Density Inference for Ranking European Research Systems in the Field of Economics

Michel Lubrano[1] Camelia Protopopescu[2]

November 2002

Abstract

This paper shows how the framework of the economics of inequality can be adapted and transformed in order to investigate the publishing habits of european economists, country by country. It underlines the limitations of usual distribution free estimators and promotes a Bayesian approach using a parametric distribution. A Bayesian model selection procedure validates the Weibull distribution as an admissible reduction of the generalised Beta-II for all the eight samples of the empirical application. The application is based on data extracted from the CD-ROM of the Journal of Economic Literature and concerns seven European countries: Belgium, France, Germany, Italy, Netherlands, Spain and UK to which California is added as a US point of comparison. Analytical upper partial moment curves and modified generalised Lorenz curves are derived for the Weibull model and serve to detect a group of three dominant european countries: The Netherlands, the UK and Belgium.

JEL Classification: C11, C12, D63, I29
Keywords: Bayesian inference, Hypothesis testing, Stochastic dominance, ranking economic departments

[1]GREQAM-CNRS, Centre de la Vieille Charité, 2 rue de la Charité, F-13002 Marseille, France; email: lubrano@ehess.cnrs-mrs.fr. Corresponding author.
[2]GREQAM, Marseille.

Support of the European Economic Association through contract "Ranking Economic Departments in Europe" is gratefully acknowledged. While remaining responsible for any error, the authors wish to thank Luc Bauwens, Charles Bos, Russell Davidson, Jacques Drèze, Jean-Pierre Florens, Nicolas Gravel, Alan Kirman, an anonymous referee and the editor for useful remarks and suggestions.

1 Introduction

There is a vast literature on how to rank the research output of economic departments, see e.g. Dusansky and Vernon (1998) for the USA and the recent paper of Kalaitzidakis, Mamuneas and Stengos (1999) for Europe. Some authors have focused on individual European countries: Combes and Linnemer (2001) on France, Bauwens (1999) on Belgium, van Damme (1996) on the Netherlands. Coupé (2000) was one of the first to face the challenge of obtaining a world ranking. Other authors have focussed on particular domains such as econometrics as did Cribari-Neto, Jensen and Novo (1999). But none of these authors explicitely acknowledged the fact that scientific production is a random variable and that consequently uncertainty has to be taken into account when producing a ranking.

Another branch of the literature investigates "patterns of research output and author concentration" (Cox and Chung 1991). It started with the historical work of Lotka (1926) who plotted on logarithmic paper the number of authors against the number of their respective contributions in the field of chemistry and physics. He found a straight line with a negative slope of 2. This corresponds to the logarithmic regression

$$\log a_n = \log a_1 - \alpha \log n \tag{1}$$

where a_n is the number of authors having made n contributions and a_1 the number of authors having made one contribution. When $\alpha = 2$, this empirical law implies that approximately 60% of the authors have made only one contribution. Several theoretical justifications have been given for Lotka's law, (see the references given in Cox and Chung 1991), but to our knowledge, it was never noticed that this is clearly connected to the Pareto distribution. Modern bibliometric research no longer proceeds by simply counting the number of publications of an author, but tries to measure his performance by taking into account the quality of the journal in which his paper is published and the number of co-authors involved. An index is thus constructed which can be considered as a proxy for an underlying continuous random variable. Thus, the change from n (a discrete variable) to y (a continuous variable) means that Lotka's law implies $f(y) = f(1)\, y^{-\alpha}$. Imposing the constraint that $f(y)$ integrates to unity yields that:

$$f(1) \int_1^\infty y^{-\alpha} dy = 1 \Rightarrow f(1) \left. \frac{y^{-\alpha+1}}{1-\alpha} \right|_1^\infty = 1 \Rightarrow f(1) = \alpha - 1. \tag{2}$$

In the general case, observations do not start from 1 but from y_0 so that the expression of the density is $f(y) = (\alpha - 1) y_0^{\alpha-1} y^{-\alpha}$ with $y > y_0$ and $\alpha > 1$. This is the Pareto density of the first kind[1]. The aim of this paper is to analyse the empirical distribution of individual scores observed on a country basis and to propose statistical tools for comparing countries. The approach taken here can be motivated by considering the paradigm of a PhD student applying for a grant under a bilateral exchange program and having to decide in which country he should write his PhD dissertation. Each country will have a distribution of research ouptut and the student has to order these distributions. This is a decision problem under uncertainty which can be solved by ranking distributions using the tool of stochastic dominance. We have chosen in this paper to estimate parametric distributions. The Pareto distribution has some very restrictive features as will be underlined in what follows. So a more general distribution is needed and the generalised Beta-II which is often used to model the income distribution is a good starting point. Adopting a parametric approach opens the way to Bayesian inference. Bayesian inference allows us to take into account in a simple way the uncertainty which is attached to the parameters of the distributions. We can then compute posterior probabilities of stochastic dominance as simple functions of these parameters.

The paper is organised as follows. In section two, we describe the data and their characteristics. In section three, we introduce the notion of stochastic dominance and its use in a decision problem.

[1] Johnson, Kotz and Balakrishnan (1994) note the Pareto density as $ak^a y^{-(a+1)}$ so that $a + 1 = \alpha$. If y is a discrete variable, the corresponding distribution is the Zypf distribution, often used to model the occurrence of words in a text.

We propose a new way of implementing it which is specific to our particular problem. We emphasize the connection between ranking academic institutions and income inequality analysis. In section four, we specify some useful parametric distributions related to the generalised Beta-II and indicate a Bayesian procedure for model selection. In section five, we derive upper partial moment functions and the Lorenz curve for the Weibull model and use these to rank countries. Section six concludes.

2 The data and their characteristics

This section is devoted to the construction of the index characterising the scientific production of an author for a given period.

2.1 Sources of information

The data can be viewed as transformations of a counting process where the objects counted are papers published over a given period of observation. They are extracted from the JEL CD-ROM over a period of ten years (1991-2000) for 7 european countries (Belgium, France, Germany, Italy, Netherlands, Spain, UK) plus California to have a US point of comparison. We obtained a total of 41 570 articles written by 21 326 different authors. These data are part of a wider project initiated by the European Economic Association. Lubrano, Bauwens, Kirman and Protopopescu (2002) contains the full details.

The second step involves ranking the journals in which these articles have been published. The JEL contains 681 journals. European countries have used essentially 506 journals. We ranked these journals on a scale between 1 for low quality to 10 for top quality. This ranking is obtained from expert opinions. We used the opinions of the panel of experts reported in Combes and Linnemer (2001) and completed it by opinions expressed by Alan Kirman. The obtained ranking was confronted with citation data available from the Journal Citation Reports for 167 journals. There were discrepancies for only 20 journals which were either professional journals at the border of the field or small academic journals. We give the top 68 journals in the appendix. In the field of econometrics, *Econometrica* scores 10, the *Journal of Econometrics* 8, *Econometric Theory* 6, the *Journal of Applied Econometrics* 6 and *Econometric Reviews* 4.

The third step aggregates this information for every author. An article i is credited to an author according to the formula

$$p_i = v_i/\sqrt{n_i} \tag{3}$$

where n_i is the number of coauthors involved and v_i the quality index of the journal where paper i is published. The total score of an author is obtained by summing all his p_i over the period of observation. This formula deserves two comments. First, it does not take into account the number of pages. As reported in Cribari-Neto *et al* (1999), the main journals used by econometricians have average page numbers which varies between 10 and 23. If we introduce the number of pages in (3), we would dramatically alter the implicit journal ranking. Second, this formula is super-additive as it gives each coauthor more than an equal share of credit for an article. Using $\sqrt{n_i}$ instead of n_i favours co-authorship[2].

2.2 Some descriptive statistics

Table 1 summarizes the main characteristics of the data set. The second column gives the total number of authors having published at least one article over the ten year period (this is the size of the sample for each country). This number divided by the total population in million gives in the third column the proportion of active authors in the population. There is a large variation in this proportion, with three outstanding top countries: UK, Netherlands and California. The other columns show other statistical characteristics of the authors by country. Compared to the maximum, the average score is very low and the median even lower. The median score is roughly

[2]Simple aggregation would use n_i instead of $\sqrt{n_i}$. But the different calculations we made for this case and that are not reported in the paper show that this option does not significantly change the results.

Table 1: Descriptive statistics for authors scores

Country	Authors	Aut./Pop.	Mean	Median	Max	$1 - F(10)$
Belgium	806	80.60	6.56	2.31	146	0.15
France	2699	45.76	5.50	2.00	216	0.11
Germany	2506	30.26	4.40	2.00	98	0.10
Italy	1921	32.87	4.19	1.71	102	0.09
Netherlands	1793	111.94	7.35	2.83	187	0.18
Spain	1527	39.15	4.64	1.42	90	0.12
UK	6656	117.55	6.78	2.83	159	0.17
Total	17908	55.05	5.84	2.31	216	0.14
California	3419	100.86	9.58	4.00	190	0.22

equivalent to 2 papers published in low ranked journals in ten years for Europe, but twice that number for California. The last column give the proportion of authors having a publishing score greater than 10.

2.3 Non-parametric density inference

Our aim is to fit a parametric density to the data. We first use a non-parametric kernel smoother[3] to get an idea about the shape of the empirical distributions. Because all the weight of the distribution is located for low values of y whereas the maximum of y can be very large, we have split Figure 1 into two panels to present independently the left tails and the major part of the right tails of the country densities. The smoothing parameter c is equal to 1.06 in the left panel and to 5 in the right panel[4].

3 Stochastic dominance and academic ranking

The empirical investigation described above does not give much of a guess as to how to choose a single measure for ranking academic systems (understood as a collection of authors located in a particular country). A look at the stochastic dominance and the expected utility literature is useful in this connection. The notion of stochastic dominance is quite old as it can be traced back to e.g. Blackwell (1953) or Lehman (1955). However there have been many recent developments.

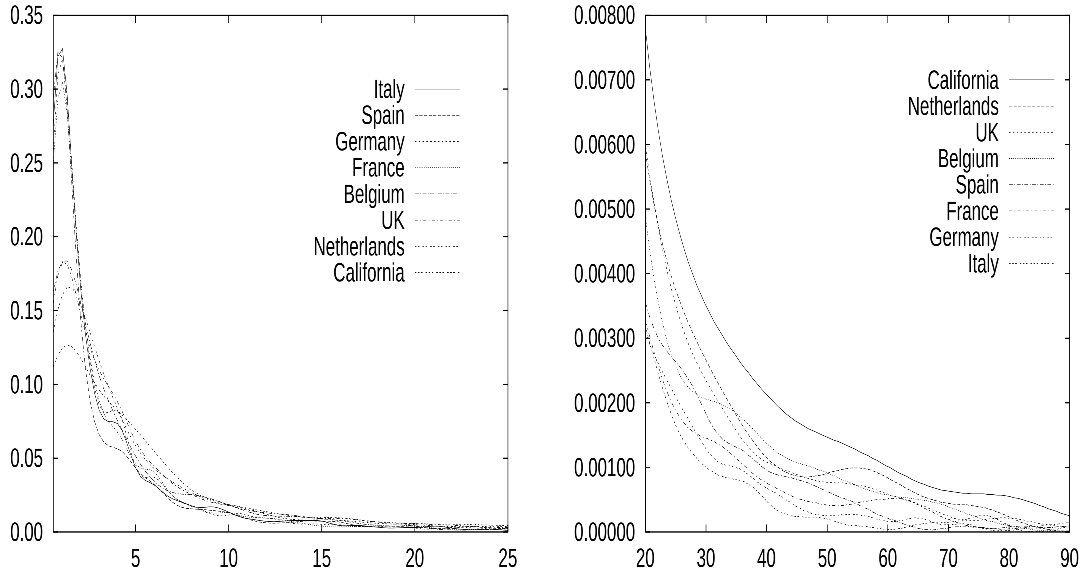
3.1 Maximizing expected utility

Let us return to the case of a PhD student and suppose that he is applying for a bilateral exchange program and has to decide in which country he should write his PhD dissertation. He is looking for two things: a PhD supervisor and a scientific environment (externalities). These two aspects are captured by the random variable Y which measures the performance of academics (their publishing score) in a given country. He is faced with a decision problem under uncertainty. Let us suppose that he has a Von-Neuman-Morgenstern utility function $u(y)$ and that he maximises his expected utility. He has to choose among different distributions $F_i(.)$ of Y and will choose the country i which maximises his expected utility

$$\max_i \int_0^\infty u(y) dF_i(y) \tag{4}$$

[3]A density kernel smoother means $\hat{f}(y) = 1/(nh) \sum K((y_i - y)/h)$ where K is a normal kernel (among many other possibilities), h is the window size determined by $h = c \times \hat{\sigma}_y / n^{1/5}$ and c is a parameter monitoring the degree of smoothing. For instance, with Silverman's rule, $c = 1.06$.

[4]The smoothing parameter is greater for the right panel because it concerns fewer observations. In this right panel the maximum value of y is equal to the minimum of all the different sample maxima.



In each panel, country names are given in the order of the heights of the modes

Figure 1: Density estimates of authors' scores

as the support of Y is $[0, +\infty[$. This class of decision problem under uncertainty was investigated in a series of papers by Quirk and Saposnik (1962), Hadar and Russell (1969), Hanoch and Levy (1969). The main result we can infer from these papers is that if the student's utility function is increasing, he will prefer country i to country j if and only if F_i stochastically dominates F_j at the first order. If the student's utility function is increasing and concave (risk aversion), he will prefer country i to country j whenever F_i stochastically dominates F_j at the second order. A side result of this literature is that the traditional Markowitz (1952) mean and variance criterion is not valid except if the utility function is quadratic or if the distribution functions are normal.

The usual definition of stochastic dominance at the order one (or first degree stochastic dominance) is (see e.g. Hadar and Russell 1969):

Definition 1 *The probability distribution F stochastically dominates the probability distribution G at the order one if and only if*

$$F(x) \leq G(x) \qquad \forall x \in [0, +\infty[. \tag{5}$$

This definition means that the probability of getting x or less is not larger with F than it is with G, whatever the value of x. Of course, when the two cumulative distribution functions intersect, this definition cannot be applied. Second order (or second degree) stochastic dominance is based on the comparison of the surface under the cumulative distribution functions and may remove this indeterminacy. We have:

Definition 2 *The probability distribution F stochastically dominates the probability distribution G at the order two if and only if*

$$\int_0^x [F(t) - G(t)]dt \leq 0 \qquad \forall x \in [0, +\infty[. \tag{6}$$

To understand what is involved mathematically, it is useful to consider the sequence of integrals for the density f

$$F_0(x) = f(x), \quad F_1(x) = \int_0^x F_0(t)dt, \quad F_2(x) = \int_0^x F_1(t)dt, \ldots \tag{7}$$

and the same sequence for the density g. Because cumulative distribution functions are positive and increasing in x, stochastic dominance at the order s, which means

$$F_s(x) \leq G_s(x) \qquad \forall x \in [0, +\infty[$$

implies stochastic dominance at all higher orders. In particular, stochastic dominance at the order two means the ordering $F_2(x) \leq G_2(x)$, $\forall x$ and implies the ordering $F_{2+j}(x) \leq G_{2+j}(x)$, $\forall j \geq 1$, but does not require the ordering $F_1(x) \leq G_1(x)$, $\forall x$.

Because it may be time consuming to check for stochastic dominance when selecting among a large set of distributions as in portfolio selection, a branch of the literature has devoted efforts to finding necessary (but not sufficient) conditions which enable one to eliminate irrelevant alternatives. In that spirit, Bawa (1975), Fishburn (1977) and Jean (1984) introduced *lower partial moments* (LPM) of order s ($s \geq 0$) for a distribution F with reference value z

$$LPM_F^s(z) = \int_0^z (z-t)^s dF(t). \tag{8}$$

The semi-variance corresponds to $s = 2$. Fishburn (1977) uses it as a measure of risk in portfolio selection. For a given z, this measure is asymmetric because it does not treat upper and lower deviations from the mean or from the target symmetrically as the variance does. It concentrates on the left tail of the distribution.

Using integration by parts, it is easy to show by recurrence, the link between the sequence of integrals (7) and the LPM definition (8):

$$F_s(z) = \frac{1}{(s-1)!} LPM_F^{s-1}(z) \qquad s \geq 1. \tag{9}$$

Stochastic dominance at the order s implies the ordering of partial moments starting from order $s-1$. For instance, stochastic dominance at the order two implies the ordering of all partial moments; but the ordering of semi-variances is not a necessary condition for stochastic dominance at the order four. See Jean (1984) for more results on partial moments.

LPM can be transformed into a function of $x \in [0, +\infty[$ as follows:

$$LPM_F^s(x) = \int_0^x (x-t)^s dF(t) \tag{10}$$

and because we have now

$$F_s(x) = \frac{1}{(s-1)!} LPM_F^{s-1}(x) \tag{11}$$

the ordering of LPM functions of order $s-1$ for distributions F and G corresponding to

$$LPM_F^{s-1}(x) \leq LPM_G^{s-1}(x) \qquad \forall x \in [0, +\infty[\tag{12}$$

is strictly equivalent to the condition for stochastic dominance of F over G at the order s.

3.2 Upper partial moment functions

Many of the authors appearing in the JEL CD-ROM are occasional writers. They have either left the academic system or are involved in administrative tasks. These persons should not be taken into account when comparing countries. This means that our student's utility function $u(x)$ has a lower bounded support, $[z, \infty]$. To avoid counting such individuals, we need to introduce z which represents a minimal level of activity which allows us to differentiate an occasional writer and someone who is sufficiently active to qualify as a PhD superviser. A bounded support for the utility function means that the condition for stochastic dominance has to be verified only for x belonging to that bounded support. This is the concept of restricted stochastic dominance as introduced by Atkinson (1987). This notion (under a different name) is already present in Hanoch and Levy (1969).

The particular shape of the lower bounded interval $[z, \infty[$ suggests that we could concentrate our interest on the right tail of the distribution and no longer on the truncated left tail. To illustrate this idea, we introduce the concept of the upper partial moment (UPM) function of a distribution which is the symmetric counterpart of the LPM function (12):

$$UPM_F^s(x) = \int_x^\infty (t-x)^s dF(t) \qquad x \in [0, +\infty[. \tag{13}$$

For $x = z$ and $s = 1$, $UPM_F^1(z)$ measures the mean publication gap over the reference level z. Intuitively, it makes more sense to try to rank countries using UPM functions which are increasing in x than using LPM functions which are decreasing in x. Roughly speaking, a ranking using UPM functions compares the number of productive academics, while a ranking using LPM functions compares the number of unproductive academics.

We have now to see whether UPM is a meaningful notion and what is its relation to stochastic dominance. Let us define $\tilde{F}_s(x)$ as

$$\tilde{F}_s(x) = \frac{1}{(s-1)!} UPM_F^{s-1}(x) = \frac{1}{(s-1)!} \int_x^\infty (t-x)^{s-1} dF(t). \tag{14}$$

Theorem 1 *A UPM function at the order s for a distribution function F is well defined (convergence of the integral) provided there exist a $j > s-1$ so that*

$$\lim_{t \to \infty} (t-x)^j [1 - F(t)] = 0.$$

The proof of this and of the subsequent theorems are given in the appendix. Let us now investigate the relation existing between ordering UPM functions and stochastic dominance.

Theorem 2 *Let us suppose that the condition stated in theorem 1 is satisfied. Stochastic dominance at the order one and at the order two are strictly equivalent to the ordering of the upper partial moments functions of order zero and one $\forall x \in [0, \infty[$.*

Remarks:

- Hanoch and Levy (1969) already noted that it was equivalent to consider the ordering $F(x) \leq G(x)$ or the ordering $1 - F(x) \geq 1 - G(x)$. A similar property was, however, not noted for second order stochastic dominance.
- When we are interested in restricted stochastic dominance, curves may intersect outside the domain of interest and in this case considering the lower partial moment or the upper partial moment functions, may not give the same ordering. This is what we were looking for.

3.3 Connection with income inequality measurement

Ordering the distributions of authors' scores is closely related to the literature of income inequality measurement. In his pioneering article Atkinson (1970) drew extensively on the same literature on decision making under uncertainty, replacing the individual utility function by a social welfare function.

This parallel has been made and can be justified by the common ground (the expected utility literature). We can therefore borrow some of the concepts developed in the field of income inequality measurement. Typically, z is a poverty level for income measurement (i.e. some minimum subsistance level) while here represents a minimum level of academic activity. Thus $z - x$ is a poverty gap for $x < z$ whilst $x - z$ is an activity gap or production gap for $x > z$.

A large class of inequality indices (those proposed by Foster, Greer and Thorbecke 1984) are mere transformations of LPM. They can thus immediately be translated into indexes expressed as functions of UPM. It can be easily shown that Kakwani (1993) inferential results on inequality indices are directly transposable to the present situation. If we now turn to partial moment functions, empirical estimates of LPM and UPM functions have essentially the same statistical properties and thus the distributional results derived, for instance by Davidson and Duclos (2000) for dominance apply.

Finally we might also borrow from the literature on income inequality measurement another concept, the generalised Lorenz curve. This is defined as the graph of the proportion of individuals (horizontal axis) having at least a given level z of income. Defining p the proportion of the population below z as $p = F(z)$, the generalised Lorenz curve is

$$GL(F(z)) = \int_0^z ydF(y). \tag{15}$$

Using the change of variable $t = F(y)$, we get the alternative form of the Lorenz curve popularized by Gastwirth (1971)

$$GL(p) = \int_0^p F^{-1}(t)\,dt. \tag{16}$$

As shown in the literature (se e.g. Foster and Shorrocks 1988), generalised Lorenz ordering is equivalent to second order stochastic ordering.

Let us now define p as the proportion of the population above a given level z, implying $p = 1 - F(z)$. We arrive at the following expression for a modified generalised Lorenz curve:

$$\begin{aligned} MGL(1-F(z)) &= \int_z^\infty ydF(y) \\ MGL(p) &= \int_{1-p}^1 F^{-1}(t)\,dt. \end{aligned} \tag{17}$$

and to the following definition:

Definition 3 *Country i Lorenz dominates country j in the modified Lorenz sense if $MGL_i(p) \geq MGL_j(p)$ for all $p \in [0,1]$.*

One can prove that generalised Lorenz dominance is equivalent to the second order stochastic dominance. Furthermore, we can prove the same type of equivalence between the modified generalised Lorenz ordering and second order stochastic dominance[5].

3.4 Comparing countries

Let us now specify a minimum admissible level of activity z. We can define this in absolute terms, by choosing a minimum number of published papers of a given quality over the 10 years of the sample. For instance, one might require one paper in a top ranked journal with a single author. This score can also be obtained with 2 papers published in a medium range journal, or 10 papers in low range journals. This would mean accepting $z = 10$ as our academic minimum level of activity. This level may seem very conservative by accepted academic standards, but column 7 of Table 1 shows that it is reached on average by only 14% of the authors in the seven European countries considered!

First order stochastic dominance of F_i over F_j means that the proportion of academics with an output above z is greater in country i than it is in country j, whatever the value of z. This measure counts the proportion of authors above the minimum level, but does not take into account their distribution. It does not change if an author above the minimum increases his production.

[5] A formal proof can be obtained by using a result adapted from Foster and Shorrocks (1988): $\int_x^\infty (F(t) - G(t))\,dt \geq 0$ for all x if and only if $\int_{1-p}^1 (G^{-1}(t) - F^{-1}(t))\,dt \geq 0$ for all $p \in [0,1]$. This result itself can be proved using integration by parts.

Second order stochastic dominance corrects for this since it involves the mean activity gap. The activity gap $\text{Max}(y-x,0)$ measures the distance between the score y of a given author and the reference level $x > z$, provided that this author is above the minimum level of activity. Contrary to the above head count measure, it changes if an author above the minimum level increases his production. In this context, the dominance of country i over country j means that the average activity gap is greater in i than in j, and that this is true for every value of x above z.

4 The choice of a parametric distribution

In many papers (see the references given in Davidson and Duclos 2000) dominance curves (or lower partial moment functions) are estimated from discrete data without making any assumption as to considering the distribution of the sample. These estimators are obtained by replacing the theoretical cumulative distribution function by its natural estimator, $\hat{F}(x) = \sum \mathbb{I}(y_i < x)/N$ and integrals by discrete sums. For LPM and UPM, we have:

$$\begin{aligned} \widehat{LPM}_F^s(x) &= \frac{1}{N}\sum_{i=1}^{N}(x-y_i)^s\,\mathbb{I}(y_i \le x), \\ \widehat{UPM}_F^s(x) &= \frac{1}{N}\sum_{i=1}^{N}(y_i-x)^s\,\mathbb{I}(y_i \ge x). \end{aligned} \tag{18}$$

where $\mathbb{I}$ is the indicator function. Statistical properties of these types of estimators and related tests for stochastic dominance can be found in Kakwani (1993), Kaur, Prakasa Rao and Singh (1994) or Davidson and Duclos (2000).

There are however many goo reasons choosing a parametric formulation for the distribution of y. The direct estimate of the curve $LPM_F^s(x)$ is sensitive to sampling errors and to observations in the extreme tails of the distribution. For instance testing for stochastic dominance requires one to define a grid over x where partial moment functions are compared. The grid chosen may lead to "the veto of the extremes" so that distributions are effectively ordered according to the observed maximum of the sample. To avoid this problem, some authors recommend doing some trimming of the data (see for instance Cowell and Victoria-Feser (2001) and the references cited therein). But this practice can be justified only when the ranges of the samples to be compared are fairly similar. Our samples have a common minimum, but their maxima vary considerably (from 90 to 216). As noted and advocated by Cowell and Victoria-Feser (1996), a parametric formulation of the distribution has several advantages. First it opens the way to classes of estimators with well defined sampling properties. Second, it allows one to detect outliers resulting from coding errors. Finally, we would add that it operates some kind of smoothing for the estimated distribution (see the non-parametric estimates reported in Figure 1 as a point of comparison). For the case considered here, the first argument is particularly important because a parametric formulation for the distribution of the data leads naturally to Bayesian inference. On top of these traditional justifications for using a parametric distribution, there is the problem of the low number of observations which is specific to our case. The population of academic authors is very small compared to the total population. Moreover, we are mainly interested in the right tail of the distribution. Thus it is useful to impose some structure by means of a parametric distribution in order to be able to extract more information from the data. The problem is of course that of choosing an adequate parametric distribution which does justice to the information contained in the data.

4.1 The family of the generalised Beta-II

In the field of income inequality measurement, the generalised Beta-II (GB2) introduced in McDonald (1984) is one of the most general distributions which have been proposed in the literature for fitting income data. The density of this four parameter distribution is given by:

$$f_{GB2}(y|a,b,p,q) = \frac{a y^{ap-1}}{b^{ap}B(p,q)[1+(y/b)^a]^{p+q}} \tag{19}$$

with $y > 0$ and where $B(p,q)$ is the Beta function. The parameters a, b, p and q have to be positive[6]. The analytical expression of its un-centred moments[7] is

$$\begin{aligned} \mathrm{E}(Y^h) = \mu'_h &= b^h \frac{B(p+h/a, q-h/a)}{B(p,q)} \\ &= b^h \frac{\Gamma(p+h/a)\Gamma(q-h/a)}{\Gamma(p)\Gamma(q)} \end{aligned} \tag{21}$$

where $\Gamma(p)$ is the Gamma function. The h^{th} moment exists provided $aq > h$. Note that b is a scale parameter which does not influence the moment ratios. The incomplete h^{th} moments and the cumulative distribution involve confluent hypergeometric functions. McDonald (1984) presents in a graph all the possible reductions of this family of densities. Three densities will be useful for us: the generalised gamma (GG), the Weibull and marginally the Pareto density.

4.2 Related distributions

The generalised gamma density is obtained as a limit of the GB2 when $q \to \infty$

$$\begin{aligned} f_{GG}(y|a,\beta,p) &= \lim_{q\to\infty} f_{GB2}(y|a, b = \beta q^{1/a}, p, q) \\ &= \frac{a y^{ap-1}}{\beta^{ap}\Gamma(p)} \exp[-(y/\beta)^a]. \end{aligned} \tag{22}$$

The cumulative distribution function is

$$F(x) = G[(x/\beta)^a; p] \tag{23}$$

where G is the incomplete Gamma function. The un-centred moments are

$$\mathrm{E}[Y^h] = \beta^h \frac{\Gamma(p+h/a)}{\Gamma(p)}. \tag{24}$$

No parametric restrictions are needed for the existence of the moments, other than positivity restrictions.

The Weibull density is obtained as a restriction of the generalised gamma for $p = 1$:

$$f_W(y|a,\beta) = f_{GG}(y|a,\beta,p=1) = \frac{a}{\beta}(y/\beta)^{a-1}\exp(-(y/\beta)^a) \tag{25}$$

with a and β positive. The cumulative distribution function is

$$F_W(x) = 1 - \exp(-(y/\beta)^a). \tag{26}$$

The un-centred moments reduce to

$$\mathrm{E}(Y^h) = \beta^h \Gamma(1 + h/a). \tag{27}$$

[6] In the most general formulation, there is no restriction on the sign of a because it appears with an absolute value in the numerator of f_{GB2}. But in all the empirical applications we have seen, a was positive, so we finally adopted this more simple notation of the density.

[7] The centred moments and moments ratios are obtained as:

$$\begin{aligned} &\mu_1 = \mu'_1 && \sigma^2 = \mu'_2 - \mu_1^2 \\ &\alpha_3 = (\mu'_3 - 3\mu'_2\mu_1 + 2\mu_1^3)/\sigma^3 && \alpha_4 = (\mu'_4 - 4\mu'_3\mu_1 + 6\mu'_2\mu_1^2 - 3\mu_1^4)/\sigma^4 \end{aligned} \tag{20}$$

The Pareto density corresponds to a completely different line of restriction of the GB2 as now q is finite and becomes the sole parameter while $a = b = p = 1$

$$\begin{aligned} f_{P2}(y|\alpha) &= f_{GB2}(y|a=1, b=1, p=1, q=\bar{\alpha}) \\ &= \bar{\alpha}[1+y]^{-(\bar{\alpha}+1)}. \end{aligned} \tag{28}$$

This is the Pareto of the second kind (or Lomax distribution). The usual Pareto is obtained after a change in variable $x/\sigma = 1 + y$ (with Jacobian $1/\sigma$) so that

$$f(x) = \bar{\alpha}\sigma^{\bar{\alpha}} x^{-(\bar{\alpha}+1)} \mathbb{I}(x > \sigma). \tag{29}$$

The cumulative distribution function is

$$F(x) = 1 - (\sigma/x)^{\bar{\alpha}}. \tag{30}$$

The un-centred moments are

$$\mathrm{E}(X^h) = \sigma^h \frac{\bar{\alpha}}{\bar{\alpha} - h} \qquad \text{if } \bar{\alpha} > h. \tag{31}$$

The existence condition $\bar{\alpha} > h$ is a rather restrictive. For instance, Lotka's law implies $\bar{\alpha} = \alpha - 1 = 2 - 1 = 1$ which precludes the existence of the mean as well as the existence of first order upper partial moment. Consequently, we cannot consider the Pareto distribution as an interesting reduction of the GB2 to conduct our analysis. We shall limit our investigations to the GB2, the generalised gamma and the Weibull, noting that for these last two distributions there are no restrictions for the existence of moments other than positivity restrictions.

4.3 Bayesian model choice

Which density fits our data best and how can we devise an interesting model choice criterion? Lubrano (2001) builds on the old idea of comparing actual data y_d to hypothetical replicated samples y_{rep} obtained by simulating the model. If the model under inspection fits the data in a satisfactory way, the observed sample and the simulated samples should not look too different. To implement this idea, the main stream Bayesian literature (see references below) defines a sample statistics $g(y_d)$ to compute a tail area probability

$$\Pr\{g(y) \geq g(y_d)|M\} \tag{32}$$

which quantifies the extremeness of the observed value $g(y_d)$ with reference to the distribution of the transformation $g(.)$ of the random variable y which represents the possible outcomes of the model. A small tail area probability casts doubts on validity of the model. For Box (1980), this probability is computed in reference to the predictive distribution associated with model M

$$p(y|M) = \int f(y|\theta, M)\, \varphi(\theta|M)\, d\theta. \tag{33}$$

In this formulation, $f(y|\theta, M)$ is the data density, θ the vector of parameters and $\varphi(\theta|M)$ the prior density. Rubin (1984) develops the posterior predictive approach arguing that, if in some situations it may be interesting to operate unconditional frequency checks, in other cases it may be preferable "to regard θ as a fixed feature of the replications" and then to average over the posterior distribution of θ. This is the posterior predictive approach. This last approach has seen a significant development in the statistical literature with Meng (1994), Gelman and Meng (1996) or Gelman, Meng and Stern (1996) and more recently with Bayarri and Berger (2000). With the posterior predictive approach, interest is focused on the fit of the model to the observed data.

The choice of the transformation $g(.)$ is crucial for the method. It is used to summarise the observed sample into a single number. Lubrano (2001) suggests that it is better to avoid choosing a particular transformation and consequently prefers to compare a non-parametric estimate $\hat{f}(x)$ of the density of the observed sample directly with the data density $f_M(x|\theta)$ of the model using the square of the Hellinger distance defined as

$$D_H(\theta) = 2(1 - \int \sqrt{\hat{f}(x)\, f_M(x|\theta)}\, dx). \tag{34}$$

We have to approximate the integral in (34) using a Simpson rule and a predefined grid x_j of k points covering the sample range. Given N posterior draws θ_i, we obtain N posterior draws of D_H. Model A will be preferred to model B if the posterior distribution of D_H^A stochastically dominates the posterior distribution of D_H^B in the traditional sense. A necessary condition is provided by the ordering of the means implying that the preferred model has the smallest posterior mean for D_H. It is rather difficult indicate a particular value for D_H over which a richer model than those considered should have to be considered (for instance mixtures) because the value of D_H is sensitive to the window size used for the non-parametric estimate. For more details see Bos and Lubrano (2002).

4.4 Bayesian inference for adjusting parametric distributions

The classical statistical literature abounds with exemples of the numerical difficulties induced in applying the maximum likelihood method to the generalised gamma distribution (see e.g. Johnson, Kotz and Balakrishnan 1994 and also Tsionas 2001) and we can reasonably suppose that these difficulties apply also to the richer GB2 distribution. In general, the statistical literature does not recommend to use the GG model for empirical work unless there are enough observations to group data in a frequency table. Considering these difficulties, the classical literature favours the use of a method of moments. We propose to use a generalised method of moments to estimate the parameters. These classical estimates are necessary as starting points in the subsequent MCMC methods used to integrate the posterior density.

We leave on one side the Pareto model which has received a particular treatment in the literature (see e.g. Arnold and Press 1983). To our knowledge, there does not exist a Bayesian treatment for inference with the generalised Beta-II. Tsionas (2001) has proposed a Bayesian analysis for the generalised gamma. He introduces a reparameterisation of the density which provides an analytical expression for the conditional posterior density of the scale parameter. He then builds a Gibbs sampler algorithm. Tsionas (2000) proposes a similar approach for the Weibull density. This elegant reparameterisation is not convenient in our context as it makes difficult the computation of partial moments functions. So we stick to the original parameterisation.

The three models we shall consider have one scale parameter b or β and one, two or three shape parameters: a (Weibull), a and p (generalised gamma), a, p and q (generalised Beta-II). Either from the classical or the Bayesian side, the likelihood function presents the same pitfalls. In the GB2 model, we have an extreme negative correlation between a and q and between b and p. For the GG model, there is an extreme negative correlation between b and p. For these two models, the posterior density cannot be feasibly integrated under a non-informative prior. Tsionas (2001) evokes the possibility of independent exponential priors. We adopt the following specification:

$$\mu(a, b, p, q) \propto \theta \exp(-\theta(a + 1/b + p + 1/q))$$

which might counterbalance the extreme negative correlations. Let us call $l(x|a, b, p, q)$ the likelihood function, the posterior density of the parameters is then

$$\mu(a, b, p, q|x) \propto \mu(a, b, p, q)\, l(x|a, b, p, q).$$

We shall use an independent Metropolis algorithm with a Student importance function centred on the moments estimates and with 4 degrees of freedom to integrate this density. We shall iterate the method several times with 20 000 draws each time until convergence which was checked using CUMSUM graphs. We took as prior values $\theta = 3$ for the GB2 model, $\theta = 2$ for the GG model and $\theta = 0$ for the Weibull model which does not present special numerical difficulties.

4.5 Posterior results

Posterior means and standard deviations are presented in Table 2 together with the posterior mean of the Hellinger distance attached to each model.

Table 2: Bayesian inference results

	Bel	Fra	Ger	Ita	Net	Spa	UK	Cal
				Generalised Beta-II				
a	3.700 [0.655]	5.650 [0.505]	5.487 [0.600]	6.266 [0.633]	1.277 [0.133]	9.413 [0.939]	1.028 [0.063]	0.789 [0.066]
b	0.492 [0.044]	0.500 [0.026]	0.492 [0.026]	0.502 [0.024]	0.539 [0.093]	0.492 [0.017]	0.266 [0.037]	0.437 [0.092]
p	3.196 [0.789]	4.156 [0.899]	4.318 [0.912]	3.881 [0.848]	4.478 [0.811]	3.733 [0.855]	8.987 [1.101]	6.361 [0.910]
q	0.209 [0.047]	0.144 [0.015]	0.154 [0.019]	0.137 [0.016]	0.830 [0.138]	0.088 [0.009]	1.103 [0.114]	1.511 [0.238]
Hel	0.485	0.500	0.377	0.407	0.346	0.464	0.239	0.334
				Generalised Gamma				
a	0.438 [0.020]	0.396 [0.008]	0.428 [0.011]	0.439 [0.012]	0.407 [0.013]	0.421 [0.013]	0.373 [0.006]	0.368 [0.009]
b	0.298 [0.101]	0.075 [0.014]	0.092 [0.019]	0.111 [0.025]	0.158 [0.040]	0.132 [0.033]	0.058 [0.010]	0.119 [0.027]
p	3.181 [0.284]	4.410 [0.192]	4.438 [0.227]	4.174 [0.237]	3.936 [0.242]	3.671 [0.228]	4.969 [0.168]	4.122 [0.196]
Hel	0.327	0.475	0.203	0.222	0.279	0.279	0.195	0.291
				Weibull				
a	0.748 [0.020]	0.758 [0.011]	0.845 [0.013]	0.840 [0.015]	0.773 [0.014]	0.764 [0.015]	0.809 [0.008]	0.738 [0.010]
b	5.215 [0.286]	4.276 [0.126]	3.924 [0.110]	3.710 [0.118]	6.045 [0.216]	3.759 [0.147]	5.855 [0.103]	7.587 [0.202]
Hel	0.292	0.267	0.172	0.185	0.250	0.238	0.171	0.266

Standard deviations are between brackets. The *Hel* line gives the posterior mean of the Hellinger distance.

Figure 2 presents the graph of the posterior density of the Hellinger distance. It is clear from this graph that there is a huge cost to complexity for the GB2 and the GG models. The Weibull model is rewarded for its parsimony and flexibility. We have tried other models, including the Burr XII and the Pareto which were also rejected, especially the latter.

5 Ranking countries using the Weibull model

5.1 Stochastic ordering

The partial moment function for stochastic dominance of order $s = 1$ has a simple analytical expression as it is a mere translation of the cumulative distribution function :

$$UPM^0(x|a,b) = 1 - F(x) = \exp(-\left(\frac{x}{b}\right)^a). \tag{35}$$

Stochastic dominance of F over G at the order $s = 1$ means $UPM_F^{s-1}(x) \geq UPM_G^{s-1}(x)$ for all x. We have the following theorems. Proofs are given in the appendix.

Theorem 3 *Within the Weibull model, distribution F stochastically dominates distribution G at the order one if $a_F \leq a_G$ and $b_F \geq b_G$ provided $b_G > 1$. G dominates F if $a_F \geq a_G$ and $b_F \leq b_G$ provided $b_F > 1$. In the other cases, we cannot conclude because the partial moment functions intersect.*

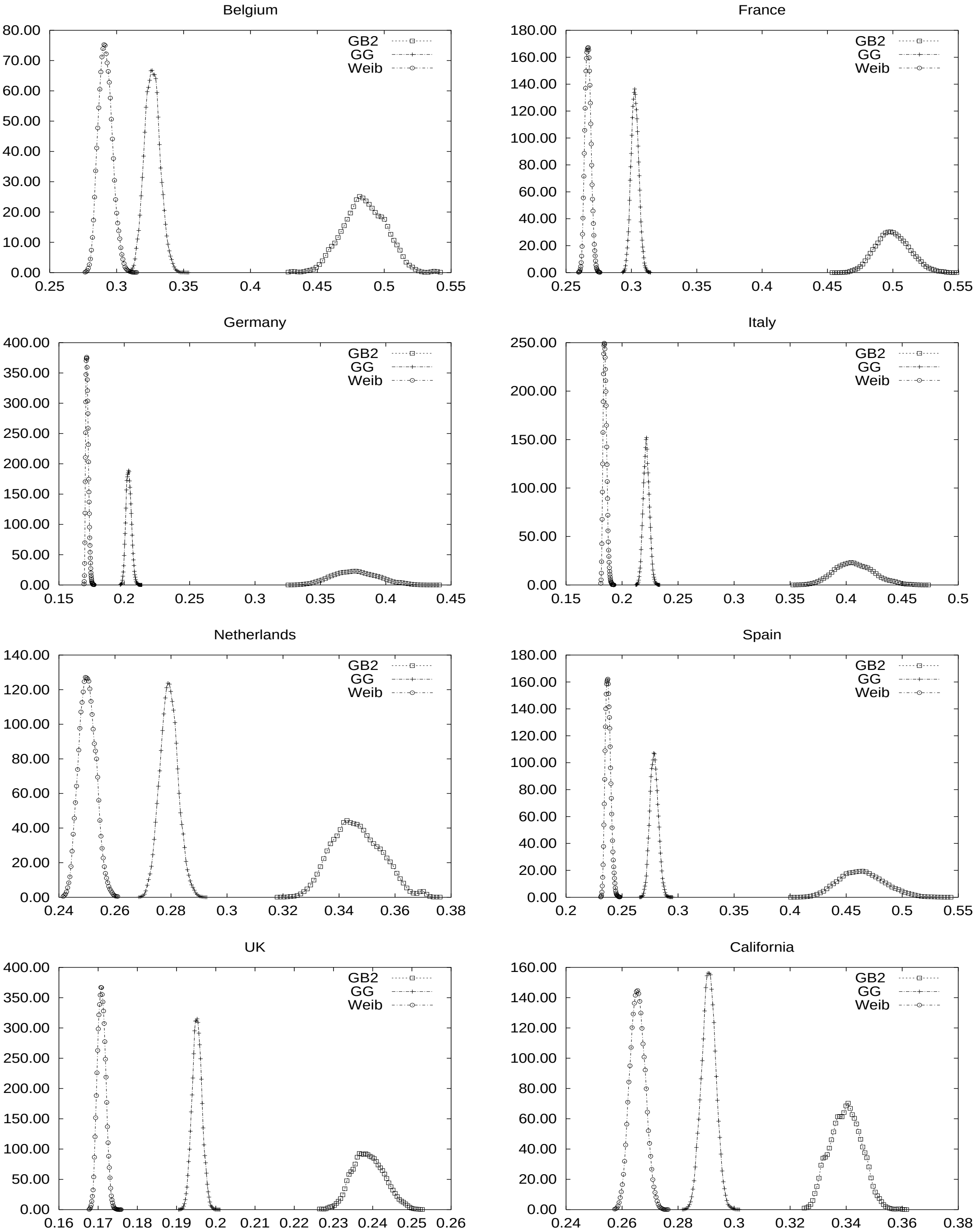


Figure 2: Posterior density of the Hellinger distance between data density and posterior predictive density

Theorem 4 *The partial moment function for stochastic dominance at the order $s = 2$ is*

$$UPM^1(x|a,b) = \frac{b}{a}\Gamma\left(\frac{1}{a}\right)\left[1 - G\left(\left(\frac{x}{b}\right)^a; \frac{1}{a}+1\right)\right] - x\exp(-\left(\frac{x}{b}\right)^a)$$

where $G(x;p)$ is the incomplete Gamma function.

Theorem 5 *The modified generalised Lorenz curve for the Weibull model is*

$$MGL(p|a,b) = \frac{b}{a}\Gamma\left(\frac{1}{a}\right)[1 - G(-\log(p); \frac{1}{a}+1)].$$

Lorenz ordering of F over G means $MGL_F(p) \geq MGL_G(p)$ for all p and is verified if $a_F \leq a_G$ and $b_F \geq b_G$.

Second order stochastic ordering conditions follow using the equivalence between Lorenz ordering and second order stochastic dominance.

5.2 Bayesian testing for stochastic dominance

We have a complete stochastic ordering of F over G if $a_F \leq a_G$ together with $b_F \geq b_G$. This type of joint inequality is relatively easy to test in a Bayesian framework. When two models are estimated by a MCMC method (or more simply when one manages to simulate draws in the posterior distribution of the parameters), the probability of the joint event $\Pr[(a_F \leq a_G) \bigcap (b_F \geq b_G)|y_F, y_G]$ is simply obtained as the proportion of times this event is true among the draws of the posterior densities of the parameters of models F and G.

When partial moment curves intersect, it is no longer possible to have a complete stochastic ordering. But the same partial moment curves may not intersect for a restricted range of x. This case is more tricky as restricted stochastic dominance is no longer obtained as a direct restriction on the parameters of two distributions. One has to compute two partial moment curves for a given grid of x. But as these curves are after all transformations of the parameters of the estimated distributions, what we have to compare are two specific transformations of the parameters. Let us define $d^s(x|\theta) = UPM_F^s(x|\theta_F) - UPM_G^s(x|\theta_G)$. We want to test for stochastic dominance of F over G when $x \in [z_*, z^*]$. We define a grid $[x_j]$ of m points for x with endpoints z_* and z^*. To characterize the event that $d^s(x|\theta) \geq 0$, we focuss our interest on the probability that $\operatorname{Min}_j d^s(x_j|\theta)$ is positive:

$$\begin{aligned} \Pr(\operatorname{Min}_j d^s(x_j|\theta) \geq 0) &= \int_\theta \mathbb{I}[\operatorname{Min}_j d^s(x_j|\theta) > 0]\, \mu(\theta|y)\, d\theta \\ &\simeq \frac{1}{N}\sum_i \mathbb{I}[\operatorname{Min}_j d^s(x_j|\theta_i) > 0] \end{aligned} \tag{36}$$

where the θ_i are the N draws from the posterior densities of θ_F and θ_G.

5.3 Ranking countries

We first ignore the value of z, the minimum level of activity and compute the probability that a country strictly dominates another one. Table 3 shows that few numbers are large, indicating that we have a total ordering only for very few cases, e.g. most partial moment functions cross. It is quite clear that Italy is dominated by all the other countries, except perhaps by Germany (but Italy does not dominate Germany). California, at the other end, dominates all the other countries. We encourage the reader to examine this table further. It appears that the heterogeneity in the academic population precludes a useful ranking. So we have to restrain our attention to the segment $[10, +\infty[$.

Table 4 presents posterior probabilities of restricted stochastic dominance at the order one. The image is now much clearer. California increases its leading position as all probabilities of its line

Table 3: Posterior probability of complete stochastic dominance

	Cal	Net	UK	Bel	Fra	Spa	Ger	Ita
Cal	1.000	0.996	1.000	0.639	0.913	0.975	1.000	1.000
Net	0.000	1.000	0.752	0.059	0.083	0.262	1.000	0.997
UK	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.990
Bel	0.000	0.006	0.011	1.000	0.632	0.861	1.000	0.999
Fra	0.000	0.000	0.000	0.000	1.000	0.797	1.000	1.000
Spa	0.000	0.000	0.000	0.000	0.000	1.000	0.141	0.616
Ger	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.390
Ita	0.000	0.000	0.000	0.000	0.000	0.000	0.049	1.000

A line indicates the probability that the horizontal country dominates the countries given in columns. A column indicates the probability a country is dominated by the line countries.

Table 4: Posterior probability of restricted stochastic dominance at the order one over the range $[10, \infty[$

	Cal	Net	UK	Bel	Fra	Spa	Ger	Ita
Cal	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Net	0.000	0.000	0.914	0.707	1.000	1.000	1.000	1.000
UK	0.000	0.000	0.000	0.032	0.996	1.000	1.000	1.000
Bel	0.000	0.014	0.080	0.000	1.000	1.000	1.000	1.000
Fra	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000
Spa	0.000	0.000	0.000	0.000	0.000	0.000	0.959	0.989
Ger	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.721
Ita	0.000	0.000	0.000	0.000	0.000	0.005	0.091	0.000

are equal to one. The Netherlands dominates all other European countries. There is an ambiguity about the ranking of Belgium and the UK. France and Spain follow. There is an ambiguity about the ranking of Germany and Italy. Going to the second order with Table 5 clarifies the situation for Germany and Italy, but ambiguity as to the relative ranks of the UK and Belgium remains.

Figure 3 displays the posterior expectation of the upper partial moment function. It illustrates the previous ranking.

Table 5: Posterior probability of restricted stochastic dominance at the order two over the range $[10, \infty[$

	Cal	Net	UK	Bel	Fra	Spa	Ger	Ita
Cal	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Net	0.000	0.000	0.994	0.667	1.000	1.000	1.000	1.000
UK	0.000	0.007	0.000	0.025	0.984	1.000	1.000	1.000
Bel	0.000	0.048	0.436	0.000	1.000	1.000	1.000	1.000
Fra	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000
Spa	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000
Ger	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.721
Ita	0.000	0.000	0.000	0.000	0.000	0.000	0.136	0.000

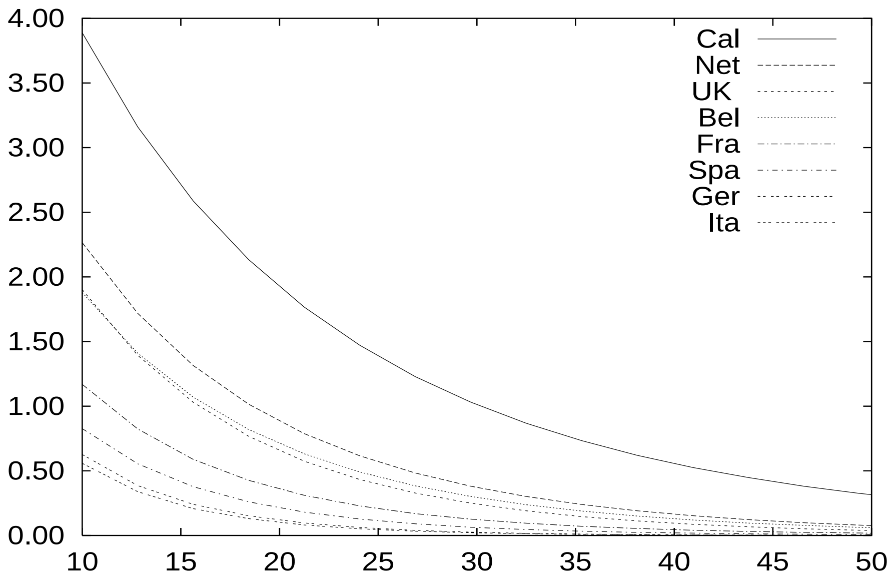


Figure 3: UPM functions for second order stochastic dominance

Remark:

Because it has a single parameter, the Pareto distribution leads to partial moment functions $UMP^0(x) = 1 - F(x) = \sigma^\alpha x^{-\alpha}$ that can never intersect, provided the σ are the same, which is the case here. Consequently it is not possible in this model to differentiate between complete and restricted stochastic ordering. Our empirical results show that this distinction may be important. This is a stylised fact which mitigates against Lotka's law. The Weibull is the most parsimonious distribution that allows curve crossing.

6 Conclusion

Using a parametric and a Bayesian approach, we were able to provide posterior probabilities for the proposed country rankings. Surprisingly the Weibull density with only two parameters is preferred to more elaborate models such as the generalised gamma density or the generalised Beta-II. Only a mixture of two Weibull densities could have slightly improved the fit according to the experiments we made, but this was at the cost of losing the nice analytical results concerning the ordering of partial moment functions. We are fairly confident in the results obtained with the Weibull model as this model is both simple and sufficiently flexible.

As for the empirical results, we have not ranked institutions, but countries. We have considered a picture of the whole academic system, because we mixed all the authors of a same country . This penalises countries having numerous economic departments like the UK. Netherlands comes first in Europe because it has few departments, but all of them are of a fairly high quality.

Which country will our PhD student select in the face of uncertainty? If he had really the choice, he would select unambiguously California. Figure 3 shows that its posterior UPM function is well above all the other curves. If we exclude this choice, Figure 3 suggests a leading group of three European countries: the Netherlands, Belgium and the UK. The surprise of this study is that the UK is not first in this leading group.

APPENDIX

A Proof of theorems

Proof 1:

$$\begin{aligned}\int_x^{\infty}(t-x)^{s-1}dF(t) &= -\int_x^{\infty}(t-x)^{s-1}d[1-F(t)] \\ &= -\left[(t-x)^{s-1}[1-F(t)]\right]_x^{\infty} \\ &\quad + (s-1)\int_x^{\infty}(t-x)^{s-2}[1-F(t)]dt.\end{aligned}$$

The first term converges to zero by the above assumption. The second integral converges because

$$(t-x)^{s-2}[1-F(t)] = (t-x)^j[1-F(t)]\frac{1}{(t-x)^{j-s+2}}$$

and $\lim_{t\to\infty}(t-x)^j[1-F(t)] = 0$ by assumption, while $1/(t-x)^{j-s+2}$ is integrable.

□

Proof 2: For $s=1$, the definition of UPM implies that

$$\tilde{F}_1(x) = \int_x^{\infty} f(t)dt = 1 - F_1(x). \tag{37}$$

So for two distributions F and G, ordering the UPM functions of order zero for all x is equivalent to first order stochastic dominance because

$$F(x) \leq G(x) \Leftrightarrow 1-F(x) \geq 1-G(x) \qquad \forall x \in [0,+\infty[\tag{38}$$

For $s=2$, as in the proof of Theorem 1, we obtain that

$$\tilde{F}_2(x) = \int_x^{\infty}(t-x)dF(t) = \int_x^{\infty}[1-F(t)]dt$$

when $F_2(x)$ is

$$\begin{aligned} F_2(x) &= \int_0^x (x-t)dF(t) = [(x-t)F(t)]_0^x - \int_0^x -F(t)dt \\ &= -\int_0^x [1-F(t)-1]dt = x - \int_0^x [1-F(t)]dt. \end{aligned}$$

Let us consider two distributions F and G for which we have $F_2(x) \leq G_2(x)\ \forall x$, which means

$$x - \int_0^x [1-F(t)]dt \leq x - \int_0^x [1-G(t)]dt.$$

This is strictly equivalent to

$$\int_x^\infty [1-F(t)]dt \geq \int_x^\infty [1-G(t)]dt.$$

Thus for two distributions F and G, $LPM_F^1(x) \leq LPM_G^1(x)\ \forall x$ is equivalent to $UPM_F^1(x) \geq UPM_G^1(x)\ \forall x$.

□

Proof 3: $\exp(-(x/b_F)^{a_F}) \geq \exp(-(x/b_G)^{a_G})$ is equivalent to $(x/b_F)^{a_F} \leq (x/b_G)^{a_G}$ and the result follows because x is positive.

□

Proof 4: Introducing the formula of the Weibull density into the expression of the upper partial moment function for $s = 2$, we have

$$UPM^1(x|a,b) = \frac{a}{b^a}\int_x^\infty (y-x)y^{a-1}\exp\left(-(y/b)^a\right)\,dy.$$

Let us consider the change of variable $x = (y/b)^a$ which implies $dy = (b/a)x^{1/a-1}$. We have

$$UPM^1(x|a,b) = b\int_{(x/b)^a}^\infty x^{1/a}\exp(-x)\,dx - x\int_{(x/b)^a}^\infty \exp(-x)\,dx$$

which is equal to

$$\frac{b}{a}\Gamma\left(\frac{1}{a}\right)[1 - \frac{1}{\Gamma(1/a+1)}\int_0^{(x/b)^a} x^{1/a}\exp(-x)\,dx] - x\exp(-(x/b)^a).$$

The result follows when we recognize the expression of the incomplete Gamma function.

□

Proof 5: Let us combine the definition of the modified generalised Lorenz curve with the expression of the inverse of the cumulative Weibull distribution

$$MGL(p|a,b) = \int_{1-p}^1 b[\log(1/(1-t))]^{1/a}dt.$$

Let us introduce the change of variable $y = \log(1/(1-t))$.

$$MGL(p|a,b) = b\int_{-\log(p)}^\infty y^{1/a}\exp(-y)dy.$$

Let us decompose this integral into the difference of two integrals so that

$$\begin{aligned} & b\Gamma(1+1/a)\left[\frac{1}{\Gamma(1+1/a)}\int_0^{\infty} y^{1/a}\exp(-y)dy \right. \\ & \left. -\frac{1}{\Gamma(1+1/a)}\int_0^{-\log(p)} y^{1/a}\exp(-y)dy\right] \\ = \; & b\Gamma(1+1/a)\left[1-\frac{1}{\Gamma(1+1/a)}\int_0^{-\log(p)} y^{1/a}\exp(-y)dy\right]. \end{aligned}$$

Lorenz ordering conditions follow because the MGL is a decreasing function in a and an increasing function in b.

□

Table 6: Top journal ranking

Journals	Grade	Journals	Grade
American-Economic-Review	10	Industrial-and-Labor-Relations-Review	6
Econometrica	10	International-Journal-of-Game-Theory	6
Journal-of-Economic-Theory	10	Journal-of-Applied-Econometrics	6
Journal-of-Political-Economy	10	Journal-of-Banking-and-Finance	6
Quarterly-Journal-of-Economics	10	Journal-of-Business-and-Economic-Statistics	6
Review-of-Economic-Studies	10	Journal-of-Comparative-Economics	6
American-Political-Science-Review	8	Journal-of-Development-Economics	6
International-Economic-Review	8	Journal-of-Economic-Behavior-and-Organization	6
Journal-of-Econometrics	8	Journal-of-Economic-Dynamics-and-Control	6
Journal-of-Economic-Literature	8	Journal-of-Economic-Growth	6
Journal-of-Finance	8	Journal-of-Economic-History	6
Journal-of-Financial-Economics	8	Journal-of-Economic-Methodology	6
Journal-of-International-Economics	8	Journal-of-Economic-Perspectives	6
Journal-of-Labor-Economics	8	Journal-of-Economics-and-Management-Strategy	6
Journal-of-Law-and-Economics	8	Journal-of-Environmental-Economics-and-Management	6
Journal-of-Monetary-Economics	8	Journal-of-Financial-and-Quantitative-Analysis	6
Journal-of-Money,-Credit,-and-Banking	8	Journal-of-Health-Economics	6
Journal-of-Public-Economics	8	Journal-of-Human-Resources	6
Journal-of-the-American-Statistical-Association	8	Journal-of-Industrial-Economics	6
Michigan-Law-Review	8	Journal-of-Law,-Economics-and-Organization	6
Rand-Journal-of-Economics	8	Journal-of-Mathematical-Economics	6
Review-of-Economics-and-Statistics	8	Journal-of-Risk-and-Insurance	6
Yale-Law-Journal	7	Journal-of-Risk-and-Uncertainty	6
Accounting-Review	6	Journal-of-Urban-Economics	6
American-Journal-of-Agricultural-Economics	6	Macroeconomic-Dynamics	6
Brookings-Papers-on-Economic-Activity	6	Marketing-Science	6
Demography	6	Mathematical-Methods-of-Operations-Research	6
Econometric-Theory	6	National-Tax-Journal	6
Economica	6	Oxford-Bulletin-of-Economics-and-Statistics	6
Economic-Journal	6	Public-Choice	6
Economics-Letters	6	Regional-Science-and-Urban-Economics	6
Economic-Theory	6	Scandinavian-Journal-of-Economics	6
European-Economic-Review	6	Social-Choice-and-Welfare	6
Games-and-Economic-Behavior	6	Urban-Studies	6

REFERENCES

Arnold B.C. and S.J. Press (1983) Bayesian inference for Pareto populations. *Journal of Econometrics* 21, 287-306.

Atkinson A.B. (1970) On the measurement of inequality. *Journal of Economic Theory* 2, 244-263.

Atkinson A.B. (1987) On the measurement of poverty. *Econometrica* 55 (4), 749-764.

Bauwens L. (1999) Economic Research In Belgian Universities. http://www.core.ucl.ac.be /econometrics /Bauwens /rankings /rankings.htm

Bawa V. (1975) Optimal rules for ordering uncertain prospects. *Journal of Financial Economics* 2, 95-121.

Bayarri M.J. and J.O. Berger (2000) P values for composite null hypothesis. *Journal of the American Statistical Association* 95, 1127-1142.

Blackwell D. (1953) Equivalent comparisons of experiments. *Annals of Mathematical Statistics* 24, 265-272.

Bos C.S. and M. Lubrano (2002) Puzzling model choice for overdispersed samples: a Bayesian approach using bibliometric data. Paper presented at the 2002 EC2 meeting in Bologna.

Box, G.E.P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society* A 143, 383–430

Combes P.P. and L. Linnemer (2001) La publication d'articles de recherche en économie en France. *Annales d'Economie et de Statistique* 62, 5-47.

Coupé Tom (2000) Revealed Performances Worldwide Rankings of Economists and Economics Departments. *Mimeo, ECARES* Université Libre de Bruxelles.

Cowell F. and M.P. Victoria-Feser (1996) Robustness Properties of Inequality Measures. *Econometrica* 64 (1), 77-101.

Cowell F. and M.P. Victoria-Feser (2001) Distributional dominance with dirty data. *DARP discussion paper* No 51, LSE, London.

Cox R.A.K. and K.H. Chung (1991) Pattern of research output and author concentration in the economic literature. *The Review of Economics and Statistics* 73(4), 740-747

Cribari-Neto F., M. Jensen and A. Novo (1999). Research in Econometric Theory: Quantitative and Qualitative Productivity Rankings. *Econometric Theory*, 15, 719-752.

Davidson R. and J.Y. Duclos (2000) Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* 68 (6), 1435-1464.

Dusansky R. and C.J. Vernon (1998) Ranking of US economic departments. *Journal of Economic Perspectives* 12, 157-170.

Fishburn P.C. (1977) Mean-risk analysis with risk associated with below-target return. *American Economic Review* 67, 116-126.

Foster J., J. Greer and E. Thorbecke (1984) A class of decomposable poverty measures, *Econometrica* 52 (3), 761-766.

Foster J.E. and A. Shorrocks (1988) Poverty orderings. *Econometrica* 56 (1), 173-177.

Gastwirth J. (1971) A general definition of the Lorenz curve. *Econometrica* 39 (6), 1037-1039.

Gelman A. and Meng X.L. (1996) Model checking and model improvement. *Practical Markov Chain Monte Carlo* (W. Gilks, S. Richardson and S. Spiegelhalter, eds.). London: Chapman and Hall, 189–201.

Gelman A., Meng X.L. and Stern H. (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6**, 733-807.

Hadar J. and W. R. Russell (1969) Rules for ordering uncertain prospects. *American Economic Review* 59, 25-34.

Hanoch G. and H. Levy (1969) The efficiency analysis of choices involving risk. *Review of Economic Studies* 36, 335-346.

Jean W.H. (1984) The harmonic mean and other necessary conditions for stochastic dominance. *The Journal of Finance* 39(2), 527-534.

Johnson N.L., S. Kotz and N. Balakrishnan (1994) *Continuous Univariate Distributions*. New-York, Wiley

Kakwani N. (1993) Statistical inference in the measurement of poverty. *Review of Economics and Statistics* 75, 632-639.

Kaur A., B.L.S. Prakasa Rao and H. Singh (1994) Testing for second order stochastic dominance of two distributions. *Econometric Theory* 10, 849-866.

Kalaitzidakis P., T.P. Mamuneas and T. Stengos (1999) European economics: an analysis based on publications in the core journals. *European Economic Review* 43, 1150-1168.

Lehmann E.L. (1955) Ordered families of distributions. *Annals of Mathematical Statistics* 26, 399-419.

Lotka A.J. (1926) The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16 (12), 317-323.

Lubrano M. (2001) Bayesian model choice for the short term US interest rate. in Edward I. George, *Bayesian methods with application to science, policy and official statistics*, Eurostat 2001, 341-350.

Lubrano M., L. Bauwens, A.P. Kirman and C. Protopopescu (2002) Some Methodological Aspects for Ranking European Economic Departments: a Statistical Approach. To appear in the *Journal of the European Economic Association*.

Markowitz H.M. (1952) Portfolio selection. *Journal of Finance* 7 (1), 77-91.

McDonald J.B. (1984) Some generalised functions for the size distribution of income. *Econometrica* 52, 647-663.

Meng, X.L. (1994). Posterior predictive p-values. *The Annals of Statistics* 22, 1142–1160.

Quirk J.P. and R. Saposnik (1962) Admissibility and measurable utility functions. *Review of Economic Studies* 29, 140-146.

Rubin, D.B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics* **12**, 1151–1172.

Tsionas E.G. (2000) Posterior analysis, prediction and reliability in three-parameter Weibull distributions. *Communications in Statistics (Theory and Methods)* 29 (7), 1435-1449.

Tsionas E.G. (2001) Exact inference in four-parameter Generalized Gamma distributions. *Communications in Statistics (Theory and Methods)* 30 (4), 747-756.

van Damme Eric (1996). Measuring quality of academic journals and scientific productivity of researchers. *Mimeo, CenTer*, Tilburg University.