDP. Innovation is statistically significant at level 0.001, and the r-squared shows that innovation adjusts very smoothly to country GDP.

Table 2: Innovation index and GDP regression

	GDP
Innovation Index	1.37e+13***

Constant	2.519e+11***
N	63
R2	0.9438

* p<.05; ** p<.01; *** p<.001

Another important correlation is the level of technology development and the innovation output of a country. To test for this correlation, in Table 3 I use Fieler's (2011) index of country technological development, which is basically a residual of Eaton and Kortum's (2002) bilateral trade gravity regression. The correlation of these two variables is 0.72, and the linear regression coefficient is presented in Table 3.

Table 3: Innovation index and technological development

	Technology
Innovation Index	0.916***
Constant	-0.00237
N	63
R2	0.5233

* p<.05; ** p<.01; *** p<.001

This relation does not imply causality between the two variables. Nevertheless, it is suggestive that, even at this very simple stage of a composite index of innovation with only components weighted by complexity, the data generated is consistent with very intuitive results.

Conclusion

Although patents are still the most popular measure of innovation, there have been important improvements to tackle the shortcomings of counting patents. Still, most composite indicators still rely heavily on patent counts. In this paper, I proposed a simple method to reduce the bias of counting patents.

By weighting patent counts, and other non-patent measure of innovation, with the complexity of the product, invention, or index, complex inventions gain a higher weight. Countries with more complex or sophisticated exports and patents rank better in the innovation ranking, and this result is consistent with how more innovative countries should correlate with GDP or technological development.

The main message of this paper is simple: instead of adding large sets of proxies with questionable relation to innovation, composite indices should weight their innovation metrics with an appropriate metric of the quality of the innovation.

References

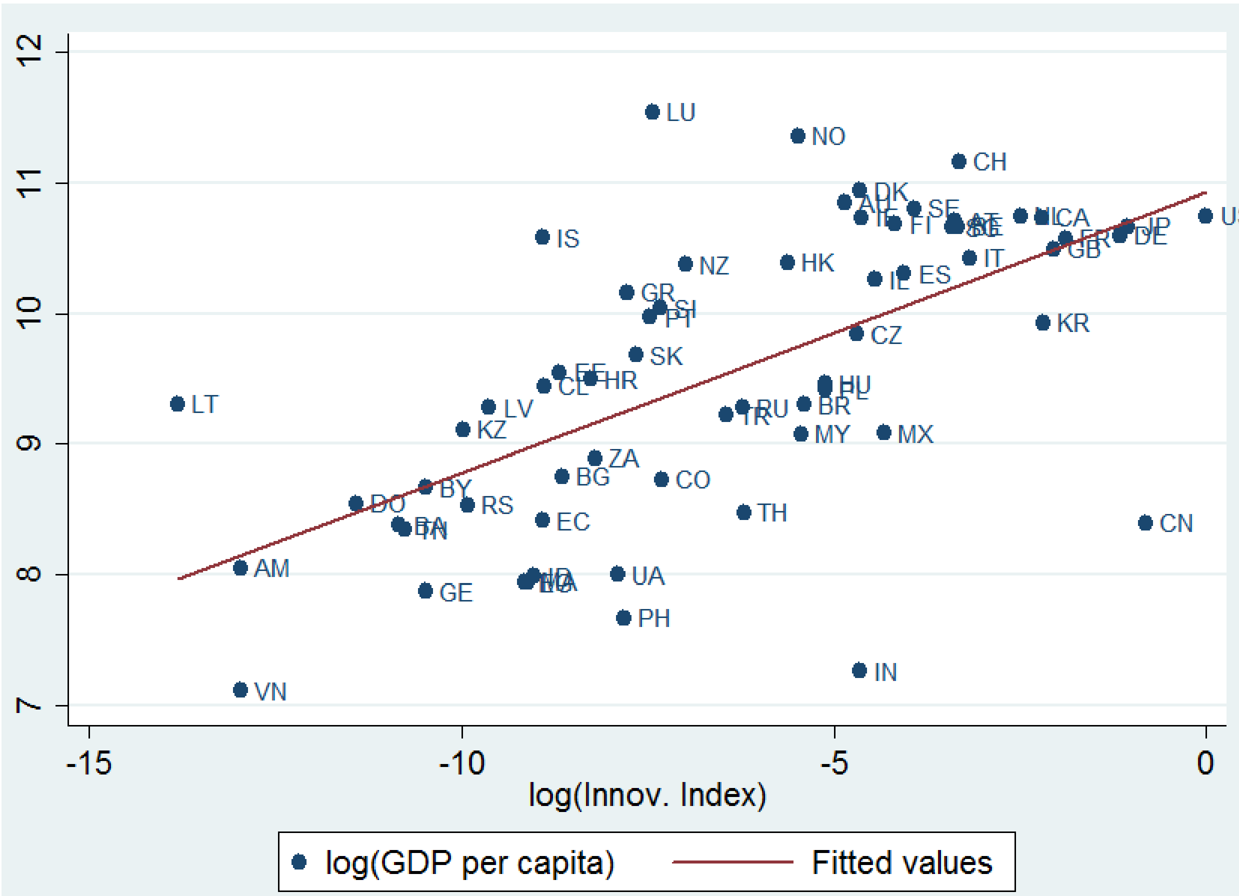
- Bilir K. 2013. Patent Laws, Product life-cycle lengths, and multinational activity. *The American Economic Review*, Vol. 104-7, pp. 1979-2013(35).
- Costinot A., Oldenski L., and Rauch J. 2011. Adaptation and the Boundary of Multinational Firms. *Review of Economics and Statistics*, vol. 93-1, pp. 298-308.
- Eaton J., and Kortum S. 2002. Technology, Geography, and Trade. *Econometrica*, Vol. 70-5, PP. 1741-1779.
- Fernandez Donoso J. 2014. Do complex inventions need less international patent protection? *Economics Letters*, Vol. 125-2, pp. 278-281.
- Fieler A.C. 2011. Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation. *Econometrica*, Vol. 79-4, pp. 1069-1101.
- Freeman C., and Soete L. 2009. Development science, technology and innovation indicators: what we can learn from the past. *Research Policy*, Vol. 38-4, pp. 583-589.
- Global Innovation Index. The local Dynamics of Innovation. 2013.
- Haumann C., Hidalgo C., Bustos S., Coscia M., Simoes A., and Yildirim M. 2013. The Atlas of Economic Complexity: Mapping Paths to Prosperity. *Massachusetts Institute of Technology and Center for International Development, Harvard University*.
- Hidalgo C., and Hausmann R. 2009. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 106-26.
- Lybbert T., and Zolas N. 2014. Getting patents and economic data to speak to each other: An 'Algorithmic Links to Probabilities' approach for joint analyses of patenting and economic activity. *Research Policy*, vol. 43-3, PP. 530-542.
- Moser P. 2013. Patents and Innovation: Evidence from Economic History. *The Journal of Economic Perspectives*, Vol. 27-1, pp. 23-44.
- Naghavi A., Spies J., and Toubal F. 2015. Intellectual Property Rights, Product Complexity, and the Organization of Multinational Firms. *Canadian Journal of Economics*, Vol. 48-3.
- Noel M., and Schankerman M. 2013. Strategic Patenting and Software Innovation. *The Journal of Industrial Economics*, Vol. 61-3, pp. 481-520.
- Nunn N. 2007. Relationship-specificity, Incomplete Contracts and the Pattern of Trade. *Quarterly Journal of Economics*, Vol. 122(2), pp. 569-600.
- Stiglitz J. 2014. Intellectual Property Rights, the Pool of Knowledge, and Innovation. *NBER Working Paper* No. 20014.
- Schmoch U., Laville F., Patel P., and Frietsch R. 2003. Linking Technology Areas to Industrial Sectors. *Final Report to the European Commission, DG Research*.

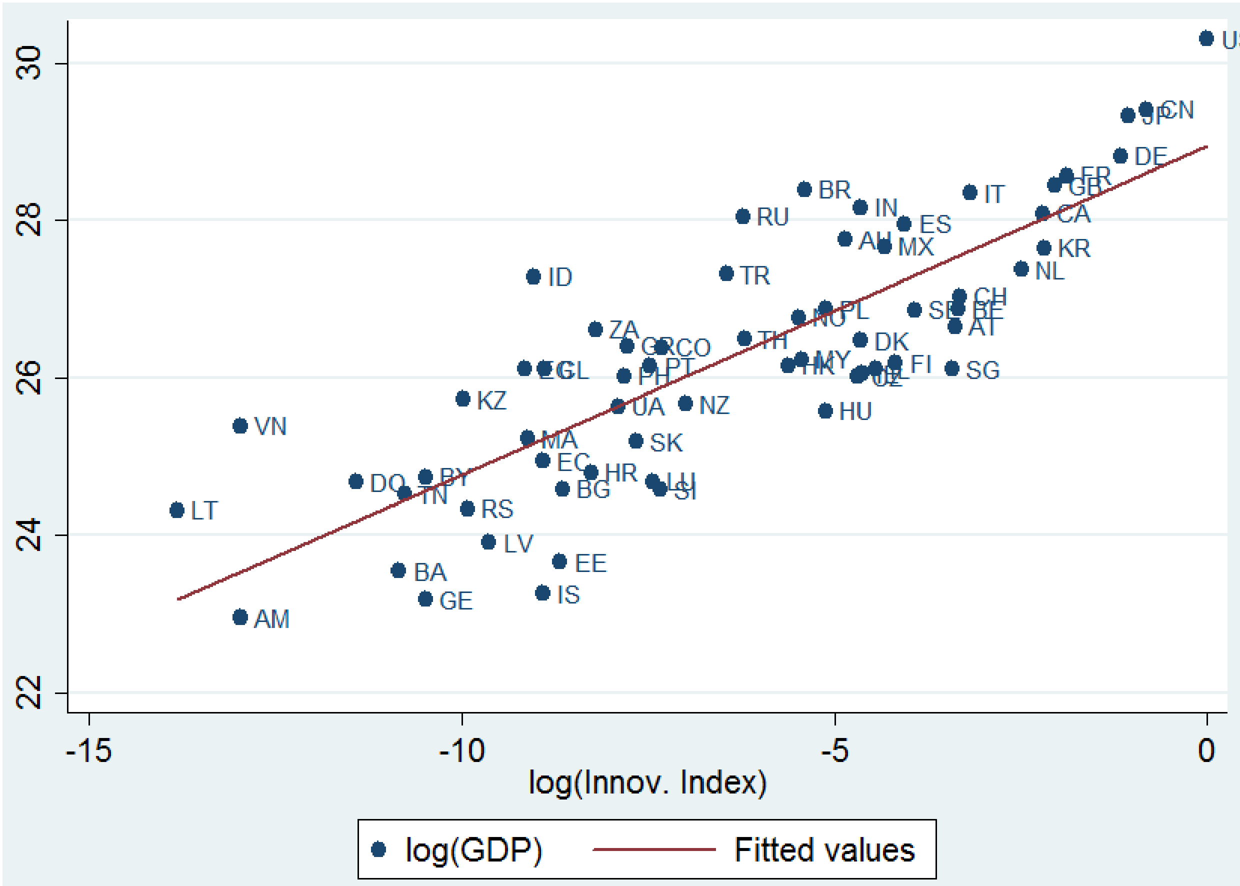
Appendix

Complexity Index: Naghavi, Spies, and Toubal (2015)

O*Net provides information on the importance and level of complex solving skills for 809 eight digit SOC occupations. Each occupation o embodies a complexity of $i_o{}^{\alpha} + l_o{}^{\beta}$, where α and β are the contributions of two complexity components: importance $i \in [1,5]$, and level $l \in [1,7]$. The different scales of complexity components are normalized to a [0,1] scale using the min-max method. Complexity is then merged with employment information from the U.S. Census of Labor Statistics' Occupational Employment Statistics (OES). The data contains the number of employees by occupation in every three digit SIC classification. The occupational intensity, $b_o{}^k$ of each industry k is given by $b_o{}^k = L_o{}^k / \sum_o L^k$, where $L_o{}^k$ is the employment level of occupation o in industry k.

Innovation Index Plots



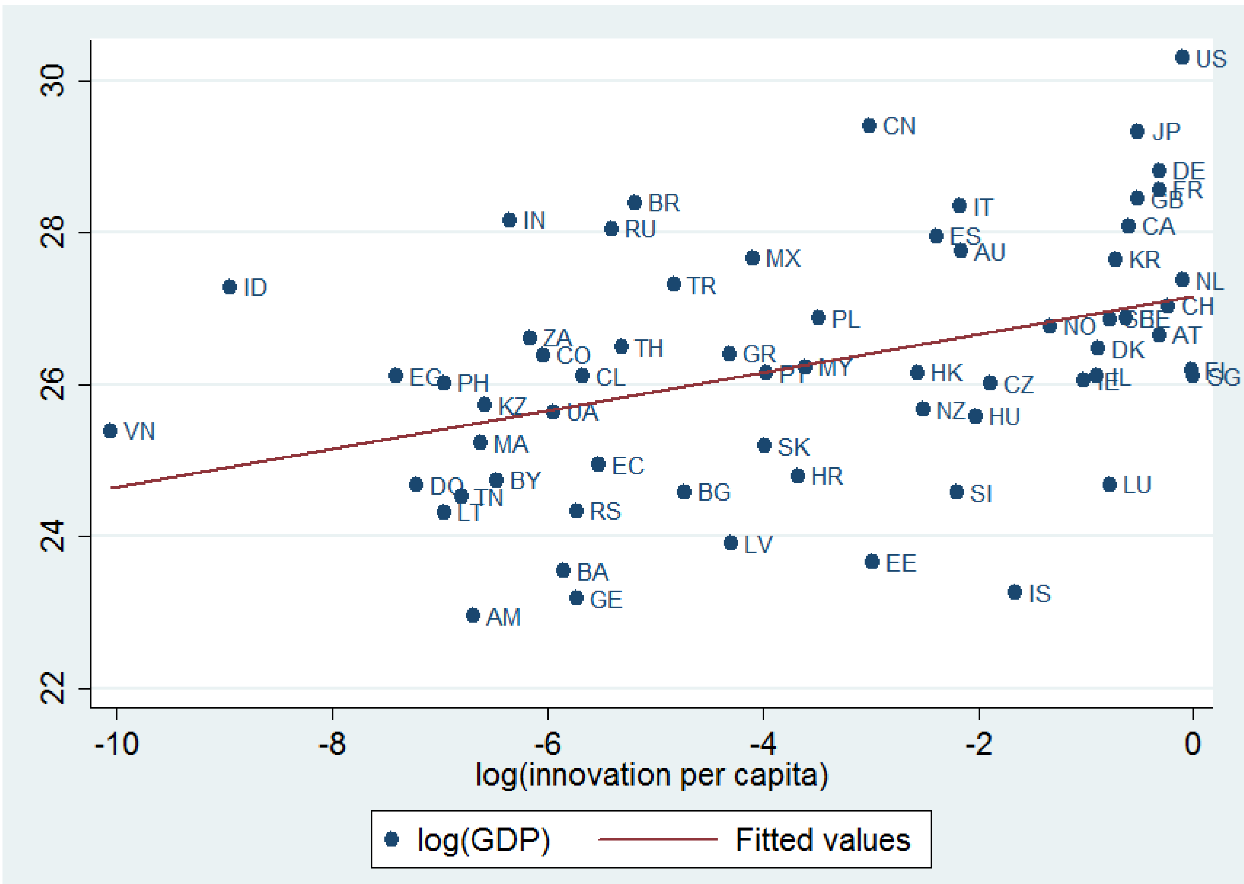


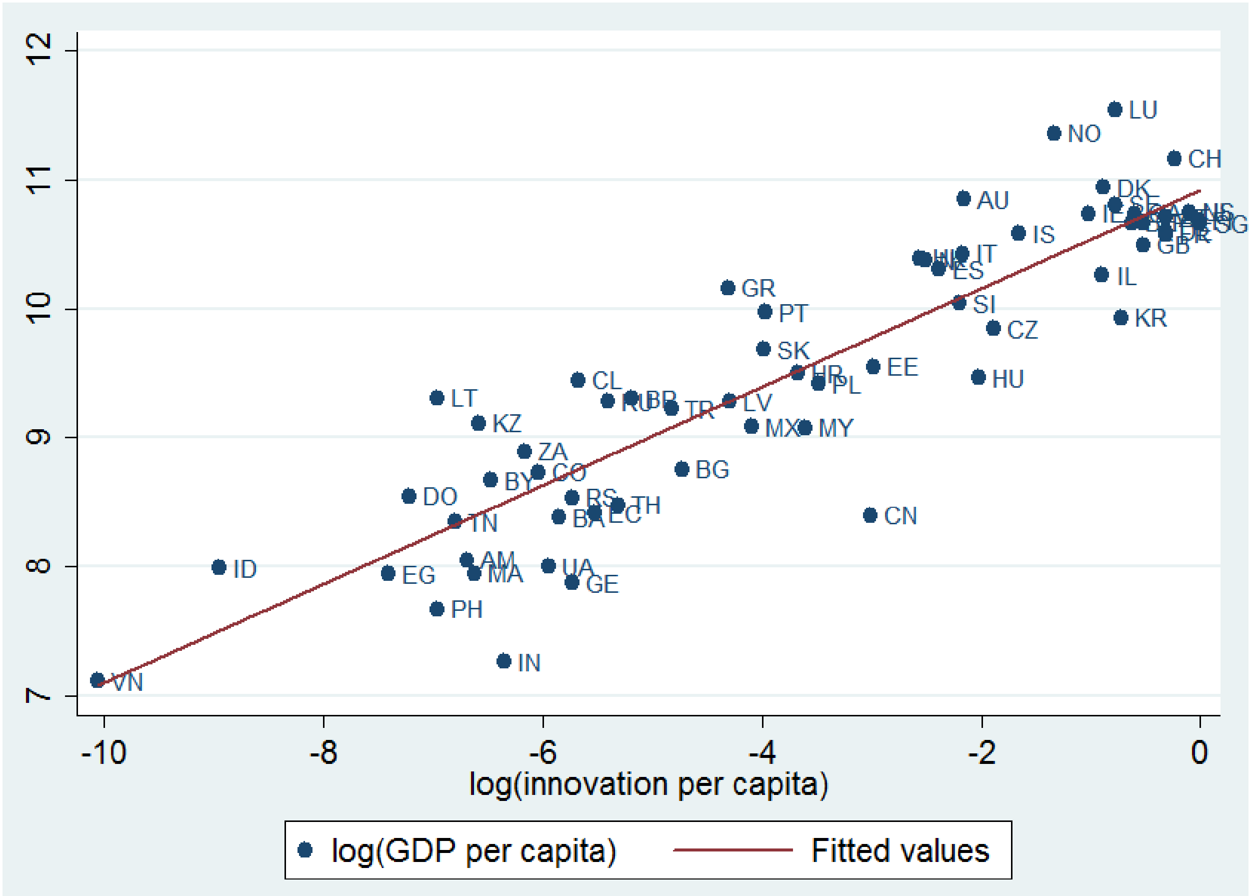
Innovation per capita index

Country ranking of innovation per capita

Rank	Country	Rank	Country	Rank	Country
1	Singapore	22	Hungary	43	Russia
2	Finland	23	Australia	44	Ecuador
3	United States	24	Italy	45	Chile
4	Netherlands	25	Slovenia	46	Georgia
5	Switzerland	26	Spain	47	Serbia
6	France	27	New Zealand	48	Bosnia and H.
7	Germany	28	Hong Kong	49	Ukraine
8	Austria	29	Estonia	50	Colombia
9	Japan	30	China	51	South Africa
10	Great Britain	31	Poland	52	India
11	Canada	32	Malaysia	53	Belarus
12	Belgium	33	Croatia	54	Kazakhstan
13	Korea	34	Portugal	55	Morocco
14	Sweden	35	Slovakia	56	Armenia

15	Luxembourg	36	Mexico	57	Tunisia
16	Denmark	37	Latvia	58	Lithuania
17	Israel	38	Greece	59	Philippines
18	Ireland	39	Bulgaria	60	Dominican R.
19	Norway	40	Turkey	61	Egypt
20	Iceland	41	Brazil	62	Indonesia
21	Czech R.	42	Thailand	63	Vietnam





12
11
10
9
8
7
-10
-8
-6
-4
-2
0
log(innovation per capita)
log(GDP per capita)
Fitted values
LU
NO
CH
DK
AU
IS
IT
ES
IL
GR
PT
SI
CZ
KR
SK
EE
HU
PL
CL
LT
TR
LV
KZ
MX
MY
ZA
BG
CO
BY
DO
RS
TH
TN
CN
ID
AM
EG
MA
UA
GE
PH
IN
VN
GB