



**UNITED NATIONS  
UNIVERSITY**

**UNU-MERIT**

**UNU-MERIT Working Paper Series**

**#2014-060**

**The size of patent categories: USPTO 1976-2006**  
François Lafond

**Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)**  
email: [info@merit.unu.edu](mailto:info@merit.unu.edu) | website: <http://www.merit.unu.edu>

**Maastricht Graduate School of Governance (MGSOG)**  
email: [info-governance@maastrichtuniversity.nl](mailto:info-governance@maastrichtuniversity.nl) | website: <http://mgsog.merit.unu.edu>

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands  
Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499

**UNU-MERIT Working Papers**

**ISSN 1871-9872**

**Maastricht Economic and social Research Institute on Innovation and Technology,  
UNU-MERIT**

**Maastricht Graduate School of Governance  
MGSOG**

*UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT and MGSOG to stimulate discussion on the issues raised.*

# The size of patent categories: USPTO 1976-2006 \*

François Lafond<sup>1,2,3</sup>

<sup>1</sup>Institute for New Economic Thinking at the Oxford Martin School

<sup>2</sup>London Institute for Mathematical Sciences

<sup>3</sup>UNU-MERIT

September 13, 2014

## Abstract

Categorization is an important phenomenon in science and society, and classification systems reflect the mesoscale organization of knowledge. The Yule-Simon-Naranan model, which assumes exponential growth of the number of categories and exponential growth of individual categories predicts a power law (Pareto) size distribution, and a power law size-rank relation (Zipf's law). However, the size distribution of patent subclasses departs from a pure power law, and is shown to be closer to a shifted power law. At a higher aggregation level (patent classes), the rank-size relation deviates even more from a pure power law, and is shown to be closer to a generalized beta curve. These patterns can be explained by assuming a shifted exponential growth of individual categories to obtain a shifted power law size distribution (for subclasses), and by assuming an asymmetric logistic growth of the number of categories to obtain a generalized beta size-rank relationship (for classes). This may suggest a shift towards incremental more than radical innovation.

---

\*I would like to thank Robin Cowan, Giorgio Triulzi, and audiences in Santa-Fe and Oxford. Financial support from METEOR is gratefully acknowledged. All errors are mine. Contact: francois.lafond@oxfordmartin.ox.ac.uk; Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Walton Well Road, Oxford OX2 6ED, U.K.; +44 (0)1865 288895.

JEL: O30; Keywords: incremental and radical innovation, Pareto, Zipf, Beta distribution, Logistic growth, rank - size distribution.

## 1 Introduction

Categorization is at the basis of reasoning. Theorizing about scientific and technological systems is no exception, and always relies on the grouping of several items into “categories”. For instance, the concepts of paradigms, research fields, school of thought, epistemic communities, etc. are all based on the idea that an underlying grouping can be meaningfully established. Elements in these categories (“bio-technologies”, “economics”, “subclass N234”, “keynesians”) are then taken to behave in the same way. At the very least, the analyst can argue that elements within a category have a degree of homogeneity which is much higher than elements taken from different categories. Since analyzing a number of categories is simpler than analyzing every single element, categorization reduces the dimension of the problem.

Categorization, therefore, is at the heart of thought processes. This implies that categories are not simply useful to describe reality, they are the main tool to *construct* it. Categories, when they are created as nouns, can have a predicate and become a subject. They enter discourses with their own identity, and shape our understanding of reality. Classification systems are essential tools in the creation of routinized habits of thoughts. Hence, when a classification is put to use, one may argue that it creates a feedback on the system it describes. Classification systems are institutions which often legitimate the items that they classify. This affects the future evolution of the items, and their relation (boundaries) with other items. Along this line of argument, the process of categorization is performative. The evolution of the technological classification system therefore provides data on how society understands its technological artefacts and legitimizes them through the process of categorization.

In this paper, I propose an attempt at clarifying some of the key processes underlying the evolution of technological and scientific classification systems by studying in detail one of the most important, relatively well defined quantity: the size distribution of categories (or the size-rank relationship, which by construction is less noisy). I study the US patents granted by the USPTO between 1976 and 2006, partitioned at the level of more than 400 classes and 100,000 subclasses.

The size distribution of patent *subclasses* is well fitted by a shifted power law, in agreement with a slightly modified version of the Yule (1925)-Simon (1955)-Naranan (1970) models. However, at the level of *classes*, the size distribution is less skewed. The small sample (428 categories) suggests to study the size-rank relationship instead of the size distribution. The size-rank relationship, at the level of classes, is not a power law (Zipf's law). An exponential relation was recently proposed (Carnabuci 2013) for this data. Here I find that a generalized beta, suggested for size-rank relationships by Martínez-Mekler et al. (2009), fits the data better. I give a simple and original explanation for this fact, answering partially an open problem stated in Egghe (2012). The reason for the departure from Zipf's law is that the number of categories tends to grow faster in the beginning than in the end, as compared to the growth of individual categories. The model of Naranan (1970), which is a simplified version of Yule's (1925) and Simon's (1955) models, derives Zipf's law by assuming an exponential growth of both the number of categories and the number of items per category. To obtain a generalized beta (of the first kind) for the size-rank relationship, I find that one should instead assume an asymmetric S-shaped curve for the number of categories.

The paper is organized as follows. Section 2 gives a background discussion on category systems and technological change, and reviews existing literature. Section 3 describes the data and the methodology. Section 4 presents empirical results. Section 5 proposes theoretical models consistent with the observed empirical laws. Section 6 discusses the results. The last section concludes.

## 2 Literature review

The purpose of this section is to provide a general discussion of what categories mean, why it is important to study them, and how this general theoretical background applies to the case of technological categories. A review of the literature on the evolution of technological domains using patent categories follows.

### 2.1 Theoretical background

The philosophy of category systems<sup>1</sup> has traditionally distinguished between Aristotle realism (categories of things do exist) and Kant conceptualism (what really

---

<sup>1</sup>see the Stanford Encyclopedia of Philosophy for *Categories* and *Natural kinds*, from where the quotes are taken (<http://plato.stanford.edu>, accessed 28/08/2014).

exists are the categories of understanding, based on experience). Husserl proposed an encompassing view where the two systems, categories of meaning and ontological categories, co-exist and are related. Foucault (1966) insisted on the idea that words, the categories making up discourses, are not descriptive tools but genuinely *construct* the world. Latour (2005) concluded that the social scientist should not overimpose her own categories over the actors she analyzes. Instead, the analyst should follow the actors, and see how they create categories themselves.

A more realist view is that of *natural kinds*. A natural kind is defined as one that “corresponds to a grouping or ordering that does not depend on humans”. Chemistry is said to provide the least controversial example of natural kinds (the periodic elements table), whereas biological species classification are less easily taken for granted. From a metaphysical point of view, one is interested in the essence of natural kinds, that is, “the property or set of properties whose possession is a necessary and sufficient condition for a particular’s being a member of a kind”. The existence and relevance of these essences is the point of view of “essentialists”. The view according to which there are “genuinely *natural* ways of classifying things” is called *naturalism*.

On the other hand, constructivists do not believe that classifications reflect the “real” world. In its weak version, constructivism does not deny the existence of natural kinds, but doubts that we can actually see them. Strong constructivism however rejects the mere existence of natural kinds. “Ontological relativism” can be defined as “the view that all entities, processes, relations and theoretical posits are relative to a certain conceptual scheme”. More generally, constructivists argue that categories are created, constructed by the observer. In some cases, it is even argued that the objects – not only the categories as a concept, but the actual objects – are constructed.

Whatever the nature of categories – real or constructed – they are the key building blocks of discourses, including scientific discourses. More generally, a criterion for the existence of a category is that it is a level of aggregation at which a given law holds. According to the “cluster kind realists”, “a natural kind is any (...) family of co-occurring properties that may be employed in inductive inference for the purpose of scientific explanation”. Quine argued that “it is the similarity or sameness of kinds between instances that permits an induction”. Putnam’s (1975) theory of semantic externalism holds that the *meaning* of what people say is not in their head, but in the head of experts - the linguistic community- who collectively know what things are.

In recent years, there has been an expanding literature in sociology about classification, notably some historical studies of controversial classification systems, such as the classification of diseases (Bowker & Star 2000). Some analytical insights from this literature are worth mentioning (Shepherd 2010), and all have to do with the fact that classification exerts an effect on the users of the classification scheme, or even on the elements being classified. First, there is evidence that the degree of institutionalization of the classification system influences the perception of hybrid members (elements with multiple categories): once a classification system is well established, there is a penalty to hybridity. Second, classification essentially consists in drawing boundaries. Boundaries should be thought of as interfaces rather than borders (Bowker & Star 2000). How communication takes place within boundaries is different than how it takes place between; hence classification institutionalizes possible interactions. Third, the extent to which the users of classification systems can shape the classification changes the way in which users interpret and use the classification system. Finally, actors exercise power to obtain classification systems to their advantage.

## 2.2 The classification of technologies

### 2.2.1 Technologies at the mesoscale

For Arthur (2009), a technology is something that relies on the mastering of natural phenomena to produce a useful artefact. In his words, technology is “a collection of phenomena captured and put to use”. Acknowledging that technology builds out of itself, we see that a technology is made of technologies, thus a technology is a “complex of interacting phenomena”. Here is the central point of the discussion: since there are families of phenomena (chemical ones, electrical ones, quantum ones), there are families of technologies based on these phenomena. Arthur goes a step further by arguing that “each grouping forms a language within which particular technologies – particular devices and methods – are put together as expressions within that language”. Hence, technologies form clusters, which he calls *domains*, because they are based on the same phenomena, or because of some other shared characteristics or purposes. Individual technologies and domains, though both hierarchical constructs (i.e. having their sub-technology or sub-domains), are distinct. To be sure,

A technology (individual, that is) does a job; it achieves a purpose -

– often a very particular purpose. A domain (technology-plural) does no job; it merely exists as a toolbox of useful components to be drawn from, a set of practices to be used. A technology defines a product, or a process. A domain defines no product; it forms a constellation of technologies –a mutually supporting set – and when these are represented by the firms that produce them, it defines an industry. A technology is invented; it is put together by someone. A domain (...) is not invented; it emerges piece by piece from its individual parts. A technology – an individual computer, say – gives a certain potency to whoever possesses it. A domain –the digital technologies – gives potential to a whole economy that can in time become transmuted into future wealth and political power.

However, it is not always obvious to attribute a unique category to a given innovation. From the inventor perspective, Arthur (2009) describes the process of choosing a category (a “palette of components”) for a new device as *domaining*. Sometimes, this is automatic, sometimes more difficult. Often, if the technology is large enough, it will belong to several domains.

### **2.2.2 Rationalizing practice: classification as legitimation**

If technologies can be meaningfully categorized, the information feedback provided by the category system will in turn influence further technological development. From a pragmatic stance, things are real if they are real in their consequences, so if firms use category systems to search for technologies and build their own, categories are ontological. But how exactly do these categories map with the human perceptions, or construction, of them? Nelson (2006) describes technological evolution as the co-evolution of a body of practice and a body of understanding. He describes the role of the body of understanding as one of “rationalizing” the practice.

(...) what makes the evolution of human practice, and especially technology, different from the evolution of animal behavior as studied by ethologists is exactly that extant human practice is generally supported by a rather elaborate body of reasons, or rationalizations.(...) To the extent that technology is seen as not simply a body of practice, but also a body of understanding, the nature of the evaluation and selection processes becomes more complicated. While the criteria for selection



on the former aspect may well ‘fit’ with user need, the criteria for the later may appear to be the ‘ability to explain observed relevant facts and enable problems to be solved and progress made’. The selection processes and those who control them, as well as the criteria, may well be different. For practice, the process is ultimately under the control of users, or their agents; for understanding, the control rests with the community of technologists”.

In this paper, I consider categorization as a process of codification of an understanding concerning the technological system, and I argue that the dynamics of patent classes and subclasses constitute a window on the “community of technologists”. Using mathematical and statistical modelling, it is possible to uncover the most fundamental principles at play in the growth of classification systems. Perhaps the most important of these principles is *creation*, i.e. the fact that *new* categories are created over time.

### **2.2.3 The emergence of new technology categories**

Technological evolution is reflected in the evolution of the classification system and this suggests to study the dynamics of categories creation (Strumsky et al. 2012). Classification is a particularly important topic in innovation studies because by definition, innovation challenges existing schemas (Hicks 2011). Economists of innovation have traditionally made a distinction between incremental and radical innovations. In the context of classification, the former can be interpreted as being a new technology which fits perfectly in the existing classification scheme. This sort of innovation falls into what Arthur (2009) calls “structural deepening”, the addition of subsystems and subassemblies to an existing technology so as to improve basic performance, adapt to wider tasks, improve reliability and safety, or simply react to changed circumstances, for instance on the demand side. On the other hand, a radical innovation can be defined as a new technology which creates a new category (perhaps by creating a need to reframe the classification system, or simply by being *sui generis*, the first innovation of its own kind) – this would be called redomaining or revolution in Arthur’s terms.

## 2.3 Growth and distribution of technological domains

The growth of technological domains has been deeply scrutinized in the economics of technical change and development (Schumpeter 1934, Dosi 1982, Pasinetti 1983, Pavitt 1984, Freeman & Soete 1997, Saviotti 1996, Malerba 2002). A recurring theme in this literature is the high heterogeneity among sectors. Besides heterogeneity due to the very nature of the underlying knowledge base, sectors are also at different stages of their life cycles (Vernon 1966, Klepper 1997).

Life-cycle theories of technological change suggest non linearities in the evolution of individual technological domains. Andersen (1999) fits logistic growth models, using USPTO data for the period 1890-1990, and aggregated at the level of 56 technological groups (collections of patent classes). Andersen (1999) does not fit logistic growth for the whole period (1890-1990) but instead for subperiods, arguing that a logistic growth episode constitutes one cycle. Hence, in this theory, an individual category follows repeated S-shaped patterns. In this paper, I do not look for detailed life-cycle patterns in the data. Instead, I seek for the most parsimonious model that gives reasonable aggregated results. Hence, the results are less precise, and less informative of individual sectors' histories. On the other hand, the models and results of this paper are more powerful in explaining why other, potentially very different datasets, exhibit similar regularities ("universal laws").

Similarly, Carnabuci (2013) detected a departure from pure multiplicative growth (Gibrat's model) for the size of individual categories, which I do not account for in the models of section 5. It turns out that at both the subclass and class levels, assuming exponential growth of individual categories gives reasonable frequency-size or rank-size distributions, if associated with an appropriate growth function for the number of categories.

Finally, using the almost complete record of US patents, Youn et al. (2014) found that the number of subclasses follows the same trend as the number of patents until about 1870, but grew less fast afterwards. Remarkably, however, the number of unique *combinations* of subclasses used in a patent keeps the same trend as the number of patents. The data used in the present study does not contain multiple classifications, but only the main patent class/subclass, and covers only the period 1976-2006, as described in the next section.

## 3 Data and method

### 3.1 The United States Patent Classification System

Falasco (2002) describes three types of rationale behind the United States Patent Classification System (USPCS). The classification by *industry* was the original method when the first USPCS appeared in the beginning of the XIX<sup>th</sup> century. The classification by *structure* (“arrangement of components”) is sometimes useful, for instance composite materials can be classified according to the arrangement of their parts. The classification by *utility* (proximate function) was “adopted in the early XX<sup>th</sup> century as the fundamental principle for classifying prior patent art in the USPCS”. By proximate function, it is meant the fundamental function of the invention, not some example application in a particular device or industry. The classification by utility includes also the classification by effect or product, which “provides collections based on *result*”, for instance measuring, illuminating, etc.

The USPCS attributes to each patent at least one subject matter. A subject matter includes a main class, delineating the main technology, and a subclass, delineating processes, structural features and functional features. All classes and most subclasses have a definition. All subclasses within a class are arranged in a class schedule, “with the most complex and comprehensive subject matter generally at the top of the schedule, and the least complex and comprehensive at the bottom” (USPTO 2012). Subclasses have indentation levels. Primary subclasses are the main type of subclass, and each patent must be assigned to at least one of them. Some subclasses are called “Alpha subclasses”. These were previously unofficial subclasses, created by examiners to ease their work. They are identified by their parent primary subclass, adding one or two letters. Finally, note that these are the patent claims which are classified, and the patent inherits the classification of its claims. The main classification is the one of its main claim.

Patent examiners follow a rather precise algorithm for finding the appropriate subclass (USPTO 2012): starting at the top, scan downward looking only at main line subclasses until one is found that provides for any portion of the claim. Within this one, scan downward subclasses which are at one higher indentation level until one is found that provides with at least a certain portion of the subject matter, and go ahead similarly to the next indentation level. When no subclasses can be found, stop the process and classify at the last subclass selected.<sup>2</sup>

---

<sup>2</sup>It is interesting to note that this is a case of a formal, codified rule that should be followed to

Classification is used mainly for the search of prior art. It is also used to determine which department, that is, which examiners will evaluate the patent. It is also widely used by business analysts.

### 3.2 The NBER patent data (1976-2006)

I use the well-known NBER patent dataset (Hall et al. 2001), updated to 2006.<sup>3</sup> It contains information on 3,210,361 utility patents. 1335 patents have a “NA” class, corresponding either to a “withdrawn” subclass (1328) or to a “D” subclass (7 patents). I removed these 1335 patents, leaving a total of 3,209,056 uniquely allocated to 428 classes and 119833 subclasses (“primary” US class/subclass). Subclasses are detailed at the 6 digits level, and I consider alpha subclasses as distinct. Throughout, I use the current classification system, that is, the category of each patent is its category in 2008 (“CCL” variables in the NBER file, i.e. the variables “cclass” and “nclass”).

## 4 Empirical observations

This section presents empirical results on the size distribution of subclasses, and the size-rank relation for classes. The size of a category is simply defined as the number of patents within it.

### 4.1 The size distribution of subclasses

The size-distribution of patent subclasses is a heavy tail distribution, and it is relatively well behaved (not too noisy). After trying a number of classical candidates (power law, stretched exponential, Weibull, etc...), two distributions, the Waring (a discrete version of the shifted power law) and the lognormal, were found to give a good fit. The parameters of these two distributions were estimated using maximum likelihood. A modified Waring distribution was used, defined so that it is properly

---

produce additional codified knowledge (a classification). It highlights that classification is done by following heuristics, as would be expected from bounded rationality and evolutionary economics theory (Simon 1947, Nelson & Winter 1982). Note that the knowledge of this procedure can in principle be exploited to understand the growth of subclasses as a function of their indentation level and position in the class schedule.

<sup>3</sup>see <https://sites.google.com/site/patentdataproyect/Home>. I used the file pat76\_06\_assg.dta.

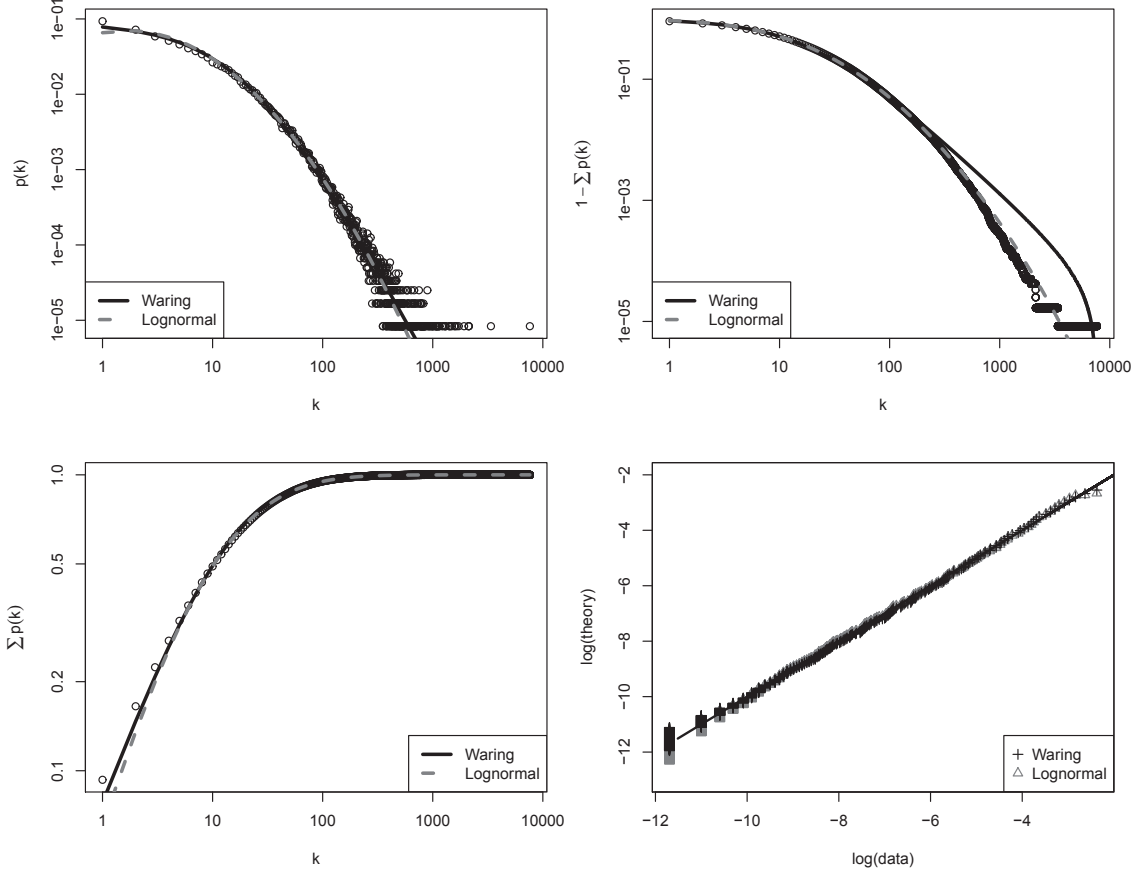


Figure 1: Size distribution of patent subclasses. Top left: Probability mass function. Top right: Complementary cumulative distribution. Bottom left: Cumulative distribution. Bottom right: QQ-plot on double log axis. The fitted curves are a Waring and a log-normal distribution.

normalized over the range  $k = 1 \dots \infty$  (instead of  $0 \dots \infty$ ):<sup>4</sup>

$$p(k) = \frac{d + \gamma - 1}{d} \frac{B(k + d, \gamma)}{B(d, \gamma - 1)}. \quad (1)$$

The parameters and  $R^2$  were computed for the cumulated system (all patents granted up to a given year), at 6 periods of 5 years interval (table 1). The visual fits for the system as of 2006 are provided in figure 1. Although this fit is visually good, the log-transformed data does not pass the Anderson-Darling test for normality.

<sup>4</sup>The normalization constant  $C = \frac{d+\gamma-1}{dB(d,\gamma-1)}$  is found as follow. First impose  $\sum_{k=1}^{\infty} CB(k + d, \gamma) = \sum_{s=0}^{\infty} CB(s + 1 + d, \gamma) = 1$ . From that  $\sum_{s=0}^{\infty} \frac{(1+d)_s}{(1+d+\gamma)_s} = 1/(CB(1 + r, \gamma))$ . The LHS is Gauss Hypergeometric Function and Gauss hypergeometric theorem can be applied to get rid of the summation symbol. Simplifying and solving for  $C$  then gives the result.

	# patents	# categories	$\langle k \rangle$	Waring		Log-normal	
				$\gamma$	$d$	mean	std. dev.
1981	378,040	84,834	4.46	4.44	7.49	1.05	0.90
1986	702,508	99,776	7.04	4.00	11.27	1.40	1.02
1991	1,145,796	109,268	10.49	3.74	15.85	1.71	1.12
1996	1,654,321	114,372	14.46	3.43	18.89	1.96	1.20
2001	2,390,834	117,772	20.30	3.01	19.63	2.20	1.29
2006	3,209,056	119,833	26.78	2.69	18.96	2.36	1.36

Table 1: Parameter estimates for the size distribution at the level of subclasses.

The values of the likelihood (not reported) were slightly but systematically higher for the Waring distribution than for the lognormal.

## 4.2 Rank-size relationship of classes

The number of classes is quite low, which makes the probability density rather noisy.<sup>5</sup> For this reason, I choose to study the rank-size distribution. Carnabuci (2013) noticed that this relationship is not Zipfian. He fitted an exponential function, that is,

$$r(g) = Ce^{-xg},$$

where  $r$  is the rank and  $g$  is the relative size (i.e. the number of patents in the category of rank  $r$ , divided by the total number of patents<sup>6</sup>). However, while this fit is better than that of a power law and gives a good first approximation ( $R^2 \approx 0.96$ ), the exponential functional form clearly underestimates the tail (as can be seen in the plots, but is only poorly reflected by the  $R^2$ ). On the other hand, since the tail is not as fat as Zipf’s law would imply, one needs to find an alternative functional form. Martínez-Mekler et al. (2009) showed that for a number of datasets the size-rank relation could be very well fitted by the following formula, which expresses size  $g$  as a function of rank  $r$ :

$$g(r) = K_1(N + 1 - r)^b r^{-a}, \quad (2)$$

---

<sup>5</sup>Sanditov (2005) fitted a log-normal and gamma distribution to the 1963-1999 database.

<sup>6</sup>Carnabuci (2013) worked with absolute size, but it is easier here to work with relative size.

where  $K_1$  is a normalization constant. To ease later comparison with a close formula derived here, I consider  $N + 1 \rightarrow N$ . Then equation 2 can be rewritten as

$$g(r) = K_2 \left(1 - \frac{r}{N}\right)^b r^{-a}, \quad (3)$$

where  $K_2$  is a normalization constant. This formula is like a power law, except for the factor  $\left(1 - \frac{r}{N}\right)^b$  which decreases as the rank  $r$  approaches the number of categories  $N$ , creating an overall decay faster than normal power laws. Martínez-Mekler et al. (2009) call it a Generalized Beta Distribution (GBD).

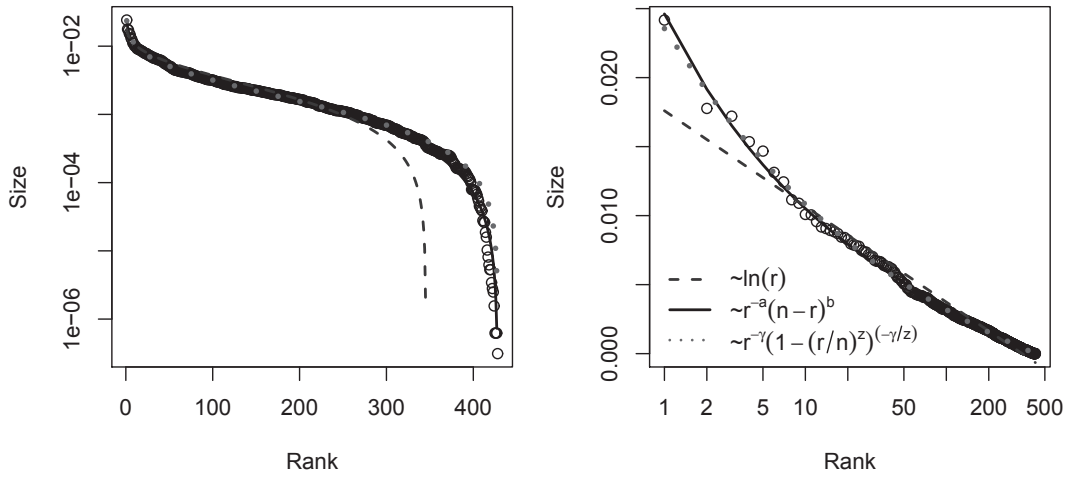


Figure 2: Fit of the size-rank relationship at the level of classes. The same data is displayed on the two panels. On the left panel, only the axis for size is on a log scale, as in Martínez-Mekler et al. (2009), and on the right, only the axis for rank is on a log scale, as in Carnabuci (2013).

	# of patents	# of categories	equation 3		equation 10	
			$a$	$b$	$\gamma$	$z$
1981	378040	423	0.33	1.14	0.24	0.28
1986	702508	425	0.35	1.08	0.27	0.34
1991	1145796	426	0.36	1.07	0.29	0.36
1996	1654321	427	0.35	1.13	0.27	0.33
2001	2390834	427	0.36	1.18	0.27	0.31
2006	3209056	428	0.36	1.36	0.23	0.23

Table 2: Parameter estimates for the size-rank relation at the level of classes.

Using the NBER patent data at the class level, I compared three hypothesis: the exponential relation between rank and size (Carnabuci 2013), the Generalized Beta Distribution (GBD) of Martínez-Mekler et al. (2009) (equation 3), and the slightly different GBD equation 10 derived in section 5 *infra*. Figure 2 shows that the two GBD clearly outperform the exponential. The  $R^2$  values are 0.9597 for the exponential, 0.9974 for the GBD (3), and 0.9978 for the GBD derived from an original model in section 5 (equation 4). These results suggest that these two different GBD perform equally well, at least for this data. Note that following Martínez-Mekler et al. (2009), I take  $N$  directly from the data, that is,  $N$  is equal to the number of classes. Hence only two parameters are estimated by the fitting procedure.<sup>7</sup> Table 2 shows the estimated parameters for the two GBD over time. It is interesting to see that the parameters seem very stable for the first four 5-years periods, and change after 2001, except for the power law exponent ( $a$ ) of Martínez-Mekler et al.'s (2009) equation, which is strikingly constant.<sup>8</sup> This may reflect a great stability in at least part of the process, as well as a slightly better performance of Martínez-Mekler et al.'s (2009) equation over the one presented here.

## 5 Theoretical models

Yule (1925) proposed the first model generating a power-law size distribution. He assumed that the number of categories grows at an exponential rate, and each category grows at an exponential rate.<sup>9</sup> Under these assumptions, the size distribution of categories is a Yule distribution, which has power law tails. Simon (1955) proposed a different version of Yule's model, in which time is not clock time but system time. However, up to a small modification, the two processes are equivalent

---

<sup>7</sup>I used nonlinear least squares with a Gauss-Newton algorithm. Alternative optimization algorithms and starting conditions do not change the results significantly. Note that for Martínez-Mekler et al.'s (2009) equation, using (2) instead of (3) makes OLS estimation possible, by taking logs. In this case, the estimated parameters are slightly different (for 2006,  $a = 0.24$  and  $b = 1.55$ ,  $R^2 = 0.986$ ).

<sup>8</sup>The same heuristic conclusion could be reached by looking at the plots of the four parameters over time for each year, not reported here.

<sup>9</sup>Yule's model is actually more complicated, in that he does not *assume* exponential growth but derives it from first principles, namely that each new species (item) has a fixed probability of generating a new species of a new genera and a fixed probability of generating a new species of an existing genera. Note that Yule's model is stochastic, but the deterministic version here can be seen as a mean-field approximation.



(Simkin & Roychowdhury 2011). Naranan (1970) proposed a deterministic version and obtained similar results.

As we have seen in the previous section, neither the subclasses nor the classes have a pure power law size distribution. How should the Yule-Simon-Naranan's model be modified to produce the distributions observed empirically? To answer this question, I use Egghe's (2012) Generalized Naranan's Framework. Egghe (2012) extended Naranan's work for arbitrary invertible growth functions. He concluded his paper by asking whether one could obtain Martínez-Mekler et al.'s (2009) formula (equation 3) using the generalized Naranan's framework. Here I show first how to derive a *shifted* power law size distribution from the Generalized Naranan's framework (to explain the phenomenology of subclasses), and second how to derive a generalized beta curve very close to that of Martínez-Mekler et al. (2009) for the size-rank function (to explain the phenomenology of classes).

Consider that the number of categories at (continuous) time  $t$  is  $\phi(t)$ , and the number of items in a category of age  $t$  is  $\psi(t)$ . Moreover, assume that  $\phi$  and  $\psi$  are invertible. Egghe (2012) proved the following:

**Theorem 1.** *In the generalized Naranan's framework, the size-rank function, which gives the size of a category as a function of its rank is*

$$G(r) = \psi(t - \phi^{-1}(r)). \quad (4)$$

**Theorem 2.** *In the generalized Naranan's framework, the frequency-size function, which gives the number of categories having a given size, is*

$$P(k) = \frac{\phi'(t - \psi^{-1}(k))}{\psi'(\psi^{-1}(k))}. \quad (5)$$

## 5.1 Shifted power law for subclasses

Let us start by analyzing the subclasses. We seek for  $\phi$  and  $\psi$  such that  $P(k) \propto (k + a)^{-\gamma}$ . It is known from growing network models that a shifted attachment kernel (i.e. where the probability that the next newborn node links to an existing node of degree  $k$  is proportional to  $k + \text{constant}$ ) gives a shifted power law. In other words, if a category accumulates items at a rate proportional to how many items it already has, growth is exponential and the size distribution is a power law. But if a category accumulates items at a rate proportional to how many items it already has *plus a constant*, we should expect a shifted power law. Hence, I assume

$$\frac{d\psi(t)}{dt} = x_1\psi(t) + x_2,$$

which implies

$$\psi(t) = \frac{-x_2 + e^{tx_1}(x_1 + x_2)}{x_1} = -q + c_2 e^{a_2 t}, \quad (6)$$

where the parameters have been rewritten in a more condensed way. My assumption for the growth of the number of categories is the same as Yule-Naranan's:

$$\phi(t) = c_1 e^{a_1 t}. \quad (7)$$

Theorem 5 with the assumptions (6) and (7) gives

$$P(k) = \frac{a_1 c_2^{a_1/a_2} c_1 e^{a_1 t}}{a_2} (k + q)^{-1 - \frac{a_1}{a_2}}.$$

The probability density  $p(k)$  is obtained by dividing by the total number of categories at time  $t$ ,

$$p(k) = P(k)/\phi(t) = \frac{a_1}{a_2} (k + q)^{-1 - \frac{a_1}{a_2}},$$

which is the shifted power law we were seeking to obtain. Moreover, it is a steady-state result. So, to obtain a shift in the power law size distribution in the generalized Naranan's framework, all that needs to be modified is the shift parameter  $q$  in the function  $\psi$  describing the growth of individual categories. It implies that in their early life, categories (here, patent subclasses) benefit from an additional growth factor that ultimately dies out. This factor is equivalent to Dorogovtsev et al.'s (2000) "initial attractiveness", which is necessary when such models are applied to citation networks (de Solla Price 1976). In fact, using a stochastic and discrete model (that is, Simon's (1955) model with a shifted attachment kernel) would allow to derive directly the Waring distribution. However, I preferred to use the generalized Naranan's framework here, because it is much easier to apply to the next case of interest: the generalized beta size-rank function observed for patent classes.

## 5.2 Generalized beta for classes

To obtain a generalized beta size-rank function, one can stick to Naranan's assumption that individual categories grow exponentially, but needs to assume a very flexible function for the number of categories. More precisely, the number of items in a category of age  $t$  is

$$\psi(t) = c_2 e^{a_2 t}, \quad (8)$$

as in Naranan's, and the number of categories is Richard's curve. It is obtained as the solution of (Tsoularis & Wallace 2002)

$$\frac{d\phi(t)}{dt} = a_1\phi(t) \left( 1 - \left( \frac{\phi(t)}{n} \right)^z \right),$$

which is

$$\phi(t) = \frac{\phi_0 n}{(\phi_0^z + e^{-a_1 z t} (n^z - \phi_0^z))^{\frac{1}{z}}}, \quad (9)$$

where the initial condition is  $\phi(0) = \phi_0$ . With  $z = 1$ , it is the classical logistic (Verhulst) equation, which is also S-shaped but symmetric. Verhulst equation is too simple to obtain a good rank-frequency function, because the inflexion point is precisely at the point where size has reached exactly half of the total possible (long term) size. The parameter  $z$  allows greater flexibility as to the position of the inflexion point, making the S-shaped asymmetric.

Theorem 5 under the assumptions (8) and (9) gives the following size-rank relation

$$G(r) = C_1 \left( 1 - \left( \frac{r}{n} \right)^z \right)^{\gamma/z} r^{-\gamma},$$

where  $C_1(t) = c_2 e^{a_2 t} (n\phi_0)^\gamma (n^z - \phi_0^z)^{-\gamma/z}$  and  $\gamma = a_2/a_1$ . Note that this equation will give 0 at the boundary  $r = n$ , which is not desirable but comes from the assumption that growth functions are continuous in this framework. The total size (total number of patents) is  $\sum_r G(r)$  (or the integral if we stick to the continuous approximation). Hence, since the relative size is  $g(r) = G(r)/\sum_r G(r)$ , we have

$$g(r) = C_2 \left( 1 - \left( \frac{r}{n} \right)^z \right)^{\gamma/z} r^{-\gamma}, \quad (10)$$

where  $C_2(t)$  is a normalization constant. Equation 10 is not exactly the same as, but is very similar to equation 3. Therefore, I believe that the mechanism described here provides a good theoretical background for the original equation proposed by Martínez-Mekler et al. (2009) (equation 2). Hence, the results presented here answer to a good extent the open problem stated in Egghe (2012), who asked if equation 2 could be derived from the Generalized Naranan's framework.

It should be noted that Martínez-Mekler et al. (2009) proposed a generative model for their equation, based on a dynamical model of boolean replication. Their model assumes that a vector of 0 and 1 is grown with a probabilistic rule, such that an entry is flipped with some probability, or repeated otherwise. They then show that the frequency with which non-overlapping sextuplets occur is well fitted by

a GBD. Their interpretation is that the parameter  $a$  is associated to permanence (when an entry is replicated), whereas  $b$  is associated to change (when an entry is modified). The model presented here presents several differences with Martínez-Mekler et al.'s (2009). First, it is simpler. Second, the derivation of the GBD is analytical. Third, the formula is not exactly the same. Finally, the interpretation of the parameters is quite different. In the model of this paper, the parameters are determined directly by the parameters of the growth functions  $\phi$  and  $\psi$ .

These results show that, using the generalized Naranan's framework, a non linear modification of the growth of the number of categories is enough to obtain a very good fit of the rank-size relationship in many cases. This does not mean that the number of categories is actually bounded, nor that we need to observe the growth functions assumed here to obtain the GBD. In fact, note that the variable  $t$  appears only in the constant of equation 10. Hence what the model means is that, whatever the growth functions are in clock time (say  $\phi(\tau)$  and  $\psi(\tau)$ ), if we can find a variable  $t(\tau)$  such that  $\phi(t)$  and  $\psi(t)$  follow (8) and (9), then we should obtain the steady-state size-rank relation (10). It is difficult to test directly for the shape of  $\phi$  and  $\psi$  here, because the data covers only about the last half of the US patents, while most classes were created very early, way before 1976 (Strumsky et al. 2012). Nevertheless, we can get some insights into the shape of  $\psi$  by taking logs of equation 8 to obtain the linear regressions:

$$\log \psi(t) = \log c_2 + a_2 t + \epsilon_t, \quad (11)$$

where  $t$  is the year. Figure 3 shows the results for the 415 classes which had positive size in 1976. We can conclude from it that exponential growth, while very rough and unfair to the heterogeneity among classes, is a decent assumption, with most  $R^2$  above 0.9, and the coefficients highly concentrated between 0.05 and 0.15. Although this exercise does not show that growth is exponential, it suggests that the assumption of exponential growth has a fairly good amount of descriptive power. Finally, note that Richards' curve (9) cannot be meaningfully estimated directly, since in 1976 there were already 415 classes out of 428. The next section discusses why such simple assumptions nevertheless allow to make a good prediction of the size-rank relationship.

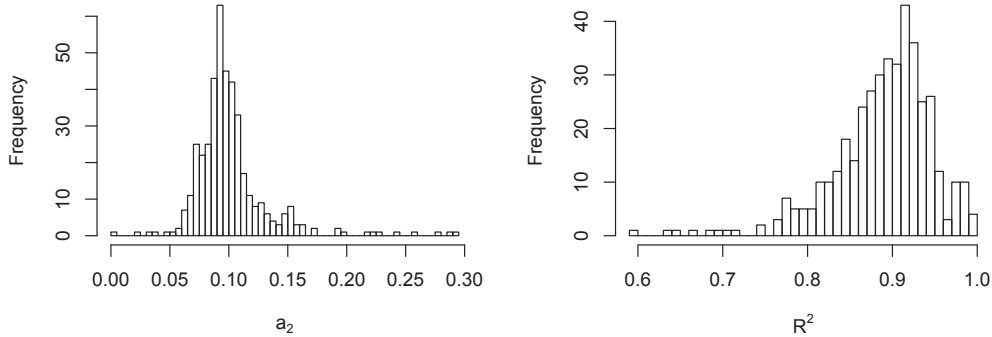


Figure 3: Test of the explanatory power of the exponential growth assumption of individual classes. OLS regressions for equation 11. The mean  $a_2$  is 0.10 and the median 0.095.

## 6 Discussion

The assumptions of the models presented here should not be taken at face value. For instance, it is likely that the functions need not be deterministic for the results on the shape of the distributions to hold true (for the same reason that Yule and Simon’s models give a result similar to Naranan’s, which can be seen as deterministic mean-field continuous approximation of these discrete stochastic models). Moreover, what matters is not the exact shape of the growth function over (real) clock time, but the relationships that the time variable implies between the growth of different items and the number of items.<sup>10</sup> For these two reasons, one can expect that the models presented here can be stated in more general terms. Pinning down the general conditions under which these laws hold is certainly an interesting avenue for future research.

It can be seen in table 1 and 2 that the estimated exponents are not always constant over time. To the extent that this is true, the models used are falsified. The models display “steady-state” results, and therefore, if they were entirely correct, we would not observe time-varying parameters. However, one possibility is that

---

<sup>10</sup>To be more explicit, consider a given model with a given  $\phi(t)$  and  $\psi(t)$ , predicting  $g(r)$  and  $p(k)$ . For the model to be correct, we need not observe  $\phi$  and  $\psi$  as assumed, with  $t$  being our experienced time in months or years. What we need to observe is  $\phi$  and  $\psi$  such that there exists a rescaling of time which makes them behaving as assumed over the rescaled time. It should be emphasized that, even for subclasses, the growth assumptions are not validated if we consider  $t$  as years (e.g. the number of subclasses grows much slower than exponentially)

the systems have not yet reached their steady-state, so that we are observing the convergence dynamics, which may not be accounted for in the models. Alternatively, and perhaps more plausibly, the system is actually in disequilibrium. There are no steady-state distribution of the US patent classes and subclasses sizes, because the way in which categories grow individually and in number is unstable. Note, in particular, that the mean number of patents per category is increasing. This trend can be worrying if one takes a realist view on technological categories. For instance, if one thinks that new categories reflect radical innovation and new patents in existing categories represent incremental innovations, then the data tell us that there is a tendency towards innovation being more and more incremental (less and less radical). echoing the concerns raised by other researchers using different theoretical frameworks (Jones 2009, Gordon 2012).

Finally, the models presented here shed light on the significant differences between the dynamics of classes and subclasses in the US patent classification system. Why is it the case that subclasses are almost power law distributed, whereas classes have a much less fat tail? The models suggest a simple explanation for that: the number of classes stops growing as the number of patents do, whereas the number of subclasses does keep a pace similar (in terms of functional form) to the growth of the number of patents. This is not a surprising result, because while classes have a purely horizontal structure (each class is at the same level of the classification tree), subclasses do not. Indeed, as explained in the literature review, subclasses have different indentation levels, so that as patents accumulate one can create a new subclass in a lower layer of the tree. Hence, if technological change is bounded horizontally (taking a realist view, there is a finite number of categories of natural phenomena, hence a finite number of categories of technologies based on them), this will be reflected in the number of classes but not in the number of subclasses. Moreover, while classes give an indication of the technological domain in which a patent is granted, subclasses are a mandatory classification for *claims*, which are the elements of novelty in a patent document. Novelty by definition cannot be novel for too long, hence as new patents arrive it should be expected that new subclasses are created to accommodate genuine novelty. Beside these explanations, we cannot dismiss other factors behind the emergence of the US patent classification structure. It is possible that the decision to create or not a category is based on search efficiency arguments more than on ontological realities regarding the technology being described. For instance, it can be decided by classification decision-makers that the classes should not be changed too often, so as to provide a stability that is

necessary to fix common understanding and expectations. This would explain the low growth of the number of classes. Albeit more speculative due to the lack of data, the non-fully-realist interpretation reminds us that patent classifications do not show directly the mesoscale organization of inventions, but the technologists' understanding of it, codified for specific purposes which are not only ontological but also functional (in particular searchability).

## 7 Conclusion

This paper proposed three results. First, considering the US patents granted between 1976 and 2006, the size distribution of subclasses is well fitted by a shifted power law. Second, the size-rank relation of classes is well fitted by two different forms of a generalized beta distribution. Third, it is possible to derive a generalized beta functional form for rank-size relationships using the generalized Naranan's framework, by assuming an asymmetric logistic growth of the number of categories.

Much work is left to be done in this area. The models used make very strong assumptions but it remains to be seen how much conclusions would change if more randomness and heterogeneity were allowed. The role of time also needs to be clarified by mathematical arguments, both to understand how general are the models based on the generalized Naranan's framework, and to allow for disequilibrium dynamics, a common feature of knowledge systems. Moreover, the dataset used here covers (utility) patents granted in the US between 1976 and 2006, which is only about a half of the total number of US patents, whereas the models assume that the complete system is observed (notably the growth of the classification system). By observing these "universal" laws more precisely and in different contexts, future research may be able to understand better their fundamental drivers.

## References

- Andersen, B. (1999), 'The hunt for S-shaped growth paths in technological innovation: a patent study', *Journal of Evolutionary Economics* **9**(4), 487–526.
- Arthur, W. (2009), *The Nature of Technology: What It Is and How It Evolves*, Free Press.

- Bowker, G. & Star, S. (2000), *Sorting things out: classification and its consequences*, The MIT Press.
- Carnabuci, G. (2013), ‘The distribution of technological progress’, *Empirical Economics* **44**(3), 1143–1154.
- de Solla Price, D. (1976), ‘A general theory of bibliometric and other cumulative advantage processes’, *Journal of the American Society for Information Science* **27**(5), 292–306.
- Dorogovtsev, S., Mendes, J. & Samukhin, A. (2000), ‘Structure of growing networks with preferential linking’, *Physical Review Letters* **85**(21), 4633–4636.
- Dosi, G. (1982), ‘Technological paradigms and technological trajectories:: A suggested interpretation of the determinants and directions of technical change’, *Research Policy* **11**(3), 147–162.
- Egghe, L. (2012), ‘Study of rank-and size-frequency functions and their relations in a generalized naranan framework’, *Mathematical and Computer Modelling* **55**(7), 1898–1903.
- Falasco, L. (2002), ‘Bases of the United States Patent Classification’, *World Patent Information* **24**(1), 31–33.
- Foucault, M. (1966), *Les mots et les choses*, Vol. 42, Gallimard Paris.
- Freeman, C. & Soete, L. (1997), *The economics of industrial innovation*, third edn, MIT Press, Cambridge, MA.
- Gordon, R. J. (2012), ‘Is US economic growth over? faltering innovation confronts the six headwinds’, *NBER Working Paper* **18315**.
- Hall, B., Jaffe, A. & Trajtenberg, M. (2001), ‘The NBER patent citation data file: Lessons, insights and methodological tools’.
- Hicks, D. (2011), ‘Structural change and industrial classification’, *Structural Change and Economic Dynamics* **22**(2), 93–105.
- Jones, B. (2009), ‘The burden of knowledge and the death of the renaissance man: Is innovation getting harder?’, *Review of Economic Studies* **76**(1), 283–317.



- Klepper, S. (1997), ‘Industry life cycles’, *Industrial and Corporate Change* **6**(1), 145–182.
- Latour, B. (2005), *Reassembling the social: An introduction to actor-network-theory*, Oxford University Press, USA.
- Malerba, F. (2002), ‘Sectoral systems of innovation and production’, *Research Policy* **31**(2), 247–264.
- Martínez-Mekler, G., Martínez, R. A., del Río, M. B., Mansilla, R., Miramontes, P. & Cocho, G. (2009), ‘Universality of rank-ordering distributions in the arts and sciences’, *PLoS One* **4**(3), e4791.
- Naranan, S. (1970), ‘Bradford’s law of bibliography of science: an interpretation’, *Nature* **227**, 631–632.
- Nelson, R. (2006), Perspectives on technological evolution, in K. Dopfer, ed., ‘The evolutionary foundations of economics’, Cambridge University Press, pp. 461–471.
- Nelson, R. R. & Winter, S. G. (1982), *An evolutionary theory of economic change*, Belknap press.
- Pasinetti, L. L. (1983), *Structural change and economic growth: a theoretical essay on the dynamics of the wealth of nations*, CUP Archive.
- Pavitt, K. (1984), ‘Sectoral patterns of technical change: towards a taxonomy and a theory’, *Research Policy* **13**(6), 343–373.
- Putnam, H. (1975), *Mind, language, and reality*, Cambridge University Press.
- Sanditov, B. (2005), ‘Patent citations, the value of innovations and path-dependency’, *KITeS Working Papers* (177).
- Saviotti, P. (1996), *Technological Evolution, Variety, and the Economy*, Edward Elgar Pub.
- Schumpeter, J. A. (1934), *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*, Harvard University Press.
- Shepherd, H. (2010), ‘Classification, cognition and context: The case of the World Bank’, *Poetics* **38**(2), 134–150.

- Simkin, M. V. & Roychowdhury, V. P. (2011), ‘Re-inventing willis’, *Physics Reports* **502**(1), 1–35.
- Simon, H. (1955), ‘On a class of skew distribution functions’, *Biometrika* **42**(3/4), 425–440.
- Simon, H. A. (1947), *Administrative Behavior: A Study of Decision-making Processes in Administrative Organization.*, Macmillan.
- Strumsky, D., Lobo, J. & van der Leeuw, S. (2012), ‘Using patent technology codes to study technological change’, *Economics of Innovation and New Technology* **21**(3), 267–286.
- Tsoularis, A. & Wallace, J. (2002), ‘Analysis of logistic growth models’, *Mathematical Biosciences* **179**(1), 21–55.
- USPTO (2012), Overview of the patent classification system, Technical report.
- Vernon, R. (1966), ‘International investment and international trade in the product cycle’, *The Quarterly Journal of Economics* **80**(2), 190–207.
- Youn, H., Bettencourt, L., Strumsky, D. & Lobo, J. (2014), ‘Invention as a combinatorial process: evidence from us patents’, *arXiv preprint 1406.2938* .
- Yule, G. (1925), ‘A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S.’, *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* **213**, 21–87.

## The UNU-MERIT WORKING Paper Series

- 2014-01 *The medium-term effect of R&D on firm growth* by Marco Capasso, Tania Treibich and Bart Verspagen
- 2014-02 *Diverse and uneven pathways towards transition to low carbon development: The case of diffusion of solar PV technology in China* Michiko Iizuka
- 2014-03 *User innovators and their influence on innovation activities of firms in Finland* by Jari Kuusisto, Mervi Niemi and Fred Gault
- 2014-04 *Migration, remittances and household welfare in Ethiopia* by Lisa Andersson
- 2014-05 *Path-breaking directions of nanotechnology-based chemotherapy and molecular cancer therapy* by Mario Coccia and Lili Wang
- 2014-06 *Poor trends - The pace of poverty reduction after the Millennium Development Agenda* Richard Bluhm, Denis de Crombrughe, Adam Szirmai
- 2014-07 *Firms' adoption of international standards: Evidence from the Ethiopian floriculture sector* by Mulu Gebreeyesu
- 2014-08 *School choice, segregation, and forced school closure* by Cheng Boon Ong and Kristof De Witte
- 2014-09 *Gender difference in support for democracy in Sub-Saharan Africa: Do social institutions matter?* by Maty Konte
- 2014-10 *Why are women less democratic than men? Evidence from Sub-Saharan African countries* by Cecilia García-Peñalosa and Maty Konte
- 2014-11 *Tipping points? Ethnic composition change in Dutch big city neighbourhoods* by Cheng Boon Ong
- 2014-12 *Technology life cycle and specialization patterns of latecomer countries. The case of the semiconductor industry* by Giorgio Triulzi
- 2014-13 *Patents as quality signals? The implications for financing constraints on R&D* by Dirk Czarnitzki, Bronwyn H. Hall and Hanna Hottenrott
- 2014-14 *Assessment of effectiveness of Chinese aid in competence building and financing development in Sudan* by Samia Satti Osman Mohamed Nour
- 2014-15 *Education, training and skill development policies in Arab Gulf countries: Macro-micro overview* by Samia Satti Osman Mohamed Nour
- 2014-16 *Structure of labour market and unemployment in Sudan* by Samia Satti Osman Mohamed Nour
- 2014-17 *Overview of knowledge transfer in MENA countries - The case of Egypt* by Samia Satti Osman Mohamed Nour
- 2014-18 *The impact of ICT in public and private universities in Sudan* by Samia Satti Osman Mohamed Nour
- 2014-19 *End-user collaboration for process innovation in services: The role of internal resources* by Mona Ashok, Rajneesh Narula and Andrea Martinez-Noya
- 2014-20 *Public investment and regional politics: The case of Turkey* by Mehmet Guney Celbis, Denis de Crombrughe and Joan Muysken
- 2014-21 *Infrastructure and the international export performance of Turkish regions* by Mehmet Guney Celbis, Peter Nijkamp and Jacques Poot
- 2014-22 *Discovering and explaining work-family strategies of parents in Luxembourg* by Nevena Zhelyazkova
- 2014-23 *Parental leave take up and return to work of mothers in Luxembourg: An application of the model of nested dichotomies* by Nevena Zhelyazkova

- 2014-24 *Millennium Development Goals: Tool or token of global social governance?* by Mueid Al Raee, Elvis Amoateng, Elvis Koroku Avenyo, Youssef Beshay, Mira Bierbaum, Charlotte Keijser and Rashmi Sinha
- 2014-25 *One Europe or several? Causes and consequences of the European stagnation* by Jan Fagerberg and Bart Verspagen
- 2014-26 *The harmony of programs package: Quasi-experimental evidence on deworming and canteen interventions in rural Senegal* by Théophile Azomahou, Fatoumata Diallo and Wladimir Raymond
- 2014-27 *Country Terms of Trade 1960-2012: Trends, unit roots, over-differencing, endogeneity, time dummies, and heterogeneity* by Thomas Ziesemer
- 2014-28 *The structure and comparative advantages of China's scientific research - Quantitative and qualitative perspectives* by Lili Wang
- 2014-29 *Transition to knowledge-based economy in Saudi Arabia* by Samia Satti Osman Mohamed Nour
- 2014-30 *Challenges and opportunities for transition to knowledge-based economy in Arab Gulf countries* by Samia Satti Osman Mohamed Nour
- 2014-31 *Migration of international students and mobilizing skills in the MENA Region* by Samia Satti Osman Mohamed Nour
- 2014-32 *Beyond product innovation; improving innovation policy support for SMEs in traditional industries* by René Wintjes, David Douglas, Jon Fairburn, Hugo Hollanders and Geoffrey Pugh
- 2014-33 *The impact of innovation support programmes on SME innovation in traditional manufacturing industries: an evaluation for seven EU regions* by Dragana Radicic, Geoffrey Pugh, Hugo Hollanders and René Wintjes
- 2014-34 *Beliefs dynamics in communication networks* by Théophile T. Azomahou and Daniel C. Opolot
- 2014-35 *Stability and strategic diffusion in networks* by Théophile T. Azomahou and Daniel C. Opolot
- 2014-36 *Epsilon-stability and the speed of learning in network games* by Théophile T. Azomahou and Daniel C. Opolot
- 2014-37 *Afghan unaccompanied minors in the Netherlands: Far away from home and protected?* by Carla Buil and Melissa Siegel
- 2014-38 *Multinational production and trade in an endogenous growth model with heterogeneous firms* by Hibret B. Maemir and Thomas Ziesemer
- 2014-39 *The political economy of research and innovation in organic photovoltaics (OPV) in different world regions* by Serdar Türkeli and René Kemp
- 2014-40 *Towards the societal system of innovation: The case of metropolitan areas in Europe* by Serdar Türkeli and René Wintjes
- 2014-41 *To return permanently or to return temporarily? Explaining migrants' intentions* by Özge Bilgili and Melissa Siegel
- 2014-42 *Catching up and lagging behind in a balance-of-payments-constrained dual economy* by Alejandro Lavopa
- 2014-43 *An introduction to the economics of rare earths* by Eva Bartekova
- 2014-44 *The unequal effect of India's industrial liberalization on firms' decision to innovate: Do business conditions matter?* By Maria Bas and Caroline Paunov
- 2014-45 *Insurgents in motion: Counterinsurgency and insurgency relocation in Iraq* by Pui-hang Wong

- 2014-46 *Successive leadership changes in the regional jet industry* by Daniel Vertesy
- 2014-47 *Demand, credit and macroeconomic dynamics: A microsimulation model* by Huub Meijers, Önder Nomaler and Bart Verspagen
- 2014-48 *Accessibility analysis as an urban planning tool: Gas station location* by D.A. Escobar , C. Cadena-Gaitán, F.J. Garcia
- 2014-49 *China's economic embrace of Africa - An international comparative perspective* by Tobias Broich and Adam Szirmai
- 2014-50 *Mapping regional social enterprise ecosystems in India: Framework and indicators* by Lina Sonne
- 2014-51 *Does shelter assistance reduce poverty in Afghanistan?* By Craig Loschmann, Christopher R. Parsons and Melissa Siegel
- 2014-52 *How important is innovation? A Bayesian factor-augmented productivity model on panel data* by Georges Bresson, Jean-Michel Etienne and Pierre Mohnen
- 2014-53 *Does too much work hamper innovation? Evidence for diminishing returns of work hours for patent grants* by Mehmet Güney Celbi, Serdar Türkeli
- 2014-54 *Globalization, the rise of biotechnology and catching up in agricultural innovation: The case of Bt technology in India* Michiko Iizuka and Ajay Thutupalli
- 2014-55 *Where are innovation indicators, and their applications, going?* by Fred Gault
- 2014-56 *Productivity in services in Latin America and the Caribbean* by Elena Arias-Ortiz, Gustavo Crespi, Alejandro Rasteletti and Fernando Vargas
- 2014-57 *Optimal public investment, growth, and consumption: Fresh evidence from African countries* by Augustin Kwasi Fosu, Yoseph Yilma Getachew and Thomas H.W. Ziesemer
- 2014-58 *International R&D alliances by firms: Origins and development* by Rajneesh Narula and Andrea Martínez-Noya
- 2014-59 *Appropriability mechanisms, innovation and productivity: Evidence from the UK* by Bronwyn H. Hall and Vania Sena
- 2014-60 *The size of patent categories: USPTO 1976-2006* by François Lafond