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

Michel Lubrano[1] Camelia Protopopescu[2]

June 2002

Abstract

This paper details how the framework of 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 valids 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 for 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 countries.

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, Jean-Pierre Florens, Nicolas Gravel, 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 in the USA. See e.g. Dusansky and Vernon (1998). If we now turn to Europe, there is the earlier work of Kirman and Dahl (1994) and the more recent paper of Kalaitzidakis, Mamuneas and Stengos (1999). 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. Some other authors have focussed on particular domains such as econometric like Cribari-Neto, Jensen and Novo (1999). But none of these authors accepted to consider the idea 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 is not concerned with ranking, but investigates the "patterns of research output and author concentration" to paraphrase the title of Cox and Chung's (1991) article. This literature started with the historical work of Lotka (1926) who plotted on a 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. The generalised law of Lotka as introduced for instance in Cox and Chung (1991) is

$$a_n = a_1/n^c \qquad n = 1, 2, 3, ... \tag{1}$$

where a_n is the number of authors having made n contributions and a_1 the number of authors having made one contribution. When $c = 2$, this empirical law implies that 60% of the authors have only one contribution[1]. Several theoretical justifications where given to Lotka's law, see the references given in Cox and Chung (1991), but to our knowledge, it was not noticed that the Pareto distribution is at the corner of the street. 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 where his paper is published and the number of co-authors involved. An index is thus constructed which acts as a proxy for an underlying continuous random variable. Thus, turning from n (discrete variable) to y (continuous variable), Lotka's law implies $f(y) = f(1)\, y^{-\alpha}$. Under the constraint that $f(y)$ integrates to unity, it comes that:

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

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[2]. Our purpose in 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 Pareto distribution has some very restrictive features as will be underlined in the sequel. So a more general distribution is needed and the Generalised Beta-II used to model the income distribution is a good starting point. Once we have found a nice class of distributions, we need a statistical tool to compare country distributions. We show in the

[1] 60.79% to be precise according to the proof given in Cox and Chung (1991).

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

paper that stochastic dominance is the right tool to use, provided we specify correctly the underlying decision problem.

The paper adopts a Bayesian approach for several reasons. The classical econometric literature devoted to tests of stochastic dominance relies on asymptotic normal approximations (see the references given e.g. in Davidson and Duclos 2000). According to Zheng (2001), these approximations are valid when the sample is larger than 1 000 observations. Our data sets may be much smaller because we consider intrinsically small populations (academics in a specific field). Bayesian inference allows us to take into account in a simple way the uncertainty which is attached to the parameters. We can then compute posterior probabilities of stochastic dominance as simple functions of the 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. We propose a new approach to its implementation which is specific to our particular problem. We underline the connection existing between ranking academic institutions and income inequality analysis. In section four, we detail some useful parametric distributions related to the Generalised Beta-II and indicate a Bayesian procedure for model selection. In section five, we use the Weibull model to rank countries while 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 on a given period. A lot of information can be obtained by a simple inspection of these data.

2.1 Sources of information

The data we shall examine are transformations of a counting process where the objects counted are papers published over a given period of observation. We have chosen to extract these observations from the JEL CD-ROM for 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. Bauwens, Kirman, Lubrano and Protopopescu (2002) contains the full details.

The second step consists in ranking the journals where these articles have been published. The JEL contains 681 journals. European countries have used mainly 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 to citation data available from the Journal Citation Reports for 167 journals. There were discrepancies only for 20 journals which were either professional journals at the boarder of the field or small academic journals. We give in the appendix the first top 68 journals. In the field of econometrics, *Econometrica* is 10, the

Journal of Econometrics 8, *Econometric Theory* 6, the *Journal of Applied Econometrics* 6 and *Econometrics Review* 4.

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 the paper is published. The total score of an author is obtained by simply summing all his p_i over the period. This formula deserves two comments. First, it does not take into account the number of pages. Cribari-Neto *et al* (1999) reports that the main journals used by econometricians have an average page number which goes from 10 to 23. If we introduce the number of pages in (3), we would dramatically alter the ranking of the journals. Second, this formula is over additive as it does not divide the merit of a paper exactly between its coauthors. Using $\sqrt{n_i}$ instead of n_i favours co-authorship[3].

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

Table 1: Statistics of authors scores

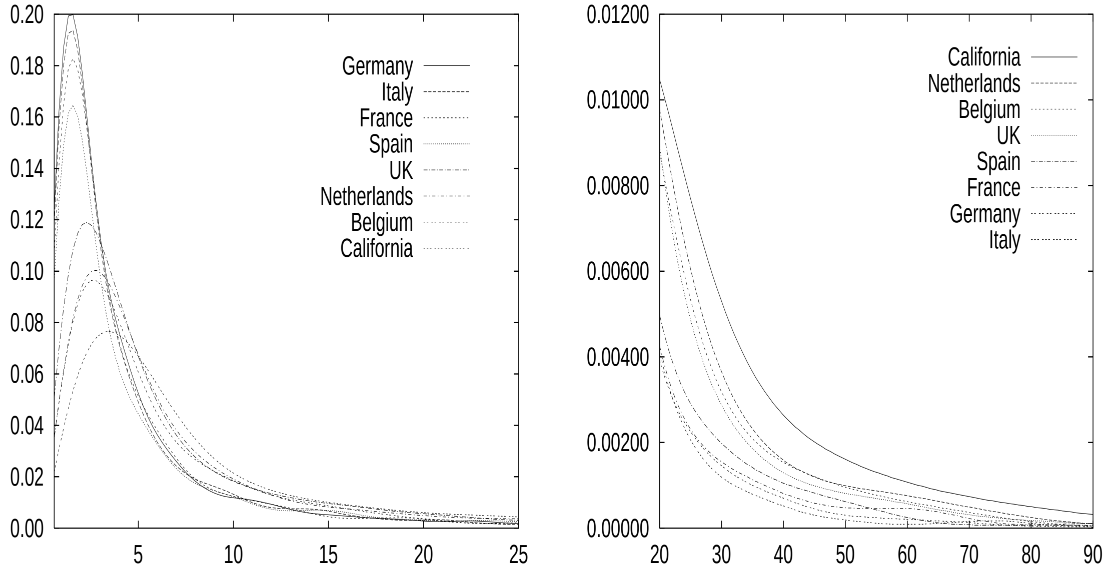
Country	Authors	Aut./Pop.	Mean	Median	Max	$1-F(10)$
Belgium	806	80.60	6.56	2.31	146	0.15
France	2699	45.76	5.50	2.00	216	0.11
Germany	2506	30.26	4.40	2.00	98	0.10
Italy	1921	32.87	4.19	1.71	102	0.09
Netherlands	1793	111.94	7.35	2.83	187	0.18
Spain	1527	39.15	4.64	1.42	90	0.12
UK	6656	117.55	6.78	2.83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

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 display complementary characteristics on the authors' score. Compared to the maximum, the average score is very low and the median even lower. The median score is roughly 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.

[3] Cribari-Neto et al (1999) argues that it is better to use $\sqrt{n_i}$ for ranking individuals and n_i for ranking institutions. But the different calculations we made and that are not reported in the paper show that this option does not change significantly the results.

2.3 Non-parametric density inference

As we aim at adjusting a parametric density on these data, a non-parametric kernel smoother[4] helps in getting an idea about the shape of the empirical distributions. Because all the weight of the distribution is located for low values of y when the maximum of y can be very large, we have split Figure 1 in 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 2 in the left panel and to 10 in the right panel[5]. The left panel of Figure 1 indicates two



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

Figure 1: Density estimates of authors' scores

groups of four countries each. The same grouping is suggested by the right panel, but in reverse order. Up to now, it is not easy to single out a criterion to rank these distributions.

3 Stochastic dominance and academic ranking

It is difficult from the previous empirical investigation to have an intuition about a single measure for ranking academic systems (understood as a collection of authors). A detour by the stochastic dominance and the expected utility literature is useful. The notion of stochastic dominance is quite old as it can be traced back to e.g. Blackwell (1953) or Lehman (1955). But its developments are more recent.

[4]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, $c = 0.9$.

[5]We have slightly over-smoothed to get a clearer image. 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.

3.1 Maximizing expected utility

Let us consider the case of a PhD student applying for a CEE grant and having to choose a country where to go in order to write his PhD thesis. He is looking for two things: a PhD supervisor and a scientific environment (externalities), both being characterised by the random variable Y which measures the performance of academics (their publishing score) in a given country. He is faced to 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

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

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 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, no conclusion is available.

Second order (or second degree) stochastic dominance is based on the comparison of the surface under the cumulative distribution functions.

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

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

In order to conduct a mathematical discussion, it is useful to consider the sequence of integrals for 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 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 one implies stochastic dominance at higher orders, which means that stochastic dominance at the order two is a less stringent requirement for ordering distributions. 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 put efforts in finding necessary (but not sufficient) conditions which, if they are violated, dismiss irrelevant alternatives. In that spirit, Bawa (1975), Fishburn (1977) and Jean (1984) introduced *lower partial moments* (LPM) of order s ($s \geq 0$) for 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 symmetrically upper and lower deviations from the mean or from the target as the variance does. It puts attention on the left tail of the distribution.

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

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

If F dominates G at the order s, a consequence is the ordering of partial moments starting from order $s-1$. For instance, the ordering of semivariances is a necessary condition for stochastic dominance at the order one, two and three, but not for the order four.

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

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

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

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

corresponds exactly to the definition of 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 not supervising PhD students. These persons should not be taken into account when comparing countries. So the student utility function $u(x)$ has a lower bounded support, $[z, \infty]$ where z represents a minimal level of activity which should make the difference between an occasional writer and a regular 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 according to the terminology introduced by Atkinson (1987). But it should be noted that this notion (without that name) is already present in Hanoch and Levy (1969).

The particular shape of the lower bounded interval $[z, \infty[$ also means that we could decide to be interested by the right tail of the distribution and no longer by the truncated left tail. To illustrate this idea, we introduce the concept of 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. Very roughly speaking, a ranking based on UPM tries to maximise the number of productive academics, while a ranking based on LPM tries to minimize the number of improductive academics.

We have now to investigate if 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 exists provided there exist a $j > s-1$ so that we have*

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

Proof:

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

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.

□

Let us now investigate the relation existing between ordering UPM and stochastic dominance for $s = 1$ and $s = 2$. For $s = 1$, the definition of UPM implies that

$$\tilde{F}_1(x) = \int_x^{\infty} f(t)dt = 1 - F_1(x) \tag{15}$$

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

$$F(x) \leq G(x) \Leftrightarrow 1 - F(x) \geq 1 - G(x) \qquad \forall x \in [0, +\infty[\tag{16}$$

This type of symmetry was already noted by Hanoch and Levy (1969). Thus first order stochastic dominance of F over G also means that the probability of getting x or more is not smaller with F than it is with G, for every x.

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 $UPM_F^1(x) \geq UPM_G^1(x)\ \forall x$.

We have shown that stochastic dominance at the order one and at the order two is strictly equivalent to the ordering of upper partial moments functions of order zero and one $\forall x \in [0, \infty[$. 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 when we proposed a new criterion.

3.3 Connexion with income inequality measurement

Ordering the distributions of authors' scores is closely related to the literature of income inequality measurement. In his founding article Atkinson (1970) drew extensively on the same literature of decision making under uncertainty, replacing the individual utility function by a social welfare function. An important question in this context is the axiomatic construction of the social welfare function and the discussion of its properties.

This parallel being made and justified by the common ground (expected utility literature), we can now borrow some of the concepts of the field of income inequality measurement. Typically, z is seen as a poverty level for income measurement when it is a minimum level of activity in the context of our study. 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 promoted by Foster, Greer and Thorbecke 1984) are mere transformations of LPM. They can thus immediately be translated in 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 contained for instance in Davidson and Duclos (2000) apply.

The last concept we can borrow from the literature of income inequality measurement is the generalised Lorenz curve. The generalised Lorenz curve is defined as the graph of the proportion of individuals (horizontal axis) having a given level of income. Considering N discrete observations y_j and their order statistics $y_{(j)}$, we have the empirical generalised Lorenz' curve

$$\widehat{GL}(i/N) = \sum_{j=1}^{i} y_{(j)} \qquad 1 \leq i \leq N \tag{17}$$

It is tightly related to the left truncated mean. If z is the truncation point, the proportion p of individuals below z is $p = F(z)$. The left truncated mean is

$$\mathrm{E}(Y|Y \leq z) = \frac{\int_0^z y dF(y)}{F(z)} \tag{18}$$

when the generalised Lorenz curve is simply

$$GL(F(z)) = p\frac{\int_0^z y dF(y)}{F(z)} = \int_0^z y dF(y) \tag{19}$$

Using the change of variable $t = F(y)$, we arrive at the usual writing of Lorenz curves popularized by Gastwirth (1971)

$$GL(p) = \int_0^p F^{-1}(t)\, dt. \tag{20}$$

As shown in the literature (se e.g. Foster and Shorrocks 1988), generalised Lorenz ordering is equivalent to second order stochastic ordering. But despite the identity of implied rankings, lower partial moments are not equal to left truncated moments as it is apparent from the comparison of (8) and (18).

Following the same mathematical steps, we can build a modified generalised Lorenz curve, starting from right truncated moments and thus considering the proportion p of the population above a given level z, implying thus this time $p = 1 - F(z)$. We arrive at the final result

$$\begin{aligned} MGL(1 - F(z)) &= \int_z^\infty y dF(y) = \int_{F^{-1}(1-p)}^\infty y dF(y) \\ MGL(p) &= \int_{1-p}^1 F^{-1}(t)\, dt. \end{aligned} \tag{21}$$

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

As well as generalised Lorenz dominance is equivalent to the second order stochastic dominance, we can prove the same type of equivalence between the modified generalised Lorenz ordering and second order stochastic dominance[6].

[6] A formal proof can be based on 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.

3.4 Comparing countries

We have understood that for comparing countries it is essential to define a minimum level of activity z. We can define an absolute level, by reference to a minimum number of published papers of a given quality over the 10 years of the sample. For instance, 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 scored journals. We can accept $z = 10$ 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 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 above z is greater in country i than it is in country j, whatever the value of z. This is a head count measure, which has been widely criticised in the economic literature of income inequality. This measure counts the proportion of authors above the minimum line, but does not make any consideration of their distribution. Neither does it change if an author above the minimum increases his production. It just says that for every value of the minimum line x above the given z, the proportion of valuable academics is greater in i than in j.

Second order stochastic dominance put forward the mean activity gap. The activity gap $\mathrm{Max}(y-x, 0)$ 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 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 for every value of x above z.

4 The choice of a parametric distribution

A large part of the literature (see references in Davidson and Duclos 2000) estimates lower partial moment functions from discrete data without making any hypothesis on the distribution of the sample. These estimators are obtained by replacing the theoretical cumulative distribution function by its natural estimator, $\hat{F}(x) = \sum \mathbb{1}(y_i < x)/N$ and integrals by discrete sums. For LPM and UPM, we have:

$$\begin{aligned} \widehat{LPM}^s_F(x) &= \frac{1}{N}\sum_{i=1}^{N}(x-y_i)^s \mathbb{1}(y_i \leq x) \\ \widehat{UPM}^s_F(x) &= \frac{1}{N}\sum_{i=1}^{N}(y_i-x)^s \mathbb{1}(y_i \geq x). \end{aligned} \tag{22}$$

where $\mathbb{1}$ is the indicator function. Statistical properties of this type 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 incentives to find a parametric formulation for the distribution of y. The direct estimate of the curve $LPM^s_F(x)$ is sensitive to sampling errors and to observations in the extreme tails of the distribution. For instance testing for stochastic dominance needs to define a grid over x where partial moment functions are compared.

The chosen grid may lead to the veto of the extremes so that distributions have to be ordered according to the observed maximum of the sample. Some authors recommend consequently to operate some trimming in the data (see for instance Cowell and Victoria-Feser (2001) and the references cited therein). But this practice can be justified only when the range of the samples to be compared is fairly the same. Our samples have a common minimum, but their maxima vary a lot (from 90 to 216). As noted and advocated by Cowell and Victoria-Feser (1996), a parametric formulation of the distribution opens first the way to classes of estimators with well defined sampling properties. Second, it allows to detect outliers resulting from coding errors. And finally, we would add that it operates some kind of smoothing for the estimated distribution (the posterior predictive distribution of the sample in the Bayesian language). For the present case, the first argument is particularly important because a parametric formulation for the distribution of the data opens naturally the way to Bayesian inference.

On top of these traditional justifications for using a parametric distribution, there is the low number of observations which is specific to our problem. As underlined in the introduction, the population of academic authors is very small compared to the total population. Moreover, we put our interest on the right tail of the distribution. Thus it is useful to impose some structure by mean of a parametric distribution in order to be able to extract more information from the data. The problem is of course the choice of an adequate parametric distribution which makes 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) and developed in McDonald and Xu (1995) 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:

$$f_{GB2}(y|a,b,p,q) = \frac{a y^{ap-1}}{b^{ap} B(p,q)[1+(y/b)^a]^{p+q}} \tag{23}$$

with $y > 0$ and where $B(p,q)$ is the Beta function. The parameters a, b, p and q have to be positive[7]. The analytical expression of its un-centred moments[8] is

$$\begin{aligned} \mathrm{E}(Y^h) = \mu'_h &= b^h \frac{B(p+h/a, q-h/a)}{B(p,q)} \\ &= b^h \frac{\Gamma(p+h/a)\Gamma(q-h/a)}{\Gamma(p)\Gamma(q)} \end{aligned} \tag{25}$$

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}

[7] In the most general writing, 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 saw, a was positive, so we finally adopted this more simple notation of the density.

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

$$\begin{aligned} &\mu_1 = \mu'_1 && \sigma^2 = \mu'_2 - \mu_1^2 \\ &\alpha_3 = (\mu'_3 - 3\mu'_2\mu_1 + 2\mu_1^3)/\sigma^3 && \alpha_4 = (\mu'_4 - 4\mu'_3\mu_1 + 6\mu'_2\mu_1^2 - 3\mu_1^4)/\sigma^4 \end{aligned} \tag{24}$$

moments involve confluent hypergeometric functions. Wilfling (1996) has derived necessary conditions for the Lorenz ordering of two GB2 distributions. 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[9].

4.2 Related distributions

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

$$\begin{aligned} f_{GG}(y|a,\beta,p) &= \lim_{q\to\infty} f_{GB2}(y|a, b=\beta q^{1/a}, p, q) \\ &= \frac{a y^{ap-1}}{\beta^{ap}\Gamma(p)} \exp[-(y/\beta)^a] \end{aligned} \tag{26}$$

The cumulative distribution function is

$$F(x) = G[(x/\beta)^a; p] \tag{27}$$

where G is the incomplete Gamma function

$$G(x;p) = \frac{1}{\Gamma(p)} \int_0^x t^{p-1} e^{-t} dt \tag{28}$$

The un-centred moments are easily derived using the properties of the gamma distribution[10]

$$\mathrm{E}[Y^h] = \beta^h \frac{\Gamma(p + h/a)}{\Gamma(p)} \tag{29}$$

There are no parametric restrictions for the existence of the moments, except positivity restrictions.

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

$$f_W(y|a,\beta) = f_{GG}(y|a,\beta,p=1) = \frac{a}{\beta}(y/\beta)^{a-1}\exp(-(y/\beta)^a) \tag{30}$$

with a and β positive. The cumulative distribution function is

$$F_W(x) = 1 - \exp(-(y/\beta)^a) \tag{31}$$

The un-centred moments reduce to

$$\mathrm{E}(Y^h) = \beta^h \Gamma(1 + h/a). \tag{32}$$

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

$$\begin{aligned} f_{P2}(y|\alpha) &= f_{GB2}(y|a=1, b=1, p=1, q=\alpha) \\ &= \alpha[1+y]^{-(\alpha+1)} \end{aligned} \tag{33}$$

[9] We have investigated several other distributions such as the Singh-Maddala or Burr XII, Dagun or Burr III. But all were either rejected by the data or dominated by other distributions.

[10] An alternative way to define the generalised gamma density is to start from a random variable Z with a standard gamma distribution and make the change of variable $Y = \beta Z^{1/a}$ with Jacobian $(a/\beta)(y/\beta)^{a-1}$.

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

$$f(x) = \alpha\sigma^{\alpha}x^{-(\alpha+1)}\mathbb{I}(x > \sigma) \tag{34}$$

The cumulative distribution function is

$$F(x) = 1 - (\sigma/x)^{\alpha}. \tag{35}$$

The un-centred moments are

$$\mathrm{E}(X^h) = \sigma^h \frac{\alpha}{\alpha - h} \qquad \text{if } \alpha > h \tag{36}$$

They are obtained as a particular case of the moments of GB2 using the previous change of variable and a recursion over h. Or they can be computed directly from the expression of the density (which is a simpler solution). The moment of order h exists only if $\alpha > h$. This is a rather restrictive condition (for instance, with the value of α predicted by Lotka's law, the mean does not exist) and may preclude the evaluation of upper partial moment functions. 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.

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 apply the idea, the main stream Bayesian literature (see references below) defines a sample statistics $g(y_d)$ to compute a tail area probability

$$\Pr\{g(y) \geq g(y_d)|M\} \tag{37}$$

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 the model. For Box (1980), this probability is computed in reference to the predictive distribution attached to model M

$$p(y|M) = \int f(y|\theta, M)\,\varphi(\theta|M)\,d\theta \tag{38}$$

which represents all the possible outcomes of the data y if model M (data density and prior distribution) were true. In this writing, $f(y|\theta, M)$ 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. Rubin (1984) says "posterior meaning conditional on observed values and predictive meaning the distribution of a future observable quantity". This last approach knew a significant development in the statistical literature with Meng (1994), Gelman and

Meng (1996) or Gelman, Meng and Stern (1996) or 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. Whereas with the pure predictive approach, concern is stressed on reliability of the model for every possible outcome, at the cost of specifying an informative prior for θ.

The posterior predictive approach is particularly convenient in a MCMC setting. Suppose that we have already obtained N draws from the posterior density. Given a draw θ_i, we draw an hypothetical sample y_i^{rep} from the sampling distribution. We compute the $p-$value as the proportion of cases in which the simulated transformation $g(y_i^{rep})$ exceed the realised value $g(y_d)$

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{I}[g(y_i^{rep}) > g(y_d)] \tag{39}$$

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) develops the idea that it is better to avoid choosing a particular transformation and consequently prefers to compare directly a non-parametric estimate of the density of the observed sample with the posterior predictive density of the model. Among the various measures of discrepancy between densities, the ϕ divergence is popular in the Bayesian literature, especially the Kullback-Leibler divergence and the χ^2 distance. However both are asymmetric. If f and g are two densities to be compared, the distance between f and g is not the same as that between g and f when measured with the above two instruments. On the contrary, the Hellinger distance defined as

$$D_H = 2(1 - \int \sqrt{f(x)\, g(x)}\, dx) \tag{40}$$

is symmetric. We have to approximate D_H on a predefined grid x_j of points covering the sample range. Given N posterior draws θ_i, posterior draws of D_H are obtained using

$$D_{H_i} = 2(1 - \frac{d}{3}\sum_j w_j \sqrt{\hat{f}(x_j) f_M(x_j|\theta_i)}) \tag{41}$$

where d is the size of the increment of x_j, w_j a weight being 1 for end points, 2 for even points and 4 for odd points (Simpson's rule). $\hat{f}(x_j)$ is a kernel estimate of the sample density evaluated at x_j and $f(x_j|\theta_i)$ is the data density also evaluated at x_j conditionally on θ_i. We obtain N values D_{H_i} which approximate the small sample distribution of D_H. With these N draws, we are free to compute several quantities. We can decide for instance to select the model producing the smallest average distance if we compute the mean of the distribution.

4.4 Classical and Bayesian inference

McDonald and Ransom (1979) or McDonald (1984) suggest estimation techniques adapted to grouped data which are based either on a multinomial likelihood function or on a minimum χ^2 distance. For individual data on the contrary, direct maximum likelihood estimators or method of moments are available. However, the statistical literature is abundant in exemples of numerical difficulties for applying the maximum likelihood method for the generalised gamma distribution. For instance, Johnson, Kotz and Balakrishnan (1994) report that it may fail to exist. Tsionas (2001) reports a list of similar exemples. We suspect easily that these difficulties extend to the GB2 distribution which can be viewed as a generalisation of the GG.

We propose a generalised method of moments (see e.g. Gouriéroux and Monfort (1996, p 21-22) for a reference) which seems to be numerically more stable than the direct maximum likelihood approach. Let us define $\hat{\mu}_h$ as the sample centred moment of order h

$$\hat{\mu}_h = \frac{1}{N}\sum_i (y_i - \bar{y})^h$$

for $h > 1$ and $\hat{\mu}_1$ as the sample mean. The theoretical centred moments corresponding to parameter θ are noted $\mu_h(\theta)$. Let us define $d_h(\theta)$ as the difference between empirical and theoretical moments

$$d_h(\theta) = \hat{\mu}_h - \mu_h(\theta)$$

and Ω as

$$\Omega = \frac{1}{N}(z_1, \cdots, z_p)'(z_1, \cdots, z_p)$$

where z_h is a vector of size N, with the ith element equal to $(y_i - \bar{y})^h - \mu_h(\tilde{\theta})$ and $\tilde{\theta}$ is any consistent estimator of the true value of the parameter θ (for instance the GMM estimator with $\Omega = Id$). An application of the generalised method of moments in this context aims at minimizing the distance between theoretical and empirical moments of the distribution within the metric Ω:

$$\hat{\theta} = \underset{\theta}{\mathrm{ArgMin}}\,(d_1(\theta), \cdots, d_p(\theta))'\Omega^{-1}(d_1(\theta), \cdots, d_p(\theta))$$

As the maximum number of parameters (the size of θ) we have is 4, we have chosen $p = 4$ to estimate three models: GB2, GG and Weibull.

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. In Tsionas (2000) a similar approach is proposed for the Weibull density. This reparameterisation is not so convenient to compute partial moments functions, so we prefer to use the original parameterisation and a Metropolis algorithm to integrate the posterior density.

4.5 Posterior results

We took a uniform prior for all the parameters. We used an independent Metropolis with a Student importance function centred on the moments estimates and with 15 degrees of freedom. We iterated the method 3 times with 20 000 draws each time. The algorithm converged nicely. This was checked using CUMSUM graphs. Posterior means and standard deviations are presented in Table 2. It also gives the posterior mean of the Hellinger distance.

Figure 2 presents the graph of the posterior Hellinger distance between a non-parameteric estimate of the data density and the posterior predictive density evaluated on the same grid. It clearly appears from this graph that the Weibull model is the preferred one because it

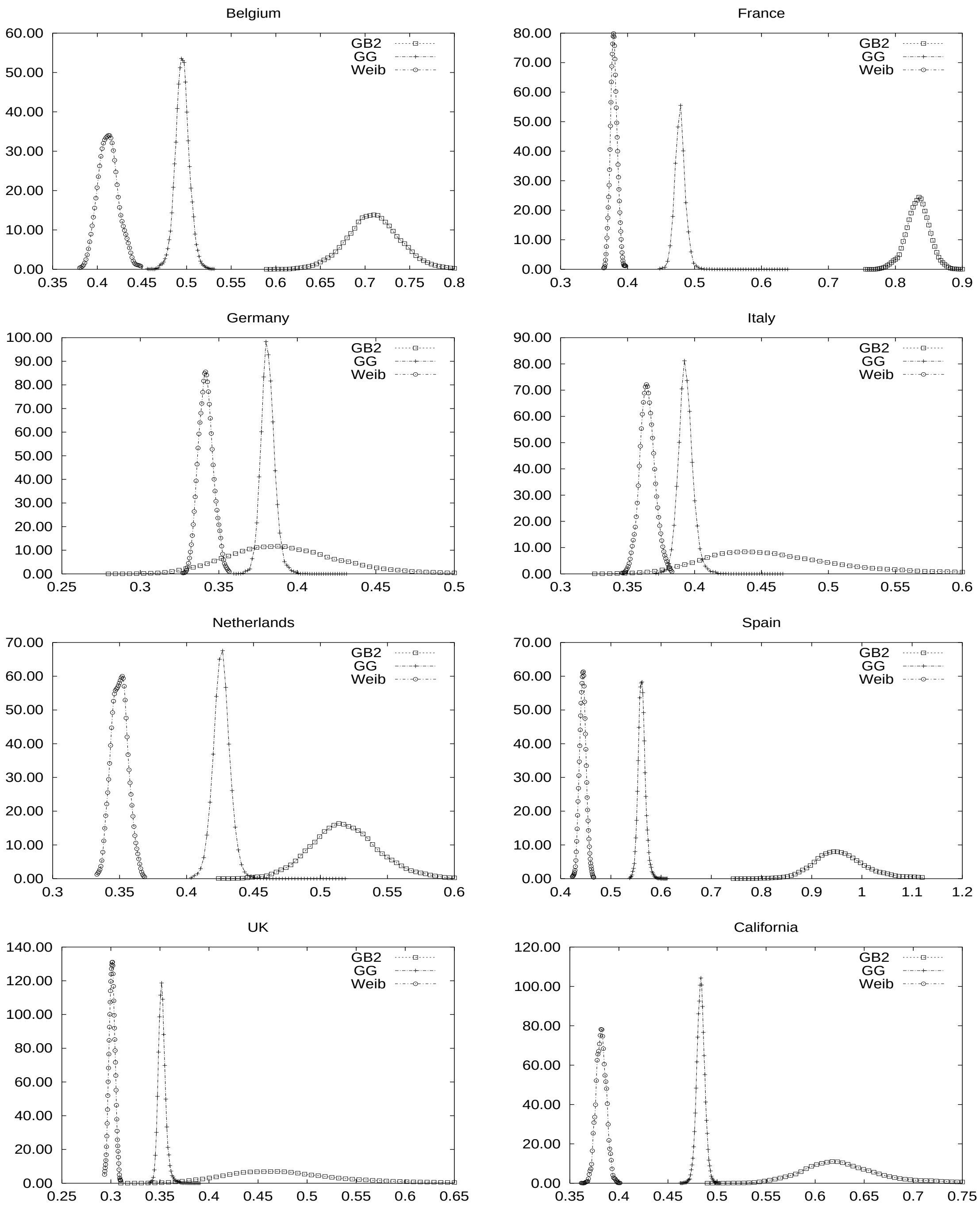


Figure 2: Posterior density of the Hellinger distance between data density and posterior predictive density

Table 2: Bayesian posterior results

	Bel	Fra	Ger	Ita	Net	Spa	UK	Cal
a	1.288 [0.107]	1.138 [0.055]	0.759 [0.076]	0.824 [0.096]	0.796 [0.064]	1.995 [0.437]	0.898 [0.154]	1.156 [0.191]
b	12.426 [0.403]	13.977 [0.346]	10.234 [1.438]	7.653 [3.73]	38.576 [0.519]	9.657 [1.147]	10.765 [0.988]	13.183 [0.607]
p	0.186 [0.019]	0.145 [0.009]	1.012 [0.179]	0.776 [0.186]	0.548 [0.073]	0.064 [0.016]	0.523 [0.147]	0.260 [0.061]
q	8.592 [0.270]	7.580 [0.157]	16.829 [2.153]	12.151 [3.139]	19.637 [0.647]	13.228 [1.658]	11.829 [0.733]	9.610 [0.504]
Hel	0.704	0.832	0.377	0.429	0.510	0.932	0.445	0.606
a	0.650 [0.025]	0.481 [0.015]	0.530 [0.012]	0.481 [0.011]	0.575 [0.019]	0.719 [0.028]	0.568 [0.016]	0.737 [0.021]
b	0.322 [0.042]	0.081 [0.016]	0.061 [0.008]	0.032 [0.005]	0.222 [0.036]	0.351 [0.044]	0.140 [0.022]	0.775 [0.067]
p	0.941 [0.049]	1.520 [0.090]	1.852 [0.072]	2.309 [0.098]	1.228 [0.070]	0.716 [0.042]	1.459 [0.078]	0.698 [0.028]
Hel	0.493	0.475	0.380	0.391	0.424	0.560	0.350	0.483
a	0.748 [0.022]	0.758 [0.014]	0.846 [0.016]	0.840 [0.018]	0.773 [0.018]	0.764 [0.020]	0.809 [0.011]	0.738 [0.013]
b	0.373 [0.022]	0.306 [0.011]	0.280 [0.010]	0.265 [0.011]	0.432 [0.020]	0.269 [0.014]	0.418 [0.010]	0.542 [0.019]
Hel	0.411	0.379	0.341	0.364	0.350	0.444	0.301	0.382

The first group of lines is related to the Generalised Beta-II distribution. The second group to the Generalised Gamma. The third group to the Weibull. Standard deviations are between brackets. The Hel line gives the posterior mean of the Hellinger distance.

displays the smallest Hellinger distance. This proves that the branch we have selected in McDonald's tree was the right one. It corresponds to a large value of q and later a p restricted to one. As this type of restriction seems to be fairly well accepted, it dismisses by the same way distributions like the Singh-Maddala (1976) distribution (also named Burr XII)(finite q and $p = 1$) or Burr III($q = 1$ and finite p).

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 :

$$UPM^0(x|a,b) = 1 - F(x) = \exp(-\left(\frac{x}{b}\right)^a) \tag{42}$$

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 theorem:

Theorem 2 *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.*

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

$\square$

Theorem 3 *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.

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

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

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.

$\square$

Theorem 4 *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 $b_F \geq b_G$.

Proof: 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.

□

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 as in the classical framework. But as these curves are in fine 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 $\underset{j}{\text{Min}}\, d^s(x_j|\theta)$ is positive:

$$
\begin{aligned}
\Pr(\underset{j}{\text{Min}}\, d^s(x_j|\theta) \geq 0) &= \int_\theta \mathbb{1}[\underset{j}{\text{Min}}\, d^s(x_j|\theta) > 0]\, \mu(\theta|y)\, d\theta \\
&\simeq \frac{1}{N}\sum_i \mathbb{1}[\underset{j}{\text{Min}}\, d^s(x_j|\theta_i) > 0]
\end{aligned} \tag{43}
$$

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 dominates another one. Table 3 shows that few numbers are large, indicating

Table 3: Posterior probability of complete stochastic dominance

	Cal	Net	UK	Bel	Fra	Spa	Ger	Ita
Cal	1.000	0.950	1.000	0.646	0.881	0.875	1.000	1.000
Net	0.000	1.000	0.702	0.187	0.230	0.377	1.000	0.997
UK	0.000	0.005	1.000	0.007	0.001	0.022	0.970	0.929
Bel	0.000	0.014	0.043	1.000	0.648	0.709	1.000	0.999
Fra	0.000	0.000	0.000	0.000	1.000	0.606	0.965	0.993
Spa	0.000	0.000	0.000	0.000	0.001	1.000	0.208	0.583
Ger	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.295
Ita	0.000	0.000	0.000	0.000	0.000	0.000	0.058	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.

that we have total ordering only for very few cases, e.g. most partial moment functions are crossing. It is quite sure that Italy is dominated by all the other countries, except perhaps by Germany (but Italy does not dominate Germany). California, on the other side of the scale, dominates all the other countries. We encourage the reader to examine further this table. 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: 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	0.859	0.668	1.000	1.000	1.000	1.000
UK	0.000	0.014	0.000	0.111	0.995	0.997	1.000	1.000
Bel	0.000	0.030	0.122	0.000	0.991	0.999	1.000	1.000
Fra	0.000	0.000	0.000	0.000	0.000	0.930	1.000	1.000
Spa	0.000	0.000	0.000	0.000	0.006	0.000	0.855	0.940
Ger	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.620
Ita	0.000	0.000	0.000	0.000	0.000	0.005	0.132	0.000

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 are equal to one. Netherlands dominates all other european countries.

There is an ambiguity between Belgium and UK. The ordering between France, Spain, Germany and Italy is obtained even if some probabilities are not very strong. Going to the second order with Table 5 does not change much the situation, but simply increases some probabilities. We still have ambiguity between the UK and Belgium.

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	0.963	0.649	0.999	1.000	1.000	1.000
UK	0.000	0.007	0.000	0.102	0.948	0.997	1.000	1.000
Bel	0.000	0.085	0.449	0.000	0.990	0.999	1.000	1.000
Fra	0.000	0.000	0.000	0.000	0.000	0.929	1.000	1.000
Spa	0.000	0.000	0.000	0.000	0.009	0.000	0.958	0.985
Ger	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.627
Ita	0.000	0.000	0.000	0.000	0.000	0.005	0.191	0.000

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

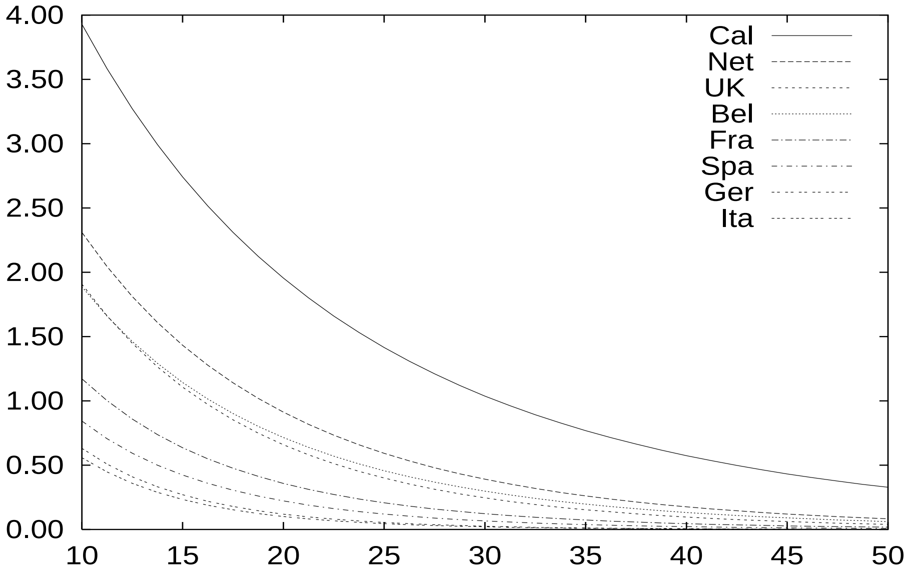


Figure 3: UPM functions for second order stochastic dominance

Remark:

Because it has a single parameter, the Pareto distribution leads to partial moment functions

$$UMP^0(x) = 1 - F(x) = \sigma^\alpha x^{-\alpha} \tag{44}$$

that can never intersect, provided the σ are the same, which is the case here. Consequently it is not possible in this model to make the difference between complete and restricted stochastic ordering. Our empirical results show that this distinction is important. This is a stylised fact against Lotka's law.

6 Conclusion

Using a parametric and a Bayesian approach, we could 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 at the cost of loosing the nice analytical results concerning the ordering of partial moment functions. We are roughly confident in the results obtained with the Weibull model as this model is both simple and flexible enough.

Concerning the empirical results, we have not ranked institutions, but countries. In doing so, we have considered in a way a picture of the whole academic system, because we took into account both very productive academics and the mass of academics having a very small scientific activity. This penalises countries having numerous economic departments, mixing good and not so good academics. 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 this choice is forbidden for him, he would be very much in trouble because UK does not show up in a better way, which is the big surprise of this study. Figure 3 suggests however a group of three leading european countries: Netherlands, Belgium and UK with the last two ones being very close from each other.

APPENDIX

Table 6: Top journal ranking

Journals	Opinions	Citations
American-Economic-Review	10	10
Econometrica	10	10
Journal-of-Economic-Theory	10	8
Journal-of-Political-Economy	10	10
Quarterly-Journal-of-Economics	10	10
Review-of-Economic-Studies	10	8

Journals	Opinions	Citations
American-Political-Science-Review	8	0
International-Economic-Review	8	6
Journal-of-Econometrics	8	8
Journal-of-Economic-Literature	8	10
Journal-of-Finance	8	10
Journal-of-Financial-Economics	8	10
Journal-of-International-Economics	8	6
Journal-of-Labor-Economics	8	6
Journal-of-Law-and-Economics	8	8
Journal-of-Monetary-Economics	8	8
Journal-of-Money,-Credit,-and-Banking	8	0
Journal-of-Public-Economics	8	8
Journal-of-the-American-Statistical-Association	8	0
Michigan-Law-Review	8	0
Rand-Journal-of-Economics	8	8
Review-of-Economics-and-Statistics	8	8
Yale-Law-Journal	7	0
Accounting-Review	6	6
American-Journal-of-Agricultural-Economics	6	6
Brookings-Papers-on-Economic-Activity	6	8
Demography	6	8
Econometric-Theory	6	4
Economica	6	6
Economic-Journal	6	8
Economics-Letters	6	4
Economic-Theory	6	0
European-Economic-Review	6	6
Games-and-Economic-Behavior	6	4
Industrial-and-Labor-Relations-Review	6	8
International-Journal-of-Game-Theory	6	0
Journal-of-Applied-Econometrics	6	4
Journal-of-Banking-and-Finance	6	6
Journal-of-Business-and-Economic-Statistics	6	8
Journal-of-Comparative-Economics	6	6
Journal-of-Development-Economics	6	6
Journal-of-Economic-Behavior-and-Organization	6	6
Journal-of-Economic-Dynamics-and-Control	6	6
Journal-of-Economic-Growth	6	0
Journal-of-Economic-History	6	6
Journal-of-Economic-Methodology	6	0
Journal-of-Economic-Perspectives	6	8
Journal-of-Economics-and-Management-Strategy	6	0

Journals	Opinions	Citations
Journal-of-Environmental-Economics-and-Management	6	6
Journal-of-Financial-and-Quantitative-Analysis	6	6
Journal-of-Health-Economics	6	6
Journal-of-Human-Resources	6	8
Journal-of-Industrial-Economics	6	6
Journal-of-Law,-Economics-and-Organization	6	0
Journal-of-Mathematical-Economics	6	4
Journal-of-Risk-and-Insurance	6	4
Journal-of-Risk-and-Uncertainty	6	6
Journal-of-Urban-Economics	6	6
Macroeconomic-Dynamics	6	0
Marketing-Science	6	6
Mathematical-Methods-of-Operations-Research	6	0
National-Tax-Journal	6	6
Oxford-Bulletin-of-Economics-and-Statistics	6	6
Public-Choice	6	6
Regional-Science-and-Urban-Economics	6	4
Scandinavian-Journal-of-Economics	6	4
Social-Choice-and-Welfare	6	4
Urban-Studies	6	6

REFERENCES

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

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

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.

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.

Gouriéroux Ch. and A. Monfort (1996) *Simulated-Based Econometric Methods.* Oxford University Press, Oxford.

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.

Kirman, A. and Dahl, M. (1994) Economic research in Europe. *European Economic Review* 38, 505-522.

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.

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

McDonald J.B. and M.R Ransom (1979) Functional forms, Estimation techniques and the distribution of income. *Econometrica* 47(6), 1513-1525.

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

McDonald J.B. and Y.J. Xu (1995) A generalisation of the Beta distribution with applications. *Journal of Econometrics* 66, 133-152.

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.

Singh S.K. and Maddala G.S. (1976). A function for size distribution of incomes. *Econometrica*, 44 (5), 963–970.

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.

Wilfling B. (1996) Lorenz ordering of generalized Beta-II income distributions. *Journal of Econometrics* 71, 381-388.

Zheng B. (2001) Statistical inference for poverty measures with relative poverty lines. *Journal of Econometrics* 101, 337-356.